



HAL
open science

Estimation of the inverse scatter matrix for a scale mixture of Wishart matrices under Efron-Morris type losses

Djamila Boukehil, Dominique Fourdrinier, Fatiha Mezoued, William Strawderman

► To cite this version:

Djamila Boukehil, Dominique Fourdrinier, Fatiha Mezoued, William Strawderman. Estimation of the inverse scatter matrix for a scale mixture of Wishart matrices under Efron-Morris type losses. 2020. hal-02503972

HAL Id: hal-02503972

<https://hal.science/hal-02503972v1>

Preprint submitted on 10 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Estimation of the inverse scatter matrix for a scale mixture of Wishart matrices under Efron-Morris type losses

Djamila Boukehil ^{*} Dominique Fourdrinier [†] Fatiha Mezoued [‡]
William E. Strawderman [§]

February 17, 2020

Abstract

We consider estimation of the inverse scatter matrix Σ^{-1} for a scale mixture of Wishart matrices under various Efron-Morris type losses, $\text{tr}\{(\hat{\Sigma}^{-1} - \Sigma^{-1})^2 S^k\}$ for $k = 0, 1, 2, \dots$, where S is the sample covariance matrix. We improve on the standard estimators $a S^+$, where S^+ denotes the Moore-Penrose inverse of S and a is a positive constant, through an unbiased estimator of the risk difference between the new estimators and $a S^+$. Thus we demonstrate that improvements over the standard estimators under a Wishart distribution can be extended under mixing. We give a unified treatment of the two cases where S is invertible ($S^+ = S^{-1}$) and where S is singular.

AMS 2010 subject classifications: 62H12, 62F10, 62C99.

Key words and phrases: Elliptically symmetric distributions, scale mixtures of normals, scale mixtures of Wisharts, Stein-Haff type identity, covariance matrix, scatter matrix, SURE estimators.

^{*}Djamila Boukehil is PHD student, École Nationale Supérieure de Statistique et d'Économie Appliquée, LAMOPS, Tipaza, Algeria, and Université de Normandie, UNIROUEN, UNIHAVRE, INSA Rouen, LITIS; avenue de l'Université, BP 12, 76801 Saint-Étienne-du-Rouvray, France. E-mail: djamila.boukehil@etu.univ-rouen.fr.

[†]Dominique Fourdrinier is Professor, Université de Normandie, UNIROUEN, UNIHAVRE, INSA Rouen, LITIS; avenue de l'Université, BP 12, 76801 Saint-Étienne-du-Rouvray, France and International Laboratory SSP & QF, National Research Tomsk State University, Russia, E-mail: Dominique.Fourdrinier@univ-rouen.fr.

[‡]Fatiha Mezoued is Professor, École Nationale Supérieure de Statistique et d'Économie Appliquée, LAMOPS, Tipaza, Algeria. E-mail: famezoued@yahoo.fr.

[§]William E. Strawderman is Professor, Rutgers University, Department of Statistics, 561 Hill Center, Busch Campus Piscataway NJ 08854, USA. E-mail: straw@stat.rutgers.edu.

1 Introduction

The estimation of precision matrices has recently received much attention. Most of the literature is devoted to the case where the observation has a multivariate normal distribution so that the sample covariance matrix S has a Wishart distribution. In that case, Σ^{-1} , the inverse of the covariance matrix Σ , is often called the precision matrix. This classical multivariate setting has been studied by Efron and Morris [4], Haff [7], Dey [2], Krishnamoorthy and Gupta [13], Dey et al. [3], Zhou et al. [19], and Tsukuma and Konno [17]. Note that, in these papers, S is assumed to be invertible. However, in the setting where the dimension of S is larger than the sample size, its inverse does not exist. Then estimation of Σ^{-1} is based on the Moore-Penrose generalized inverse S^+ . This approach was, for instance, developed in Kubokawa and Srivastava [14].

Estimation of Σ^{-1} has been extended to the distributional framework of elliptically contoured distributions. Thus Fourdrinier, Mezoued and Wells [6] considered a large subclass of the elliptically contoured distributions and improvements over the usual estimators (proportional to S^{-1} or S^+ according to the invertibility or singularity of S) were provided under quadratic loss. Note that the form of the underlying distributions was such that no unbiased estimator of the risk was available.

In this article, we focus on the subclass of scale mixture of Wishart distributions for S . Mixtures of Wishart are more and more popular in modeling. Thus Jian et al. [11] use mixtures of Wishart distributions for modeling diffusion weighted magnetic resonance imaging while Yang et al. [18] consider such distributions for the change of detection of polarimetric synthetic aperture radar images. Also, Nielsen et al. [15], in a Bayesian context, deal with a Wishart mixture model for modeling dynamic functional connectivity. In a more theoretical paper, Haff et al. [10] consider the problem of estimating the mixing density of a continuous mixture of Wishart distributions, the parameter being a covariance matrix; they also provide an application to finance.

Here, we observe S from the mixture model

$$\begin{cases} S \mid V \sim \mathcal{W}_p(n, V \Sigma) \\ V \sim \mathcal{H}(\cdot) \end{cases} \quad (1.1)$$

where $\mathcal{W}_p(n, V \Sigma)$ denotes the Wishart distribution with n degrees of freedom and covariance matrix $V \Sigma$ and where $\mathcal{H}(\cdot)$ is a distribution on \mathbb{R}_+ . Note that we may view the distribution of S as that of $V \tilde{S}$, where \tilde{S} has the Wishart distribution $\mathcal{W}_p(n, \Sigma)$ and is independent of V . Thus, for any function g such that its expectation with respect to (1.1) exists, we have

$$E_{\Sigma}[g(S)] = E_{\mathcal{H}} \left[E_{\Sigma} \left[g(\tilde{S} V) \mid V \right] \right],$$

where E_{Σ} denotes the expectation with respect to the model in (1.1) and $E_{\mathcal{H}}$ the expectation with respect to the mixing distribution \mathcal{H} . We will often use this form in the calculation below. Note also that, under the mixture model in (1.1), the population covariance matrix is proportional to Σ (the expectation of S is $n E[V] \Sigma$). For convenience, we refer to Σ as a scatter matrix and to Σ^{-1} as a precision matrix.

Estimating Σ^{-1} , we consider a wide class of losses which are data-based losses in the sense that they are function of S . More precisely, we deal with the family of loss functions

$$L_k\{\Sigma^{-1}, \hat{\Sigma}^{-1}\} = \text{tr}\{[\hat{\Sigma}^{-1} - \Sigma^{-1}]^2 S^k\}, \quad (1.2)$$

for $k = 0, 1, 2, \dots$ and the associated risks

$$R_k(\Sigma^{-1}, \hat{\Sigma}^{-1}) = E_{\Sigma}[L_k\{\Sigma^{-1}, \hat{\Sigma}^{-1}\}]. \quad (1.3)$$

Note that these losses are reminiscent of the losses used by Efron and Morris [4]. In the Wishart case, loss L_k has been mostly considered for $k = 0, 1$ and 2 . For example, for the non singular Wishart distribution, loss L_1 was considered by Efron and Morris [4], while both L_0 and L_1 were considered by Haff [8] (actually, as for L_0 , the identity matrix S^0 was replaced by an arbitrary positive definite matrix Q). Kubokawa and Srivastava [14] considered L_0, L_1, L_2 for the singular case. The consideration of the case where $k \geq 3$ is natural and we will see that, in the cases where $k = 1, 2$ and 3 , optimal estimators in (1.4) below exist while, the case $k \geq 4$ is similar to the case $k = 0$, where there are no such estimators.

Note also that we provide a unified approach to the settings where S is invertible and S is singular. To this end, the notation S^+ for the Moore-Penrose inverse of S when S is non invertible is also used when S is invertible ($S^+ = S^{-1}$). For Model (1.1), the usual, or reference, estimators of Σ^{-1} are of the form

$$\hat{\Sigma}_a^{-1} = a S^+, \quad (1.4)$$

where a is a positive constant. It is well known that such estimators may be inappropriate (see the references above). The alternative estimators that we consider are of the form

$$\hat{\Sigma}_{a,c}^{-1} = a S^+ + c S G(S), \quad (1.5)$$

where c is a constant and the matrix function $G(S)$ is homogeneous in the sense that

$$\exists \alpha \in \mathbb{R} \quad G(V S) = V^\alpha G(S). \quad (1.6)$$

The paper is organized as follows. In Section 2, we develop an unbiased estimator of the risk difference between $\hat{\Sigma}_{a,c}^{-1}$ in (1.5) and $\hat{\Sigma}_a^{-1}$ in (1.4), relying on a Stein-Haff type identity given in Fourdrinier, Haddouche and Mezoued [5] and valid for the two cases where S is invertible and S is singular. Note that this unbiased estimator of risk difference holds for all the losses in (1.2) and implies directly conditions under which $\hat{\Sigma}_{a,c}^{-1}$ improves over $\hat{\Sigma}_a^{-1}$. In Section 3, we consider the possibility of the existence of an optimal estimator in the class (1.4) of reference estimators, for each loss in (1.2). When such an optimal estimator does not exist, we propose an alternative estimator. In Section 4, when the order of homogeneity in (1.6) is -2 , we provide explicit conditions for improvement of estimators of the form (1.5) over the reference estimators in (1.4). In addition, we give analogous results for estimators of the form $a S^+ + c r(\text{tr}\{S\}) S G(S)$ when G is orthogonally invariant and r is a real valued function. In Section 5, we illustrate the theory with examples reminiscent of those studied by Haff [7] and Dey [2], for each loss in (1.2). Some concluding remarks are given in Section 6. Finally, we provide an Appendix which gathers most of the proofs of the results presented in this article.

2 A unified expression for the risk difference

Under model (1.1), assume finiteness of the risk in (1.3) of the estimators $\hat{\Sigma}_{a,c}^{-1} = a S^+ + c S G(S)$ and $\hat{\Sigma}_a^{-1} = a S^+$. Then the risk difference between $\hat{\Sigma}_{a,c}^{-1}$ and $\hat{\Sigma}_a^{-1}$ is given by

$$\begin{aligned} \Delta(G) &= R_k(\Sigma^{-1}, a S^+ + c S G(S)) - R_k(\Sigma^{-1}, a S^+) \\ &= \mathbb{E}_\Sigma [\text{tr}\{\{a S^+ - \Sigma^{-1} + c S G(S)\}^2 S^k - (a S^+ - \Sigma^{-1})^2 S^k\}] \\ &= \mathbb{E}_\Sigma [c^2 \text{tr}\{S^k [S G(S)]^2\} + 2 c \text{tr}\{S G(S)(a S^+ - \Sigma^{-1}) S^k\}] \\ &= \mathbb{E}_\Sigma [c^2 \text{tr}\{S^k [S G(S)]^2\} + 2 a c \text{tr}\{S^+ S^{k+1} G(S)\} - 2 c \text{tr}\{\Sigma^{-1} S^{k+1} G(S)\}]. \end{aligned} \quad (2.1)$$

As the integrand term of $\Delta(G)$ in (2.1) depends on the unknown covariance matrix Σ , we need the so-called Stein-Haff identity for the Wishart distribution $\mathcal{W}_p(n, \Sigma)$. This identity was derived by Stein [16] and Haff [9] when S is invertible while Kubokawa and Srivastava [14], Konno [12] and Chetelat and Wells [1] considered the case where S is singular. In the following lemma, we present a unified approach to these two cases. Its statement (under an equivalent form where $\tilde{S} G(\tilde{S})$ is replaced by $\tilde{S}^+ \tilde{S} G(\tilde{S})$) and its proof can be found in Fourdrinier, Haddouche and Mezoued [5]. Although it is reduced to the Wishart case, it will turn out to be sufficient to tackle the mixing model (1.1).

Lemma 2.1 (Stein-Haff-type-identity). *Let \tilde{S} be a $p \times p$ matrix having a Wishart distribution. For any $p \times p$ matrix function $G(\tilde{S})$ which is weakly differentiable with respect to \tilde{S} and such that $\mathbb{E}_\Sigma [\text{tr}\{\Sigma^{-1} \tilde{S} G(\tilde{S})\}] < \infty$, we have*

$$\mathbb{E}_\Sigma [\text{tr}\{\Sigma^{-1} \tilde{S} G(\tilde{S})\}] = \mathbb{E}_\Sigma [\text{tr}\{(n - (n \wedge p) - 1) \tilde{S}^+ \tilde{S} G(\tilde{S}) + 2 \tilde{S}^+ \tilde{S} D_{\tilde{S}} \{G^\top(\tilde{S}) \tilde{S}\}\}],$$

with $n \wedge p = \min(n, p)$ and where the differential operator $D_{\tilde{S}}$ for a matrix \tilde{S} is defined by

$$D_{\tilde{S}} = \left(\frac{1}{2} (1 + \delta_{ij}) \frac{\partial}{\partial \tilde{S}_{ij}} \right)_{1 \leq i, j \leq p},$$

with $\delta_{ij} = 1$ when $i = j$ and $\delta_{ij} = 0$ when $i \neq j$.

In order to apply Lemma 2.1 to Model (1.1), note that, for any appropriate function $G(S)$ and for any $v > 0$,

$$\mathbb{E}_\Sigma [G(S) | V = v] = \mathbb{E}_\Sigma [G(V \tilde{S}) | V = v] = \mathbb{E}_\Sigma [G(v \tilde{S})].$$

Theorem 2.1. Under Model (1.1) and loss (1.2), assume that the risk of $\hat{\Sigma}_a^{-1}$ in (1.4) and that of $\hat{\Sigma}_{a,c}^{-1}$ in (1.5) are finite. Assume also that the function $G(S)$ in (1.5) satisfies (1.6) and is such that $S^k G(S)$ satisfies the conditions of Lemma 2.1. Then the risk difference between $\hat{\Sigma}_{a,c}^{-1}$ in (1.5) and $\hat{\Sigma}_a^{-1} = a S^+$ is expressed through \tilde{S} as

$$\begin{aligned} \Delta(G) = & \mathbb{E}_\Sigma \left[c^2 \mu_{k+2\alpha+2} \text{tr}\{\tilde{S}^k [\tilde{S} G(\tilde{S})]^2\} \right. \\ & + 2c [a \mu_{k+\alpha} - \mu_{k+\alpha+1}(n - (n \wedge p) - 1)] \text{tr}\{\tilde{S}^+ \tilde{S}^{k+1} G(\tilde{S})\} \\ & \left. - 4c \mu_{k+\alpha+1} \text{tr}\{\tilde{S}^+ \tilde{S} D_{\tilde{S}} \{G^\top(\tilde{S}) \tilde{S}^{k+1}\}\} \right], \end{aligned} \quad (2.2)$$

where α is the order of homogeneity in (1.6) and where, for $\beta \in \mathbb{R}$, $\mu_\beta = \mathbb{E}_\mathcal{H}[V^\beta]$ is the moment of order β of V .

Proof. Note that $\Delta(G)$ in (2.1) can be written as,

$$\Delta(G) = \mathbb{E}_\Sigma [\eta(S) + \beta_\Sigma(S)],$$

where

$$\eta(S) = c^2 \text{tr}\{S^k [S G(S)]^2\} + 2ac \text{tr}\{S^+ S^{k+1} G(S)\},$$

and

$$\beta_\Sigma(S) = -2c \text{tr}\{\Sigma^{-1} S^{k+1} G(S)\}.$$

Conditioning on V , we have

$$\Delta(G) = \mathbb{E}_\mathcal{H} [\gamma(V)], \quad (2.3)$$

where $\mathbb{E}_\mathcal{H}$ is the expectation with respect to \mathcal{H} and

$$\gamma(V) = \mathbb{E}_\Sigma [\eta(S) | V] + \mathbb{E}_\Sigma [\beta_\Sigma(S) | V].$$

According to the change of variable $S = V \tilde{S}$, we have

$$\gamma(V) = \mathbb{E}_\Sigma [\eta(V \tilde{S}) | V] + \mathbb{E}_\Sigma [\beta_\Sigma(V \tilde{S}) | V].$$

where

$$\eta(V \tilde{S}) = c^2 V^{k+2\alpha+2} \text{tr}\{\tilde{S}^k [\tilde{S} G(\tilde{S})]^2\} + 2ac V^{k+\alpha} \text{tr}\{\tilde{S}^+ \tilde{S}^{k+1} G(\tilde{S})\}, \quad (2.4)$$

and

$$\beta_\Sigma(V \tilde{S}) = -2c V^{k+\alpha+1} \text{tr}\{\Sigma^{-1} \tilde{S}^{k+1} G(\tilde{S})\}.$$

Now, by Lemma 2.1,

$$\mathbb{E}_\Sigma [\beta_\Sigma(V \tilde{S}) | V] = \mathbb{E}_\Sigma [\beta(V \tilde{S}) | V],$$

where

$$\begin{aligned} \beta(V \tilde{S}) = & -2c V^{k+\alpha+1} (n - (n \wedge p) - 1) \text{tr}\{\tilde{S}^+ \tilde{S}^{k+1} G(\tilde{S})\} \\ & - 4c V^{k+\alpha+1} \text{tr}\{\tilde{S}^+ \tilde{S} D_{\tilde{S}} \{G^\top(\tilde{S}) \tilde{S}^{k+1}\}\}. \end{aligned} \quad (2.5)$$

Then, according to (2.4) and (2.5),

$$\gamma(V) = \mathbb{E}_\Sigma \left[\eta(V \tilde{S}) + \beta(V \tilde{S}) \mid V \right],$$

where the integrand term is independent of Σ , and hence,

$$\begin{aligned} \gamma(V) &= \mathbb{E}_\Sigma [c^2 V^{k+2\alpha+2} \text{tr}\{\tilde{S}^k [\tilde{S} G(\tilde{S})]^2\} \\ &\quad + 2c [a V^{k+\alpha} - V^{k+\alpha+1} (n - (n \wedge p) - 1)] \text{tr}\{\tilde{S}^+ \tilde{S}^{k+1} G(\tilde{S})\} \\ &\quad - 4c V^{k+\alpha+1} \text{tr}\{\tilde{S}^+ \tilde{S} D_{\tilde{S}} \{G^\top(\tilde{S}) \tilde{S}^{k+1}\}\} \mid V]. \end{aligned} \quad (2.6)$$

Since \tilde{S} is independent of V , unconditioning gives the desired result in (2.2). \square

The following corollary is immediate.

Corollary 2.1. *Under the conditions of Theorem 2.1,*

$$\begin{aligned} \delta(G) &= c^2 \mu_{k+2\alpha+2} \text{tr}\{\tilde{S}^k [\tilde{S} G(\tilde{S})]^2\} \\ &\quad + 2c [a \mu_{k+\alpha} - \mu_{k+\alpha+1} (n - (n \wedge p) - 1)] \text{tr}\{\tilde{S}^+ \tilde{S}^{k+1} G(\tilde{S})\} \\ &\quad - 4c \mu_{k+\alpha+1} \text{tr}\{\tilde{S}^+ \tilde{S} D_{\tilde{S}} \{G^\top(\tilde{S}) \tilde{S}^{k+1}\}\}, \end{aligned} \quad (2.7)$$

is an unbiased estimator of the risk difference between $\Sigma_{a,c}^{-1}$ and Σ_a^{-1} , and hence, a sufficient condition for improvement of $\Sigma_{a,c}^{-1}$ over Σ_a^{-1} is that $\delta(G)$ in (2.7) is non positive.

The result of Corollary 2.1 is in contrast to the setting in Fourdrinier, Mezoued and Wells [6] (when $k = 0$) where there is no unbiased estimator of the risk difference and upper bounds of $\Delta(G)$ are needed. Note that, when the mixture distribution is the Dirac distribution at a fixed point V , Model (1.1) reduces to the Wishart distribution $\mathcal{W}_p(n, V \Sigma)$. Thus the result in Theorem 2.1 expresses that the existence of an unbiased estimator of $\Delta(G)$ is preserved under mixing, under the assumption (1.6) of homogeneity of G .

In the next section, we consider the existence of optimal estimators in the class $\{a S^+ / a > 0\}$ for the losses in (1.2). When there are no such estimators, we propose alternatives.

3 Reference estimators $a S^+$

In this subsection, for the usual class of estimators of Σ^{-1} of the form $a S^+$ with $a > 0$, we consider the possibility of the existence of an optimal constant a^* , that is, such that, for $k = 0, 1, 2, \dots$, its risk in (1.3) satisfies

$$\forall a > 0, R_k(\Sigma^{-1}, a^* S^+) \leq R_k(\Sigma^{-1}, a S^+), \quad (3.1)$$

under Model (1.1). In the case where such an optimal constant a^* does not exist, we suggest constants a_0 which possess satisfying properties.

For $k = 0, 1, 2, \dots$, the risk of the estimator $a S^+$ under the model (1.1) is given by

$$\begin{aligned} R_k(\Sigma^{-1}, a S^+) &= \mathbb{E}_\Sigma[\text{tr}\{\{\Sigma^{-1} - a S^+\}^2 S^k\}] \\ &= \mathbb{E}_\Sigma[a^2 \text{tr}\{S^k (S^+)^2\} - 2a \text{tr}\{\Sigma^{-1} S^k S^+\}] + \mathbb{E}_\Sigma[\text{tr}\{\Sigma^{-2} S^k\}]. \end{aligned}$$

Note that minimizing $R_k(a S^+, \Sigma^{-1})$ in a reduces to minimizing

$$a^2 \mathbb{E}_\Sigma[\text{tr}\{S^k (S^+)^2\}] - 2a \mathbb{E}_\Sigma[\text{tr}\{\Sigma^{-1} S^k S^+\}],$$

that is, to minimizing

$$a^2 \mu_{k-2} \mathbb{E}_\Sigma[\text{tr}\{\tilde{S}^k (\tilde{S}^+)^2\}] - 2a \mu_{k-1} \mathbb{E}_\Sigma[\text{tr}\{\Sigma^{-1} \tilde{S}^k \tilde{S}^+\}], \quad (3.2)$$

which gives as a minimizing value

$$a^* = \frac{\mu_{k-1} \mathbb{E}_\Sigma[\text{tr}\{\Sigma^{-1} \tilde{S}^k \tilde{S}^+\}]}{\mu_{k-2} \mathbb{E}_\Sigma[\text{tr}\{\tilde{S}^k (\tilde{S}^+)^2\}]} \quad (3.3)$$

Applying Lemma 2.1 to the numerator in (3.3) with $G(\tilde{S}) = \tilde{S}^k (\tilde{S}^+)^2$, using the fact that $\tilde{S}^+ = \tilde{S}^+ \tilde{S} \tilde{S}^+ = \tilde{S} (\tilde{S}^+)^2$, gives

$$\mathbb{E}_\Sigma[\text{tr}\{\Sigma^{-1} \tilde{S}^k \tilde{S}^+\}] = (n - (n \wedge p) - 1) \mathbb{E}_\Sigma[\text{tr}\{\tilde{S}^k (\tilde{S}^+)^2\}] + 2 \mathbb{E}_\Sigma[\text{tr}\{\tilde{S}^+ \tilde{S} D_{\tilde{S}} \{\tilde{S}^k \tilde{S}^+\}\}].$$

Hence the optimal value in (3.3) is expressed as

$$a^* = \frac{\mu_{k-1}}{\mu_{k-2}} [n - (n \wedge p) - 1] + 2 \frac{\mu_{k-1}}{\mu_{k-2}} \frac{\mathbb{E}_\Sigma[\text{tr}\{\tilde{S}^+ \tilde{S} D_{\tilde{S}} \{\tilde{S}^k \tilde{S}^+\}\}]}{\mathbb{E}_\Sigma[\text{tr}\{\tilde{S}^k (\tilde{S}^+)^2\}]} \quad (3.4)$$

The optimal value a^* in (3.4) may depend on Σ according to the values of k . When this is not the case, it is a constant which corresponds to an optimal estimator among the reference estimators. When it does depend on Σ , we provide alternative values of a which give rise to estimators having reasonable properties. This is specified in the following proposition whose proof is postponed to the appendix.

Proposition 3.1. *Assume that S has the mixture of Wishart distribution in (1.1) and that, under (1.3), the risk of S^+ is finite.*

When $k = 1, 2$ and 3 , the optimal value a^ in (3.4) is constant (so that (3.1) is satisfied) and equals*

- $\frac{1}{\mu_{-1}} [(n \vee p) - (n \wedge p) - 1]$ if $k = 1$, with $n \wedge p = \min(n, p)$ and $n \vee p = \max(n, p)$;
- $\mu_1 (n \vee p)$ if $k = 2$;
- $\frac{\mu_2}{\mu_1} [n + p + 1]$ if $k = 3$.

When $k = 0$ and $k \geq 4$, the optimal value a^* in (3.4) depends on Σ . However, a lower bound a_0^- and an upper bound a_0^+ exist for a^* so that, in each case, for $a < a_0^-$, $a_0^- S^+$ improves over a S^+ and, for $a > a_0^+$, $a_0^+ S^+$ improves over a S^+ . Specifically,

- $a_0^- = \frac{\mu_{-1}}{\mu_{-2}} [(n \vee p) - 2((n \wedge p) + 1)]$ and $a_0^+ = \frac{\mu_{-1}}{\mu_{-2}} [(n \vee p) - (n \wedge p) - 3]$ if $k = 0$;
- $a_0^- = \frac{\mu_{k-1}}{\mu_{k-2}} [(n \vee p) + (n \wedge p) + k - 2]$ and $a_0^+ = \frac{\mu_{k-1}}{\mu_{k-2}} [(n \vee p) + (k - 2)(n \wedge p) + k - 2]$ if $k \geq 4$.

Comment It is worth noting that, taking the mixture as degenerate, that is, in the Wishart case, when $p \leq n$, $E_\Sigma[S^+] = E_\Sigma[\tilde{S}^{-1}]$ exists if and only if $n - p - 1 > 0$; when $k = 1$, this condition coincides with the finiteness of the risk of \tilde{S}^+ in (1.3), and hence the optimal constant $a^* = n - p - 1$ is positive. Similarly, when $k = 0$, the finiteness of the risk corresponds to the existence of $E_\Sigma[(\tilde{S}^+)^2]$ which is satisfied if and only if $n - p - 3 > 0$, that is, a_0^+ is positive (however, a_0^- may be non positive). Also note that, when $k \geq 2$, the risk is always finite.

Finiteness of the risk of S^+ in the mixture model (1.1) requires the additional assumption of finiteness of $\mu_\beta = E_{\mathcal{H}}[V^\beta]$ for $\beta = k - 2$, $k - 1$ and k .

Finally, note that, in the Wishart case, when $k = 0$ and \tilde{S} is invertible, an unbiased estimator of Σ^{-1} is available; this is $(n - p - 1)\tilde{S}^{-1}$. As $a_0^+ = n - p - 3$, we have $n - p - 1 \geq a_0^+$, so that the ‘‘reasonable’’ estimator $a_0\tilde{S}^{-1}$ dominates the unbiased estimator $(n - p - 1)\tilde{S}^{-1}$.

4 Improved estimators

In this section, for a fixed $a > 0$, we consider improved estimators over aS^+ . We focus on functions $G(S)$ in (1.5) which are homogeneous of order $\alpha = -2$. In this case, the following result helps to establish a unified connection between improved estimators in the Wishart case and the general mixture case (1.1). Although it is not used in the rest of the paper, it is of independent interest.

Lemma 4.1. *Let $k = 0, 1, 2, \dots$ be fixed. Suppose that the function $G(S)$ in (1.6) is homogeneous of order -2 and weakly differentiable, and that all moments in Theorem 2.1 exist. Then, for any fixed $a \geq 0$, if $aS^+ + cSG(S)$ improves on aS^+ for some c , in the case of the Wishart model, then $(\mu_{k-1}/\mu_{k-2})(aS^+ + cSG(S))$ improves on $(\mu_{k-1}/\mu_{k-2})aS^+$ in the mixture model (1.1).*

Proof. For $\alpha = -2$, Theorem 2.1 gives an expression for the risk difference in the mixture case as

$$\begin{aligned}\Delta(G) &= \mathbb{E}_\Sigma \left[c^2 \mu_{k-2} \operatorname{tr}\{\tilde{S}^k [\tilde{S} G(\tilde{S})]^2\} \right. \\ &\quad + 2c [a \mu_{k-2} - \mu_{k-1} (n - (n \wedge p) - 1)] \operatorname{tr}\{\tilde{S}^+ \tilde{S}^{k+1} G(\tilde{S})\} \\ &\quad \left. - 4c \mu_{k-1} \operatorname{tr}\{\tilde{S}^+ \tilde{S} D_{\tilde{S}} \{G^\top(\tilde{S}) \tilde{S}^{k+1}\}\} \right] \\ &= c \mu_{k-1} \mathbb{E}_\Sigma \left[c \frac{\mu_{k-2}}{\mu_{k-1}} M(\tilde{S}) + 2N \left(a \frac{\mu_{k-2}}{\mu_{k-1}}, \tilde{S} \right) \right]\end{aligned}\quad (4.1)$$

where

$$M(\tilde{S}) = \operatorname{tr}\{\tilde{S}^k [\tilde{S} G(\tilde{S})]^2\}$$

and

$$N \left(a \frac{\mu_{k-2}}{\mu_{k-1}}, \tilde{S} \right) = a \frac{\mu_{k-2}}{\mu_{k-1}} - (n - (n \wedge p) - 1) \operatorname{tr}\{\tilde{S}^+ \tilde{S}^{k+1} G(\tilde{S})\} - 2 \operatorname{tr}\{\tilde{S}^+ \tilde{S} D_{\tilde{S}} \{G^\top(\tilde{S}) \tilde{S}^{k+1}\}\}.$$

If $S \sim \mathcal{W}_p(n, \Sigma)$ then $\mathcal{H}(\cdot)$ is degenerate at $V = 1$ and $\mu_k = 1$ for all k . Hence, if the final expression is negative for $\mu_{k-2} = \mu_{k-1} = 1$ and particular values of a and c (i.e. in the Wishart case), then it is also negative in the mixture case (1.1) with a replaced by $a \mu_{k-1}/\mu_{k-2}$ and c by $c \mu_{k-1}/\mu_{k-2}$. \square

It is worth noting, from Section 3, that the optimal constant a^* for $k = 1, 2$ and 3 (or the “reasonable” constant a_0 for $k = 0$ and $k \geq 4$) in the Wishart case corresponds to the optimal constant $a^* \mu_{k-1}/\mu_{k-2}$ (or $a_0 \mu_{k-1}/\mu_{k-2}$) in the mixture case, so that improvement over the optimal (or “reasonable”) estimator for the Wishart case corresponds to an improvement over the optimal (or “reasonable”) estimator in the mixture case through multiplication by μ_{k-1}/μ_{k-2} .

Now we consider a useful modification of the estimators in (1.5) where the function G is orthogonally invariant and homogeneous of order $\alpha = -2$. These estimators are of the form

$$\hat{\Sigma}_{a,c,r}^{-1} = a S^+ + c r(\operatorname{tr}\{S\}) S G(S), \quad (4.2)$$

where c is a constant, the matrix function $G(S)$ is homogeneous of order -2 and r is a real valued function. We assume that G is orthogonally invariant, that is, of the form

$$G = H_1 \Psi(L) H_1^\top, \quad (4.3)$$

where $L = (\operatorname{diag}(l_i))_{1 \leq i \leq n \wedge p}$ with $l_1 > \dots > l_{n \wedge p} > 0$ and H_1 is the diagonalizing matrix satisfying $H_1^\top H_1 = I_{n \wedge p}$ in the singular value decomposition¹ of S , i.e. $S = H_1 L H_1^\top$, and $\Psi(L)$ is a $(n \wedge p) \times (n \wedge p)$ diagonal matrix, $\operatorname{diag}(\psi_1(L), \dots, \psi_{n \wedge p}(L))$. Note that homogeneity of order -2 implies that $\Psi(VL) = V^{-2} \Psi(L)$.

¹When $p \leq n$, H_1 is a $p \times p$ orthogonal matrix (i.e. $H_1 H_1^\top = H_1^\top H_1 = I_p$) while, when $p > n$, H_1 is a $p \times n$ semi-orthogonal matrix (i.e. $H_1^\top H_1 = I_n$).

Lemma 4.2. *Let $k = 0, 1, 2, \dots$ be fixed. Under model (1.1) and loss (1.2), suppose that the function $G(S)$ is orthogonally invariant as in (4.3), homogeneous of order -2 and weakly differentiable, and that all moments in Theorem 2.1 exist. Suppose also that, for any $i = 1, \dots, n \wedge p$, $\psi_i(L) > 0$.*

Suppose that, for some $c > 0$, expression (2.7) is non positive so that the estimator $\hat{\Sigma}_{a,c}^{-1} = aS^+ + cSG(S)$ in (1.5) improves over the estimator $\hat{\Sigma}_a^{-1} = aS^+$ in (1.4). Then the estimator $\hat{\Sigma}_{a,c,r}^{-1} = aS^+ + cr(\text{tr}\{S\})SG(S)$ in (4.2) improves over $\hat{\Sigma}_a^{-1} = aS^+$ provided that the function r is differentiable and that, for any $t \geq 0$, $0 \leq r(t) \leq 1$ and $r'(t) \geq 0$.

Proof. Following the proof of Theorem 2.1, expression of $\gamma(V)$ in (2.6) where $G(\tilde{S})$ is replaced by $r(V \text{tr}\{\tilde{S}\})G(\tilde{S})$ and where $\alpha = -2$ becomes

$$\begin{aligned} \gamma(V) = & \mathbb{E}_\Sigma [c^2 r^2 (V \text{tr}\{\tilde{S}\}) V^{k-2} \text{tr}\{\tilde{S}^k [\tilde{S} G(\tilde{S})]^2\} \\ & + 2cr(V \text{tr}\{\tilde{S}\}) [aV^{k-2} - V^{k-1}(n - (n \wedge p) - 1)] \text{tr}\{\tilde{S}^+ \tilde{S}^{k+1} G(\tilde{S})\} \\ & - 4cV^{k-1} \text{tr}\{\tilde{S}^+ \tilde{S} D_{\tilde{S}}\{r(V \text{tr}\{\tilde{S}\}) G^\top(\tilde{S}) \tilde{S}^{k+1}\} \mid V\}]. \end{aligned} \quad (4.4)$$

As $0 \leq r(t) \leq 1$, using Lemma A.3 applied to $r(V \text{tr}\{\tilde{S}\})$ (so that $\partial/\partial l_i r(V \text{tr}\{L\}) = V r'(V \text{tr}\{L\})$ replaces $r'(\text{tr}\{L\})$), $\gamma(V)$ in (4.4) is bounded above by

$$\gamma_1(V) = \mathbb{E}_\Sigma [r(V \text{tr}\{\tilde{S}\}) (A(\tilde{S}) V^{k-2} - B(\tilde{S}) V^{k-1}) \mid V] \quad (4.5)$$

where

$$A(\tilde{S}) = (c^2 \text{tr}\{\tilde{S}^k [\tilde{S} G(\tilde{S})]^2\} + 2ca \text{tr}\{\tilde{S}^+ \tilde{S}^{k+1} G(\tilde{S})\}) \geq 0$$

and

$$B(\tilde{S}) = 2c((n - (n \wedge p) - 1) \text{tr}\{\tilde{S}^+ \tilde{S}^{k+1} G(\tilde{S})\} + 2 \text{tr}\{\tilde{S}^+ \tilde{S} D_{\tilde{S}}\{G^\top(\tilde{S}) \tilde{S}^{k+1}\}\}).$$

We have used the fact that the term involving $r'(\text{tr}\{\tilde{S}\})$ in Lemma A.3 is eliminated since, by assumption, $c > 0$, $r'(\text{tr}\{\tilde{S}\}) \geq 0$ and $\text{tr}\{G^\top(\tilde{S}) \tilde{S}^{k+1}\} = \text{tr}\{\Psi L^{k+1}\} \geq 0$. Since $\delta(G)$ in (2.7) is non positive by assumption, and since for, $\alpha = -2$,

$$\delta(G) = A(\tilde{S}) \mu_{k-2} - B(\tilde{S}) \mu_{k-1} \leq 0, \quad (4.6)$$

the term $B(\tilde{S})$ is non negative.

Now expressing the difference in risk in (2.3)

$$\Delta(G) = \mathbb{E}_\mathcal{H} [\gamma(V)] \leq \mathbb{E}_\mathcal{H} [\gamma_1(V)] = \mathbb{E}_\mathcal{H} \left[\mathbb{E}_\Sigma [r(V \text{tr}\{\tilde{S}\}) (A(\tilde{S}) V^{k-2} - B(\tilde{S}) V^{k-1}) \mid V] \right]$$

according to (4.5). Then, reversing the conditioning,

$$\begin{aligned} \Delta(G) & \leq \mathbb{E}_\Sigma \left[\mathbb{E}_\mathcal{H} [r(V \text{tr}\{\tilde{S}\}) (A(\tilde{S}) V^{k-2} - B(\tilde{S}) V^{k-1}) \mid \tilde{S}] \right] \\ & \leq \mathbb{E}_\Sigma \left[r(V_0(\tilde{S}) \text{tr}\{\tilde{S}\}) \mathbb{E}_\Sigma \left[\mathbb{E}_\mathcal{H} [(A(\tilde{S}) V^{k-2} - B(\tilde{S}) V^{k-1}) \mid \tilde{S}] \right] \right], \end{aligned}$$

as $A(\tilde{S})V^{k-2} - B(\tilde{S})V^{k-1}$ changes sign once from $+$ to $-$ at $V_0(\tilde{S}) = A(\tilde{S})/B(\tilde{S})$ and $r(V \operatorname{tr}\{\tilde{S}\})$ is non decreasing in V . Hence, using (4.6),

$$\begin{aligned} \Delta(G) &\leq \mathbb{E}_\Sigma \left[r(V_0(\tilde{S}) \operatorname{tr}\{\tilde{S}\}) \right] \mathbb{E}_\Sigma \left[A(\tilde{S}) \mu_{k-2} - B(\tilde{S}) \mu_{k-1} \right] \\ &\leq 0. \end{aligned}$$

□

5 Examples

In this section, we illustrate the theory developed in Sections 3 and 4 with estimators similar to those proposed by Haff and Dey.

5.1 Haff type estimators

We consider estimators

$$\hat{\Sigma}_{HF}^{-1} = a S^+ + c S S^+ / \operatorname{tr}\{S\}, \quad (5.1)$$

that is, of the form $a S^+ + c S G(S)$ with $G(S) = S^+ / \operatorname{tr}\{S\}$. Note that the homogeneity order of $G(S)$ is $\alpha = -2$. This type of estimator was proposed by Haff [7] in the Gaussian case where S is invertible under a loss proportional to L_1 . Note that $G(S)$ is orthogonally invariant, and hence, thanks to Lemma 4.2, the development below applies to estimators of the form $a S^+ + c r(\operatorname{tr}\{S\}) S S^+ / \operatorname{tr}\{S\}$ where r is a real valued function such that, for any $t \geq 0$, $0 \leq r(t) \leq 1$ and $r'(t) \geq 0$.

Applying Theorem 2.1 with $G(S) = S^+ / \operatorname{tr}\{S\}$ and $\alpha = -2$, the term $\delta(G)$ in (2.7) can be written for any $k \geq 0$ as

$$\begin{aligned} \delta(G) &= c^2 \mu_{k-2} \frac{\operatorname{tr}\{\tilde{S}^k (\tilde{S} \tilde{S}^+)^2\}}{(\operatorname{tr}\{\tilde{S}\})^2} + 2c [a \mu_{k-2} - \mu_{k-1} (n - (n \wedge p) - 1)] \frac{\operatorname{tr}\{\tilde{S}^{k+1} (\tilde{S}^+)^2\}}{\operatorname{tr}\{\tilde{S}\}} \\ &\quad - 4c \mu_{k-1} \operatorname{tr} \left\{ \tilde{S}^+ \tilde{S} D_{\tilde{S}} \left\{ \frac{\tilde{S}^+ \tilde{S}^{k+1}}{\operatorname{tr}\{\tilde{S}\}} \right\} \right\} \\ &= c^2 \mu_{k-2} \frac{\operatorname{tr}\{\tilde{S}^{k+1} \tilde{S}^+\}}{(\operatorname{tr}\{\tilde{S}\})^2} + 2c [a \mu_{k-2} - \mu_{k-1} (n - (n \wedge p) - 1)] \frac{\operatorname{tr}\{\tilde{S}^k \tilde{S}^+\}}{\operatorname{tr}\{\tilde{S}\}} \\ &\quad - 4c \mu_{k-1} \operatorname{tr} \left\{ \tilde{S}^+ \tilde{S} D_{\tilde{S}} \left\{ \frac{\tilde{S}^+ \tilde{S}^{k+1}}{\operatorname{tr}\{\tilde{S}\}} \right\} \right\}, \end{aligned} \quad (5.2)$$

since $\tilde{S}^+ = \tilde{S}^+ \tilde{S} \tilde{S}^+$ is symmetric and $\tilde{S}^+ \tilde{S} = \tilde{S} \tilde{S}^+$. For the differential term $D_{\tilde{S}}\{\tilde{S}^+ \tilde{S}^{k+1} / \operatorname{tr}\{\tilde{S}\}\}$ in (5.2), thanks to Equality (A.1) based on the singular value decomposition of \tilde{S} ,

we have

$$\begin{aligned}
D_{\tilde{S}} \left\{ \frac{\tilde{S}^+ \tilde{S}^{k+1}}{\text{tr}\{\tilde{S}\}} \right\} &= D_{\tilde{S}} \left\{ \frac{H_1 L^k H_1^\top}{\text{tr}\{L\}} \right\} \\
&= \frac{1}{2} [p - (n \wedge p)] \frac{\tilde{S}^k \tilde{S}^+}{\text{tr}\{\tilde{S}\}} + k \frac{\tilde{S}^k \tilde{S}^+}{\text{tr}\{\tilde{S}\}} - \frac{\tilde{S}^{k+1} \tilde{S}^+}{(\text{tr}\{\tilde{S}\})^2} \\
&\quad + \frac{1}{2} \frac{1}{\text{tr}\{\tilde{S}\}} \sum_{i \neq j}^{n \wedge p} \frac{l_i^k - l_j^k}{l_i - l_j} + \frac{1}{2} \frac{\text{tr}\{\tilde{S}^k \tilde{S}^+\}}{\text{tr}\{\tilde{S}\}} (I_p - H_1 H_1^\top), \quad (5.3)
\end{aligned}$$

where here

$$\phi_i = \frac{l_i^k}{\text{tr}\{L\}} \quad \text{and} \quad \frac{\partial \phi_i}{\partial l_i} = k \frac{l_i^{k-1}}{\text{tr}\{L\}} - \frac{l_i^k}{(\text{tr}\{L\})^2}.$$

Substituting (5.3) in (5.2) and noticing that $\tilde{S} \tilde{S}^+ (I_p - H_1 H_1^\top) = 0$, we have

$$\begin{aligned}
\delta(G) &= c^2 \mu_{k-2} \frac{\text{tr}\{\tilde{S}^{k+1} \tilde{S}^+\}}{(\text{tr}\{\tilde{S}\})^2} + 2c [a \mu_{k-2} - \mu_{k-1} (n - (n \wedge p) - 1)] \frac{\text{tr}\{\tilde{S}^k \tilde{S}^+\}}{\text{tr}\{\tilde{S}\}} \\
&\quad - 2c \mu_{k-1} [2k + p - (n \wedge p)] \frac{\text{tr}\{\tilde{S}^k \tilde{S}^+\}}{\text{tr}\{\tilde{S}\}} + 4c \mu_{k-1} \frac{\text{tr}\{\tilde{S}^{k+1} \tilde{S}^+\}}{(\text{tr}\{\tilde{S}\})^2} \\
&\quad - 2c \mu_{k-1} \frac{1}{\text{tr}\{\tilde{S}\}} \sum_{i=1}^{n \wedge p} \sum_{j \neq i}^{n \wedge p} \frac{l_i^k - l_j^k}{l_i - l_j} \\
&= c^2 \mu_{k-2} \frac{\text{tr}\{\tilde{S}^{k+1} \tilde{S}^+\}}{(\text{tr}\{\tilde{S}\})^2} + 2c [a \mu_{k-2} - \mu_{k-1} (n - p + 2(n \wedge p) + 1 - 2k)] \frac{\text{tr}\{\tilde{S}^k \tilde{S}^+\}}{\text{tr}\{\tilde{S}\}} \\
&\quad + 4c \mu_{k-1} \frac{\text{tr}\{\tilde{S}^{k+1} \tilde{S}^+\}}{(\text{tr}\{\tilde{S}\})^2} - 2c \mu_{k-1} \frac{1}{\text{tr}\{\tilde{S}\}} \sum_{i=1}^{n \wedge p} \sum_{j \neq i}^{n \wedge p} \frac{l_i^k - l_j^k}{l_i - l_j},
\end{aligned}$$

that is, expressing $\delta(G)$ in terms of L ,

$$\begin{aligned}
\delta(G) &= \frac{c}{\text{tr}\{L\}} \left\{ c \mu_{k-2} \frac{\text{tr}\{L^k\}}{\text{tr}\{L\}} - 2 [\mu_{k-1} ((n \vee p) - (n \wedge p) - 1 + 2k) - a \mu_{k-2}] \text{tr}\{L^{k-1}\} \right. \\
&\quad \left. + 4 \mu_{k-1} \frac{\text{tr}\{L^k\}}{\text{tr}\{L\}} - 2 \mu_{k-1} B_{k+1} \right\} \quad (5.4)
\end{aligned}$$

where

$$B_{k+1} = \sum_{i=1}^{n \wedge p} \sum_{j \neq i}^{n \wedge p} \frac{l_i^k - l_j^k}{l_i - l_j}$$

(see Lemma A.2 in the Appendix). Then we have $\delta(G) \leq 0$ in (5.4) for c between 0 and

$$C_a = 2 \left[\frac{\mu_{k-1}}{\mu_{k-2}} ((n \vee p) - (n \wedge p) - 1 + 2k) - a \right] \frac{\text{tr}\{L^{k-1}\} \text{tr}\{L\}}{\text{tr}\{L^k\}} - 4 \frac{\mu_{k-1}}{\mu_{k-2}} + 2 \frac{\mu_{k-1}}{\mu_{k-2}} B_{k+1} \frac{\text{tr}\{L\}}{\text{tr}\{L^k\}}. \quad (5.5)$$

This gives rise to the following proposition.

Proposition 5.1. *For $a = a^*$ given in Proposition 3.1, any Haff type estimator $\hat{\Sigma}_{HF}^{-1} = a^* S^+ + c S S^+ / \text{tr}\{S\}$ improves on the optimal estimator $a^* S^+$, and hence, improves on any estimator $a S^+$ with $a > 0$, when $k = 1$ and 2, as soon as*

- $0 < c < 2 \frac{1}{\mu_{-1}} ((n \wedge p)[(n \wedge p) + 1] - 2)$ if $k = 1$ and $n \wedge p \geq 2$;
- $0 < c < 2 \mu_1 [(n \wedge p) + 1] - 4 \mu_1$ if $k = 2$ and $n \wedge p \geq 2$.

When $k = 3$, although $C_{a^*} > 0$, that quantity is not constant and cannot be bounded from below by a positive constant. However, for $a < a^*$, as soon as

- $0 < c < 2 \frac{\mu_2}{\mu_1} (a^* - a) = 2 \frac{\mu_2}{\mu_1} (n + p + 1 - a)$,

the estimator $\hat{\Sigma}_{HF}^{-1}$ improves on a S^+ .

For $a = a_0^-$ given in Proposition 3.1, any Haff type estimator $\hat{\Sigma}_{HF}^{-1} = a_0^- S^+ + c S S^+ / \text{tr}\{S\}$ improves on $a_0^- S^+$, and hence, improves on any estimator $a S^+$ with $a < a_0^-$, as soon as

- $0 < c < 2 \frac{\mu_{-1}}{\mu_{-2}} [(n \wedge p) - 1]$ if $k = 0$ and $n \wedge p \geq 2$;
- $0 < c < 2 \frac{\mu_{k-1}}{\mu_{k-2}} (k - 3)$ if $k \geq 4$.

Comment. In the cases $k = 3$ and $a \geq a^*$, $k = 0$ and $a > a_0^+$, $k \geq 4$ and $a > a_0^+$, our method does not lead to any improved Haff type estimator. In each case, expression of C_a in (5.5) is non negative but is not bounded below by a positive constant.

Proof. According to Proposition 3.1, when $k = 1$, we have $a^* = \frac{1}{\mu_{-1}} [(n \vee p) - (n \wedge p) - 1]$ so that the quantity C_{a^*} in (5.5) equals

$$C_{a^*} = \frac{4}{\mu_{-1}} (n \wedge p) - \frac{4}{\mu_{-1}} + \frac{2}{\mu_{-1}} (n \wedge p) [(n \wedge p) - 1] = \frac{2}{\mu_{-1}} ((n \wedge p) [(n \wedge p) + 1] - 2),$$

which gives the first result.

When $k = 2$, we have $a^* = \mu_1 (n \vee p)$, and hence,

$$\begin{aligned} C_{a^*} &= 2 \mu_1 [-(n \wedge p) + 3] \frac{(\text{tr}\{L\})^2}{\text{tr}\{L^2\}} - 4 \mu_1 + 4 \mu_1 [(n \wedge p) - 1] \frac{(\text{tr}\{L\})^2}{\text{tr}\{L^2\}} \\ &= 2 \mu_1 [(n \wedge p) + 1] \frac{(\text{tr}\{L\})^2}{\text{tr}\{L^2\}} - 4 \mu_1 \\ &\geq 2 \mu_1 [(n \wedge p) + 1] - 4 \mu_{-1}, \end{aligned}$$

since $(\text{tr}\{L\})^2 \geq \text{tr}\{L^2\}$. This gives the second result.

Now, when $k = 3$, as $a^* = \frac{\mu_2}{\mu_1} [(n \vee p) + (n \wedge p) + 1]$, although $C_{a^*} > 0$, that quantity is not constant and cannot be bounded from below by a positive constant. Indeed we have

$$\begin{aligned} C_{a^*} &= 2 \frac{\mu_2}{\mu_1} [-2(n \wedge p) + 4] \frac{\text{tr}\{L^2\} \text{tr}\{L\}}{\text{tr}\{L^3\}} - 4 \frac{\mu_2}{\mu_1} \\ &\quad + 2 \frac{\mu_2}{\mu_1} ((\text{tr}\{L\})^2 + [2(n \wedge p) - 3] \text{tr}\{L^2\}) \frac{\text{tr}\{L\}}{\text{tr}\{L^3\}} \\ &= 2 \frac{\mu_2}{\mu_1} \frac{\text{tr}\{L^2\} \text{tr}\{L\}}{\text{tr}\{L^3\}} - 4 \frac{\mu_2}{\mu_1} + 2 \frac{\mu_2}{\mu_1} \frac{(\text{tr}\{L\})^3}{\text{tr}\{L^3\}}, \end{aligned}$$

according to the expression of B_4 given in Lemma A.2. As for the positivity of C_{a^*} , note that, according to Inequality (A.5) in Lemma (A.4) with $k = 1$ and $m = 1$ and 2, we have

$$\frac{(\text{tr}\{L\})^3}{\text{tr}\{L^3\}} \geq \frac{\text{tr}\{L^2\} \text{tr}\{L\}}{\text{tr}\{L^3\}} \geq 1,$$

Hence, it is clear that $C_{a^*} \geq 0$ with probability 1. However, it is clear from that expression, that, for any $a > 0$, the corresponding C_a can be written as

$$C_a = C_{a^*} + 2 \frac{\mu_2}{\mu_1} (a^* - a) \frac{\text{tr}\{L^2\} \text{tr}\{L\}}{\text{tr}\{L^3\}},$$

so that

$$C_a \geq 2 \frac{\mu_2}{\mu_1} (a^* - a),$$

as soon as $a < a^*$. Then the result for $k = 3$ follows.

When $k = 0$, we have $a_0^- = \frac{\mu_{-1}}{\mu_{-2}} [(n \vee p) - 2((n \wedge p) + 1)]$. It is easily seen from (5.5) that

$$C_{a_0^-} = 2 \frac{\mu_{-1}}{\mu_{-2}} [(n \wedge p) + 1] \frac{\text{tr}\{L^{-1}\} \text{tr}\{L\}}{n \wedge p} - 4 \frac{\mu_{-1}}{\mu_{-2}} \geq 2 \frac{\mu_{-1}}{\mu_{-2}} [(n \wedge p) - 1],$$

since $\text{tr}\{L^{-1}\} \text{tr}\{L\} \geq n \wedge p$. Hence, according to Proposition 3.1, the result for $k = 0$ follows.

When $k \geq 4$, we have $a_0^- = \frac{\mu_{k-1}}{\mu_{k-2}} [(n \vee p) + (n \wedge p) + k - 2]$. According to the expression

of B_k in Lemma A.2, it follows from (5.5) that

$$\begin{aligned}
C_{a_0^-} &= 2 \frac{\mu_{k-1}}{\mu_{k-2}} [-2(n \wedge p) + 2k - 1] \frac{\text{tr}\{L^{k-1}\} \text{tr}\{L\}}{\text{tr}\{L^k\}} - 4 \frac{\mu_{k-1}}{\mu_{k-2}} \\
&\quad + 2 \frac{\mu_{k-1}}{\mu_{k-2}} \left\{ \sum_{m=1}^{k-2} \text{tr}\{L^m\} \text{tr}\{L^{k-1-m}\} + [2 \frac{\mu_{k-1}}{\mu_{k-2}} (n \wedge p) - k] \text{tr}\{L^{k-1}\} \right\} \frac{\text{tr}\{L\}}{\text{tr}\{L^k\}} \\
&= 2 \frac{\mu_{k-1}}{\mu_{k-2}} \frac{\text{tr}\{L^{k-1}\} \text{tr}\{L\}}{\text{tr}\{L^k\}} - 4 \frac{\mu_{k-1}}{\mu_{k-2}} + 2 \frac{\mu_{k-1}}{\mu_{k-2}} \frac{\text{tr}\{L\}}{\text{tr}\{L^k\}} \sum_{m=1}^{k-2} \text{tr}\{L^m\} \text{tr}\{L^{k-1-m}\} \\
&\geq 2 \frac{\mu_{k-1}}{\mu_{k-2}} \frac{\text{tr}\{L^{k-1}\} \text{tr}\{L\}}{\text{tr}\{L^k\}} - 4 \frac{\mu_{k-1}}{\mu_{k-2}} + 2 \frac{\mu_{k-1}}{\mu_{k-2}} (k-2) \frac{\text{tr}\{L^{k-1}\} \text{tr}\{L\}}{\text{tr}\{L^k\}} \\
&= 2 \frac{\mu_{k-1}}{\mu_{k-2}} (k-1) \frac{\text{tr}\{L^{k-1}\} \text{tr}\{L\}}{\text{tr}\{L^k\}} - 4 \frac{\mu_{k-1}}{\mu_{k-2}} \\
&\geq 2 \frac{\mu_{k-1}}{\mu_{k-2}} (k-3),
\end{aligned}$$

since, according to (A.5), $\sum_{m=1}^{k-2} \text{tr}\{L^m\} \text{tr}\{L^{k-1-m}\} \geq (k-2) \text{tr}\{L^{k-1}\}$ and $\text{tr}\{L^{k-1}\} \text{tr}\{L\} \geq \text{tr}\{L^k\}$. This gives the result for $k \geq 4$. \square

5.2 Dey type estimators

Consider estimators

$$\hat{\Sigma}_{DY}^{-1} = a S^+ + c \frac{S}{\text{tr}\{S^2\}}, \quad (5.6)$$

that is, of the form $a S^+ + c S G(S)$ with

$$G(S) = \frac{I_p}{\text{tr}\{S^2\}}. \quad (5.7)$$

Note that the homogeneity order of $G(S)$ is $\alpha = -2$. Estimators of the form (5.6) were proposed by Dey [2] in the Gaussian case where S is invertible, under loss L_0 . Note that $G(S)$ is orthogonally invariant, and hence, thanks to Lemma 4.2, the development below applies to estimators of the form $a S^+ + c r(\text{tr}\{S\}) S / \text{tr}\{S^2\}$ where r is a real valued function such that, for any $t \geq 0$, $0 \leq r(t) \leq 1$ and $r'(t) \geq 0$.

Substituting in (2.7), the term in (5.7) gives

$$\begin{aligned}
\delta(G) &= c^2 \mu_{k-2} \frac{\text{tr}\{\tilde{S}^{k+2}\}}{(\text{tr}\{\tilde{S}^2\})^2} + 2c [a \mu_{k-2} - \mu_{k-1} (n - (n \wedge p) - 1)] \frac{\text{tr}\{\tilde{S}^+ \tilde{S}^{k+1}\}}{\text{tr}\{\tilde{S}^2\}} \\
&\quad - 4c \mu_{k-1} \text{tr} \left\{ \tilde{S}^+ \tilde{S} D_{\tilde{S}} \left\{ \frac{\tilde{S}^{k+1}}{\text{tr}\{\tilde{S}^2\}} \right\} \right\}. \quad (5.8)
\end{aligned}$$

Through Equality (A.1), we have

$$\begin{aligned}
D_{\tilde{S}} \left\{ \frac{\tilde{S}^{k+1}}{\text{tr}\{\tilde{S}^2\}} \right\} &= D_{\tilde{S}} \left\{ \frac{H_1 L^{k+1} H_1^\top}{\text{tr}\{L^2\}} \right\} \\
&= H_1 \left\{ \frac{1}{2} (p - (n \wedge p)) \frac{L^k}{\text{tr}\{L^2\}} + (k+1) \frac{L^k}{\text{tr}\{L^2\}} - 2 \frac{L^{k+2}}{(\text{tr}\{L^2\})^2} \right\} H_1^\top \\
&\quad + H_1 \text{diag} \left\{ \frac{1}{2 \text{tr}\{L^2\}} \sum_{j \neq i}^{(n \wedge p)} \frac{l_i^{k+1} - l_j^{k+1}}{l_i - l_j} \right\} H_1^\top \\
&\quad + \frac{1}{2} \text{tr} \left\{ \frac{L^k}{\text{tr}\{L^2\}} \right\} (I_p - H_1 H_1^\top)
\end{aligned} \tag{5.9}$$

where

$$\phi_i = \frac{l_i^{k+1}}{\text{tr}\{L^2\}} \quad \text{and} \quad \frac{\partial \phi_i}{\partial l_i} = (k+1) \frac{l_i^k}{\text{tr}\{L^2\}} - 2 \frac{l_i^{k+2}}{(\text{tr}\{L^2\})^2}.$$

Substituting (5.9) in (5.8) we have

$$\begin{aligned}
\delta(G) &= c^2 \mu_{k-2} \frac{\text{tr}\{\tilde{S}^{k+2}\}}{(\text{tr}\{\tilde{S}^2\})^2} + 2c [a \mu_{k-2} - \mu_{k-1} (n - (n \wedge p) - 1)] \frac{\text{tr}\{\tilde{S}^+ \tilde{S}^{k+1}\}}{\text{tr}\{\tilde{S}^2\}} \\
&\quad - 4c \mu_{k-1} \text{tr} \left\{ \tilde{S}^+ \tilde{S} H_1 \left\{ \frac{1}{2} (p - (n \wedge p)) \frac{L^k}{\text{tr}\{L^2\}} + (k+1) \frac{L^k}{\text{tr}\{L^2\}} \right. \right. \\
&\quad \quad \left. \left. - 2 \frac{L^{k+2}}{(\text{tr}\{L^2\})^2} H_1^\top \right\} \right\} \\
&\quad - 4c \mu_{k-1} \text{tr} \left\{ \tilde{S}^+ \tilde{S} H_1 \text{diag} \left\{ \frac{1}{2 \text{tr}\{L^2\}} \sum_{j \neq i}^{n \wedge p} \frac{l_i^{k+1} - l_j^{k+1}}{l_i - l_j} \right\} H_1^\top \right\} \\
&\quad - 4c \mu_{k-1} \text{tr} \left\{ \frac{1}{2} \tilde{S}^+ \tilde{S} \text{tr} \left\{ \frac{L^k}{\text{tr}\{L^2\}} \right\} (I_p - H_1 H_1^\top) \right\}.
\end{aligned} \tag{5.10}$$

Expressing the terms in \tilde{S} through L , $\delta(G)$ in (5.10) becomes

$$\begin{aligned}
\delta(G) &= \frac{c}{\text{tr}\{L^2\}} \left\{ c \mu_{k-2} \frac{\text{tr}\{L^{k+2}\}}{\text{tr}\{L^2\}} + 2 [a \mu_{k-2} - \mu_{k-1} ((n \vee p) - (n \wedge p) + 2k + 1)] \text{tr}\{L^k\} \right. \\
&\quad \left. + 8 \mu_{k-1} \frac{\text{tr}\{L^{k+2}\}}{\text{tr}\{L^2\}} - 2 \mu_{k-1} \sum_{i=1}^{n \wedge p} \sum_{j \neq i}^{n \wedge p} \frac{l_i^{k+1} - l_j^{k+1}}{l_i - l_j} \right\},
\end{aligned} \tag{5.11}$$

since $\tilde{S} \tilde{S}^+ (I_p - H_1 H_1^\top) = 0$ and $(n \vee p) = n + p - (n \wedge p)$. Then we have $\delta(G) \leq 0$ for c between 0 and

$$C_a = 2 \left[\frac{\mu_{k-1}}{\mu_{k-2}} [(n \vee p) - (n \wedge p) + 2k + 1] - a \right] \frac{\text{tr}\{L^k\} \text{tr}\{L^2\}}{\text{tr}\{L^{k+2}\}} - 8 \frac{\mu_{k-1}}{\mu_{k-2}} + 2 B_{k+2} \frac{\mu_{k-1}}{\mu_{k-2}} \frac{\text{tr}\{L^2\}}{\text{tr}\{L^{k+2}\}}. \tag{5.12}$$

where

$$B_{k+2} = \sum_{i=1}^{n \wedge p} \sum_{j \neq i}^{n \wedge p} \frac{l_i^{k+1} - l_j^{k+1}}{l_i - l_j}.$$

This leads to the following proposition.

Proposition 5.2. *For $a = a^*$ given in Proposition 3.1, any Dey type estimator $\hat{\Sigma}_{DY}^{-1} = a^* S^+ + c S / \text{tr}\{S^2\}$ improves on the optimal estimator $a^* S^+$, and hence, improves on any estimator $a S^+$ with $a > 0$, when $k = 1$ and 2, as soon as*

- $0 < c < \frac{4}{\mu_{-1}} [(n \wedge p) - 1]$ if $k = 1$ and $n \wedge p \geq 2$;
- $0 < c < 2\mu_1 [(n \wedge p) - 1]$ if $k = 2$ and $n \wedge p \geq 2$.

When $k = 3$, although $C_{a^*} > 0$, that quantity is not constant and cannot be bounded from below by a positive constant. However, for $a < a^*$, as soon as

- $0 < c < 2 \frac{\mu_2}{\mu_1} (a^* - a) = 2 \frac{\mu_2}{\mu_1} (n + p + 1 - a)$,

the estimator $\hat{\Sigma}_{DY}^{-1}$ improves on $a S^+$.

When $k = 0$, for $a = a_0^-$ and $a = a_0^+$ given in Proposition 3.1, any Dey type estimator $\hat{\Sigma}_{DY}^{-1} = a_0^- S^+ + c / \text{tr}\{S^2\}$ and $\hat{\Sigma}_{DY}^{-1} = a_0^+ S^+ + c / \text{tr}\{S^2\}$ improves on $a_0^- S^+$, $a_0^+ S^+$ respectively, and hence, improves on any estimator $a S^+$ with $a < a_0^-$ and $a > a_0^+$ respectively, as soon as

- $0 < c < 4 \frac{\mu_{-1}}{\mu_{-2}} [\{(n \wedge p) + 1\} (n \wedge p) - 2]$ and $0 < c < 2 \frac{\mu_{-1}}{\mu_{-2}} [\{(n \wedge p) + 3\} (n \wedge p) - 4]$ if $n \wedge p \geq 2$;

When $k \geq 4$, for $a = a_0^-$ given in Proposition 3.1, any Dey type estimator $\hat{\Sigma}_{DY}^{-1} = a_0^- S^+ + c / \text{tr}\{S^2\}$ improves on $a_0^- S^+$, and hence, improves on any estimator $a S^+$ with $a < a_0^-$, as soon as

- $0 < c < 2 \frac{\mu_{k-1}}{\mu_{k-2}} (k - 3)$.

Comment. In the cases $k = 3$ and $a \geq a^*$, $k \geq 4$ and $a > a_0^+$, our method does not lead to any improved Dey type estimator. In each case, expression of C_a in (5.12) is non negative but is not bounded below by a positive constant.

Proof. When $k = 1$, we have $a^* = \frac{1}{\mu_{-1}} [(n \vee p) - (n \wedge p) - 1]$, then C_a in (5.12) becomes

$$C_{a^*} = \frac{8}{\mu_{-1}} \frac{\text{tr}\{L\} \text{tr}\{L^2\}}{\text{tr}\{L^3\}} - \frac{8}{\mu_{-1}} + \frac{2}{\mu_{-1}} B_3 \frac{\text{tr}\{L^2\}}{\text{tr}\{L^{k+2}\}}.$$

using the expression of B_3 in Lemma A.2, it follows from the above equality that

$$\begin{aligned} C_{a^*} &= \frac{8}{\mu_{-1}} \frac{\operatorname{tr}\{L\} \operatorname{tr}\{L^2\}}{\operatorname{tr}\{L^3\}} - \frac{8}{\mu_{-1}} + [(n \wedge p) - 1] \frac{4}{\mu_{-1}} \frac{\operatorname{tr}\{L\} \operatorname{tr}\{L^2\}}{\operatorname{tr}\{L^3\}} \\ &= \frac{4}{\mu_{-1}} [(n \wedge p) + 1] \frac{\operatorname{tr}\{L\} \operatorname{tr}\{L^2\}}{\operatorname{tr}\{L^3\}} - \frac{8}{\mu_{-1}} \\ &\geq \frac{4}{\mu_{-1}} [(n \wedge p) - 1], \end{aligned}$$

where the above inequality is due to Lemma (A.4) with $k = 1$ and $m = 2$. This gives the first result.

When $k = 2$, we have $a^* = \mu_1 (n \vee p)$. By using the expression of B_k in Lemma A.2 with $k = 4$, C_a in (5.12) equals

$$\begin{aligned} C_{a^*} &= 2\mu_1 [-(n \wedge p) + 5] \frac{(\operatorname{tr}\{L^2\})^2}{\operatorname{tr}\{L^4\}} - 8\mu_1 + 2\mu_1 B_4 \frac{\operatorname{tr}\{L^2\}}{\operatorname{tr}\{L^4\}} \\ &= 2\mu_1 [-(n \wedge p) + 5] \frac{(\operatorname{tr}\{L^2\})^2}{\operatorname{tr}\{L^4\}} - 8\mu_1 + 2\mu_1 [(\operatorname{tr}\{L\})^2 + (2(n \wedge p) - 3) \operatorname{tr}\{L^2\}] \frac{\operatorname{tr}\{L^2\}}{\operatorname{tr}\{L^4\}} \\ &= 2\mu_1 [(n \wedge p) + 2] \frac{(\operatorname{tr}\{L^2\})^2}{\operatorname{tr}\{L^4\}} - 8\mu_1 + 2\mu_1 \frac{(\operatorname{tr}\{L\})^2 \operatorname{tr}\{L^2\}}{\operatorname{tr}\{L^4\}} \\ &\geq 2\mu_1 [(n \wedge p) + 3] \frac{(\operatorname{tr}\{L^2\})^2}{\operatorname{tr}\{L^4\}} - 8\mu_1. \end{aligned}$$

since $(\operatorname{tr}\{L\})^2 \geq \operatorname{tr}\{L^2\}$. Now, according to (A.5) with $k = 2$ and $m = 2$, we have

$$C_{a^*} \geq 2\mu_1 [(n \wedge p) - 1],$$

which gives the second result.

When $k = 3$, as $a^* = \frac{\mu_2}{\mu_1} [(n \vee p) + (n \wedge p) + 1]$, although $C_{a^*} > 0$, that quantity is not constant and cannot be bounded from below by a positive constant. Indeed we have

$$\begin{aligned} C_{a^*} &= 2 \frac{\mu_2}{\mu_1} [-2(n \wedge p) + 6] \frac{\operatorname{tr}\{L^3\} \operatorname{tr}\{L^2\}}{\operatorname{tr}\{L^5\}} - 8 \frac{\mu_2}{\mu_1} + 2 \frac{\mu_2}{\mu_1} B_5 \frac{\operatorname{tr}\{L^2\}}{\operatorname{tr}\{L^5\}} \\ &= 2 \frac{\mu_2}{\mu_1} [-2(n \wedge p) + 6] \frac{\operatorname{tr}\{L^3\} \operatorname{tr}\{L^2\}}{\operatorname{tr}\{L^5\}} - 8 \frac{\mu_2}{\mu_1} \\ &\quad + 2 \frac{\mu_2}{\mu_1} [2 \operatorname{tr}\{L\} \operatorname{tr}\{L^2\} + (2(n \wedge p) - 4) \operatorname{tr}\{L^3\}] \frac{\operatorname{tr}\{L^2\}}{\operatorname{tr}\{L^5\}} \\ &= 4 \frac{\mu_2}{\mu_1} \frac{\operatorname{tr}\{L^3\} \operatorname{tr}\{L^2\}}{\operatorname{tr}\{L^5\}} + 4 \frac{\mu_2}{\mu_1} \frac{\operatorname{tr}\{L\} (\operatorname{tr}\{L^2\})^2}{\operatorname{tr}\{L^5\}} - 8 \frac{\mu_2}{\mu_1}, \end{aligned} \tag{5.13}$$

by substituting the expression of B_5 given in Lemma A.2. As for the positivity of C_{a^*} , note that, according to Inequality (A.5) in Lemma (A.4) with $m = 2$ and $k = 1$ and 3 , we have

$$\frac{\operatorname{tr}\{L\} (\operatorname{tr}\{L^2\})^2}{\operatorname{tr}\{L^5\}} \geq \frac{\operatorname{tr}\{L^3\} \operatorname{tr}\{L^2\}}{\operatorname{tr}\{L^5\}} \geq 1.$$

Hence, it is clear that $C_{a^*} \geq 0$ with probability 1. However, from (5.13), for any $a > 0$, the corresponding C_a in (5.12) can be written as

$$C_a = C_{a^*} + 2 \frac{\mu_2}{\mu_1} (a^* - a) \frac{\text{tr}\{L^3\} \text{tr}\{L^2\}}{\text{tr}\{L^5\}},$$

Then, from (A.5) with $k = 3$ and $m = 3$, it is clear that

$$C_a \geq 2 \frac{\mu_2}{\mu_1} (a^* - a)$$

as soon as $a < a^*$. Then the result for $k = 3$ follows.

When $k = 0$, C_a in (5.12) equals

$$\begin{aligned} C_a &= 2 \left[\frac{\mu_{-1}}{\mu_{-2}} [(n \vee p) - (n \wedge p) + 1] - a \right] (n \wedge p) - 8 \frac{\mu_{-1}}{\mu_{-2}} + 2 \frac{\mu_{-1}}{\mu_{-2}} B_2 \\ &= 2 \left[\frac{\mu_{-1}}{\mu_{-2}} [(n \vee p) - (n \wedge p) + 1] - a \right] (n \wedge p) - 8 \frac{\mu_{-1}}{\mu_{-2}} + 2 \frac{\mu_{-1}}{\mu_{-2}} (n \wedge p) [(n \wedge p) - 1] \\ &= 2 \left[\frac{\mu_{-1}}{\mu_{-2}} (n \vee p) - a \right] (n \wedge p) - 4 \frac{\mu_{-1}}{\mu_{-2}}, \end{aligned} \quad (5.14)$$

where B_2 is given in Lemma A.2. For a equal to the value a_0^- in Proposition 3.1, that is, for $a_0^- = \frac{\mu_{-1}}{\mu_{-2}} [(n \vee p) - 2(n \wedge p) - 2]$, it follows from (5.14) that

$$C_{a_0^-} = 2 \frac{\mu_{-1}}{\mu_{-2}} [\{2(n \wedge p) + 2\} (n \wedge p) - 4] = 4 \frac{\mu_{-1}}{\mu_{-2}} [\{(n \wedge p) + 1\} (n \wedge p) - 2].$$

For a equal to the value a_0^+ in Proposition 3.1, that is, for $a_0^+ = \frac{\mu_{-1}}{\mu_{-2}} [(n \vee p) - (n \wedge p) - 3]$ it follows from (5.14) that

$$C_{a_0^+} = 2 \frac{\mu_{-1}}{\mu_{-2}} [\{(n \wedge p) + 3\} (n \wedge p) - 4].$$

Hence, according to Proposition 3.1, the result for $k = 0$ follows.

When $k \geq 4$, we have $a_0^- = \frac{\mu_{k-1}}{\mu_{k-2}} [(n \vee p) + (n \wedge p) + k - 2]$. Then, according to the

expression of B_k in Lemma A.2, it follows from C_a in (5.12) that

$$\begin{aligned}
C_{a_0^-} &= 2 \frac{\mu_{k-1}}{\mu_{k-2}} [-2(n \wedge p) + k + 3] \frac{\text{tr}\{L^k\} \text{tr}\{L^2\}}{\text{tr}\{L^{k+2}\}} - 8 \frac{\mu_{k-1}}{\mu_{k-2}} \\
&\quad + 2 \frac{\mu_{k-1}}{\mu_{k-2}} \left\{ \sum_{m=1}^{k-1} \text{tr}\{L^m\} \text{tr}\{L^{k-m}\} + [2(n \wedge p) - (k+1)] \text{tr}\{L^k\} \right\} \frac{\text{tr}\{L^2\}}{\text{tr}\{L^{k+2}\}} \\
&= 4 \frac{\mu_{k-1}}{\mu_{k-2}} \frac{\text{tr}\{L^k\} \text{tr}\{L^2\}}{\text{tr}\{L^{k+2}\}} - 8 \frac{\mu_{k-1}}{\mu_{k-2}} + 2 \frac{\mu_{k-1}}{\mu_{k-2}} \frac{\text{tr}\{L^2\}}{\text{tr}\{L^{k+2}\}} \sum_{m=1}^{k-1} \text{tr}\{L^m\} \text{tr}\{L^{k-m}\} \\
&\geq 4 \frac{\mu_{k-1}}{\mu_{k-2}} \frac{\text{tr}\{L^k\} \text{tr}\{L^2\}}{\text{tr}\{L^{k+2}\}} - 8 \frac{\mu_{k-1}}{\mu_{k-2}} + 2 \frac{\mu_{k-1}}{\mu_{k-2}} (k-1) \frac{\text{tr}\{L^k\} \text{tr}\{L^2\}}{\text{tr}\{L^{k+2}\}} \\
&= 2 \frac{\mu_{k-1}}{\mu_{k-2}} (k+1) \frac{\text{tr}\{L^k\} \text{tr}\{L^2\}}{\text{tr}\{L^{k+2}\}} - 8 \frac{\mu_{k-1}}{\mu_{k-2}} \\
&\geq 2 \frac{\mu_{k-1}}{\mu_{k-2}} (k-3),
\end{aligned}$$

applying twice Inequality (A.5) in Lemma A.4. This gives the result for $k \geq 4$. □

6 Concluding remarks

In this paper, we have considered estimation of the inverse scatter matrix Σ^{-1} of a scale mixture of Wishart distribution under Efron-Morris type losses. We have shown that the standard estimators of the form aS^+ can be improved by alternative estimators through an unbiased estimator of risk difference. Our approach unifies the two cases where the sample covariance matrix S is singular and invertible thanks to a new Stein-Haff type identity developed by Fourdrinier, Haddouche and Mezoued [5]. Our results extend several classical domination results for the Wishart case to the scale mixture of Wishart case, and also extend results for specific Efron and Morris type losses to the entire class.

Acknowledgments. The authors would like to thanks two anonymous referees for constructive comments and sharp corrections. D. Boukehil, D. Fourdrinier and F. Mezoued are partially supported by the franco-algerian program Tassili #18MDU105. D. Fourdrinier is also partially supported by the Russian program #1.472.2016/1.4 Ministry of Education and Science of the Russian Federation. W. E. Strawderman is supported by the Simons Foundation #...

Appendix

In this section, we focus on the technical aspects of the article. Thus we give the proofs of results presented in Sections 3 and 4 and some supplementary lemmas used in these proofs.

A.1 Proof of Proposition 3.1

Recall the singular value decomposition of S , i.e. $S = H_1 L H_1^\top$ (see Section 4). The proof relies on the following lemma from Fourdrinier, Haddouche and Mezoued [5].

Lemma A.1. *Let $G = H_1 \Psi H_1^\top$ where $\Psi = \Psi(L)$ is a $(n \wedge p) \times (n \wedge p)$ diagonal matrix, $\text{diag}(\psi_1(L), \dots, \psi_{n \wedge p}(L))$. Then $SS^+ G = G$ and*

$$\mathcal{D}_s\{SS^+ G\} = \mathcal{D}_s\{H_1 \Psi H_1^\top\} = H_1 \Psi^{(1)} H_1^\top + \frac{1}{2} \text{tr}(L^{-1} \Psi) (I_p - H_1 H_1^\top) \quad (\text{A.1})$$

where $\Psi^{(1)}$ is a $(n \wedge p) \times (n \wedge p)$ diagonal matrix with

$$\psi_i^{(1)} = \frac{1}{2}(p - (p \wedge n)) \frac{\psi_i}{l_i} + \frac{\partial \psi_i}{\partial l_i} + \frac{1}{2} \sum_{j \neq i}^{p \wedge n} \frac{\psi_i - \psi_j}{l_i - l_j}.$$

Recall that $\tilde{S} = S/V$. Without loss of generality, we prove Proposition 3.1 in the case where $V = 1$, that is, $\tilde{S} = S$. First, we have $\text{tr}\{\tilde{S}^k (\tilde{S}^+)^2\} = \text{tr}\{L^{k-2}\}$. Then, applying Lemma A.1 with $\Psi(L) = L^{k-1}$, we have

$$\begin{aligned} D_{\tilde{S}}\{\tilde{S}^k \tilde{S}^+\} &= D_{\tilde{S}}\{H_1 L^{k-1} H_1^\top\} \\ &= \frac{1}{2} [p - (n \wedge p) + 2(k-1)] H_1 L^{k-2} H_1^\top + \frac{1}{2} H_1 \text{diag} \left(\sum_{i \neq j}^{n \wedge p} \frac{l_i^{k-1} - l_j^{k-1}}{l_i - l_j} \right)_{1 \leq i \leq n \wedge p} H_1^\top \\ &\quad + \frac{1}{2} \text{tr}\{L^{k-2}\} (I_p - H_1 H_1^\top). \end{aligned} \quad (\text{A.2})$$

Substituting (A.2) in the numerator of the fraction in (3.4) and $\text{tr}\{L^{k-2}\}$ in the denominator and using the fact that $\tilde{S} \tilde{S}^+ (I_p - H_1 H_1^\top) = 0$, the optimal value of a^* in (3.4) is

$$\begin{aligned} a^* &= \frac{\mu_{k-1}}{\mu_{k-2}} [n - (n \wedge p) - 1 + [p - (n \wedge p) + 2(k-1)]] + \frac{\mu_{k-1}}{\mu_{k-2}} \frac{\mathbb{E}_\Sigma \left[\frac{\sum_{i=1}^{n \wedge p} \sum_{j \neq i}^{n \wedge p} \frac{l_i^{k-1} - l_j^{k-1}}{l_i - l_j}}{\text{tr}\{L^{k-2}\}} \right]}{\mathbb{E}_\Sigma [\text{tr}\{L^{k-2}\}]} \\ &= \frac{\mu_{k-1}}{\mu_{k-2}} [(n \vee p) - (n \wedge p) + 2k - 3] + \frac{\mu_{k-1}}{\mu_{k-2}} \frac{\mathbb{E}_\Sigma \left[\sum_{i=1}^{n \wedge p} \sum_{j \neq i}^{n \wedge p} \frac{l_i^{k-1} - l_j^{k-1}}{l_i - l_j} \right]}{\mathbb{E}_\Sigma [\text{tr}\{L^{k-2}\}]} \end{aligned} \quad (\text{A.3})$$

Finally, Proposition 3.1 follows immediately from the following lemma applied to (A.3).

Lemma A.2. For $k \in \mathbb{N}$, let

$$B_k = \sum_{i=1}^{n \wedge p} \sum_{j \neq i}^{n \wedge p} \frac{l_i^{k-1} - l_j^{k-1}}{l_i - l_j}.$$

Then, for $k = 0$,

$$B_0 = \operatorname{tr}\{L^{-2}\} - (\operatorname{tr}\{L^{-1}\})^2$$

and satisfies

$$[1 - (n \wedge p)] \operatorname{tr}\{L^{-2}\} \leq B_0 \leq 0;$$

for $k = 1$,

$$B_1 = 0;$$

for $k = 2$,

$$B_2 = (n \wedge p) [(n \wedge p) - 1];$$

for $k = 3$,

$$B_3 = 2[(n \wedge p) - 1] \operatorname{tr}\{L\};$$

for $k \geq 4$,

$$B_k = \sum_{m=1}^{k-3} \operatorname{tr}\{L^m\} \operatorname{tr}\{L^{k-2-m}\} + [2(n \wedge p) - (k-1)] \operatorname{tr}\{L^{k-2}\}$$

and satisfies

$$[2(n \wedge p) - (k-1)] \operatorname{tr}\{L^{k-2}\} \leq B_k \leq (k-1) [(n \wedge p) - 1] \operatorname{tr}\{L^{k-2}\}.$$

Proof. Note that the inequalities in Lemma A.2 use the inequality

$$\operatorname{tr}\{L^a\} \operatorname{tr}\{L^b\} \leq (n \wedge p) \operatorname{tr}\{L^{a+b}\}$$

for $a \in \mathbb{R}$ and $b \in \mathbb{R}$ with $ab \geq 0$, which can be established thanks to the covariance inequality. Thus, for $a = b = -1$, we have

$$(\operatorname{tr}\{L^{-1}\})^2 \leq (n \wedge p) \operatorname{tr}\{L^{-2}\},$$

which gives the first inequality satisfied by B_0 . As for the second inequality, it follows from Lemma A.4 with $k = m = -1$. Also, setting $a = m$ and $b = k - 2 - m$, we have

$$\operatorname{tr}\{L^m\} \operatorname{tr}\{L^{k-2-m}\} \leq (n \wedge p) \operatorname{tr}\{L^{k-2}\},$$

which gives rise to the second inequality satisfied by B_k for $k \geq 4$ (the first inequality being obvious).

As for the equalities, through straightforward calculations, we have

$$\begin{aligned} B_0 &= \sum_{i=1}^{n \wedge p} \sum_{j \neq i}^{n \wedge p} \frac{l_i^{-1} - l_j^{-1}}{l_i - l_j} = - \sum_{i=1}^{n \wedge p} \sum_{j \neq i}^{n \wedge p} \frac{1}{l_i l_j} = - \sum_{i=1}^{n \wedge p} \left(\sum_{j=1}^{n \wedge p} \frac{1}{l_i l_j} - \frac{1}{l_i^2} \right) \\ &= - \sum_{i=1}^{n \wedge p} \left(\frac{1}{l_i} \right) \sum_{j=1}^{n \wedge p} \left(\frac{1}{l_j} \right) + \sum_{i=1}^{n \wedge p} \left(\frac{1}{l_i^2} \right) = -(\text{tr}\{L^{-1}\})^2 + \text{tr}\{L^{-2}\}. \end{aligned}$$

Also

$$\begin{aligned} B_1 &= \sum_{i=1}^{n \wedge p} \sum_{j \neq i}^{n \wedge p} \frac{1 - 1}{l_i - l_j} = 0, \\ B_2 &= \sum_{i=1}^{n \wedge p} \sum_{j \neq i}^{n \wedge p} \frac{l_i - l_j}{l_i - l_j} = \sum_{i=1}^{n \wedge p} \sum_{j \neq i}^{n \wedge p} 1 = (n \wedge p) [(n \wedge p) - 1] \end{aligned}$$

and

$$\begin{aligned} B_3 &= \sum_{i=1}^{n \wedge p} \sum_{j \neq i}^{n \wedge p} \frac{l_i^2 - l_j^2}{l_i - l_j} = \sum_{i=1}^{n \wedge p} \sum_{j \neq i}^{n \wedge p} (l_i + l_j) = [(n \wedge p) - 1] \sum_{i=1}^{n \wedge p} l_i + \sum_{i=1}^{n \wedge p} \left(\sum_{j=1}^{n \wedge p} l_j - l_i \right) \\ &= [(n \wedge p) - 1] \text{tr}\{L\} + (n \wedge p) \text{tr}\{L\} - \text{tr}\{L\} = 2[(n \wedge p) - 1] \text{tr}\{L\}. \end{aligned}$$

Finally, for $k \geq 4$,

$$\begin{aligned} B_k &= \sum_{i=1}^{n \wedge p} \sum_{j \neq i}^{n \wedge p} \frac{l_i^{k-1} - l_j^{k-1}}{l_i - l_j} = \sum_{i=1}^{n \wedge p} \sum_{j \neq i}^{n \wedge p} \sum_{m=0}^{k-2} l_i^m l_j^{k-2-m} \\ &= \sum_{m=1}^{k-3} \sum_{i=1}^{n \wedge p} \sum_{j \neq i}^{n \wedge p} l_i^m l_j^{k-2-m} + \sum_{i=1}^{n \wedge p} \sum_{j \neq i}^{n \wedge p} (l_j^{k-2} + l_i^{k-2}) \\ &= \sum_{m=1}^{k-3} \sum_{i=1}^{n \wedge p} l_i^m \left(\sum_{j=1}^{n \wedge p} l_j^{k-2-m} - l_i^{k-2-m} \right) + \sum_{i=1}^{n \wedge p} \left(\sum_{j=1}^{n \wedge p} l_j^{k-2} - l_i^{k-2} \right) + [(n \wedge p) - 1] \sum_{i=1}^{n \wedge p} l_i^{k-2} \\ &= \sum_{m=1}^{k-3} (\text{tr}\{L^m\} \text{tr}\{L^{k-2-m}\} - \text{tr}\{L^{k-2}\}) + 2[(n \wedge p) - 1] \text{tr}\{L^{k-2}\} \\ &= \sum_{m=1}^{k-3} \text{tr}\{L^m\} \text{tr}\{L^{k-2-m}\} + [2(n \wedge p) - (k-1)] \text{tr}\{L^{k-2}\}. \end{aligned}$$

□

A.2 An expression for $\text{tr}\{\tilde{S}^+ \tilde{S} D_{\tilde{S}}\{r(\text{tr}\{\tilde{S}\}) G^\top(\tilde{S}) \tilde{S}^{k+1}\}\}$

The following lemma was used in the proof of Lemma 4.2.

Lemma A.3. For $G(\tilde{S})$ of the form $G(\tilde{S}) = H_1 \Psi(L) H_1^\top$, we have

$$\begin{aligned} \text{tr}\{\tilde{S}^+ \tilde{S} D_{\tilde{S}}\{r(\text{tr}\{\tilde{S}\}) G^\top(\tilde{S}) \tilde{S}^{k+1}\}\} &= r(\text{tr}\{\tilde{S}\}) \text{tr}\{\tilde{S}^+ \tilde{S} D_{\tilde{S}}\{G^\top(\tilde{S}) \tilde{S}^{k+1}\}\} \\ &\quad + r'(\text{tr}\{\tilde{S}\}) \text{tr}\{G^\top(\tilde{S}) \tilde{S}^{k+1}\} \end{aligned} \quad (\text{A.4})$$

Proof. Since $\tilde{S} \tilde{S}^+ G^\top(\tilde{S}) = G^\top(\tilde{S})$, we have

$$\begin{aligned} D_{\tilde{S}}\{r(\text{tr}\{\tilde{S}\}) G^\top(\tilde{S}) \tilde{S}^{k+1}\} &= D_{\tilde{S}}\{\tilde{S}^+ \tilde{S} r(\text{tr}\{\tilde{S}\}) G^\top(\tilde{S}) \tilde{S}^{k+1}\} \\ &= D_{\tilde{S}}\{H_1 r(\text{tr}\{\tilde{S}\}) \Psi L^{k+1} H_1^\top\} \\ &= H_1 \Psi^{(1)} H_1^\top + \frac{1}{2} \text{tr}\left(L^{-1} r(\text{tr}\{\tilde{S}\}) \Psi L^{k+1}\right) (I_p - H_1 H_1^\top), \end{aligned}$$

according to Lemma A.1, where

$$\begin{aligned} \psi_i^{(1)} &= \frac{1}{2} (p - (p \wedge n)) r(\text{tr}\{L\}) \frac{(\psi L^{k+1})_i}{l_i} + \frac{\partial (r(\text{tr}\{L\}) (\psi L^{k+1})_i)}{\partial l_i} \\ &\quad + \frac{1}{2} r(\text{tr}\{L\}) \sum_{j \neq i}^{p \wedge n} \frac{(\psi L^{k+1})_i - (\psi L^{k+1})_j}{l_i - l_j} \\ &= r(\text{tr}\{L\}) \left(\frac{1}{2} (p - (p \wedge n)) \frac{(\psi L^{k+1})_i}{l_i} + \frac{\partial (\psi L^{k+1})_i}{\partial l_i} + \frac{1}{2} \sum_{j \neq i}^{p \wedge n} \frac{(\psi L^{k+1})_i - (\psi L^{k+1})_j}{l_i - l_j} \right) \\ &\quad + r'(\text{tr}\{L\}) (\psi L^{k+1})_i. \end{aligned}$$

Noticing that $\tilde{S} \tilde{S}^+ (I_p - H_1 H_1^\top) = 0$, it follows that

$$\tilde{S} \tilde{S}^+ D_{\tilde{S}}\{r(\text{tr}\{\tilde{S}\}) G^\top(\tilde{S}) \tilde{S}^{k+1}\} = r(\text{tr}\{\tilde{S}\}) \tilde{S} \tilde{S}^+ D_{\tilde{S}}\{G^\top(\tilde{S}) \tilde{S}^{k+1}\} + r'(\text{tr}\{\tilde{S}\}) \tilde{S} \tilde{S}^+ \Psi L^{k+1}.$$

Finally, taking the trace gives the desired result. \square

A.3 A trace inequality

Lemma A.4. For $k \in \mathbb{R}$ and $m \in \mathbb{R}$, we have

$$\text{tr}\{L^{k+m}\} \leq \text{tr}\{L^k\} \text{tr}\{L^m\}. \quad (\text{A.5})$$

Proof. The result follows immediately from the inequality

$$\sum_{i=1}^{p \wedge n} l_i^{k+m} \leq \sum_{i=1}^{p \wedge n} l_i^k \sum_{j=1}^{p \wedge n} l_j^m.$$

\square

References

- [1] D. Chételat and M. T. Wells. Improved multivariate normal mean estimation with unknown covariance when p is greater than n . *Annals of Statistics*, 40:3137–3160, 2012.
- [2] D. K. Dey. Improved estimation of a multinormal precision matrix. *Statistics & Probability Letters*, 6:125–128, 1987.
- [3] D. K. Dey, M. Ghosh, and C. Srinivasan. A new class of improved estimators of a multinormal precision matrix. *Statistics & Risk Modeling*, 8:141–152, 1990.
- [4] B. Efron and C. Morris. Multivariate empirical bayes and estimation of covariance matrices. *Annals of Statistics*, 4:22–32, 1976.
- [5] D. Fourdrinier, A. M. Haddouche, and F. Mezoued. Covariance matrix estimation of an elliptically symmetric distribution. *Technical report, Université de Rouen and ENSSEA de Tipaza*, 2019.
- [6] D. Fourdrinier, F. Mezoued, and M.T. Wells. Estimation of the inverse scatter matrix of an elliptically symmetric distribution. *Journal of Multivariate Analysis*, 143:32–55, 2016.
- [7] L. R. Haff. Minimax estimators for a multinormal precision matrix. *Journal of Multivariate Analysis*, 7:374–385, 1977.
- [8] L. R. Haff. Estimation of the inverse covariance matrix: random mixtures of the inverse Wishart matrix and the identity. *Annals of Statistics*, 7(6):1264–1276, 1979.
- [9] L. R. Haff. An identity for the Wishart distribution with applications. *Journal of Multivariate Analysis*, 9:531–544, 1979.
- [10] L. R. Haff, Kim P. T., Koo J.-Y., and Richards D. St. P. Minimax estimation for mixtures of Wishart distributions. *The Annals of Statistics*, 39(6):3417–3440, 2011.
- [11] B. Jian, B.C. Vemuri, and E. Ozaeslan. A mixture of wisharts (mow) model for multifiber reconstruction. In *Visualisation and Processing of Tensor Fields*, pages 39–56, 2009.
- [12] Y. Konno. Shrinkage estimators for large covariance matrices in multivariate real and complex normal distributions under an invariant quadratic loss. *Journal of Multivariate Analysis*, 100:2237–2253, 2009.
- [13] K. Krishnamoorthy and A. K. Gupta. Improved minimax estimation of a normal precision matrix. *Canadian Journal of Statistics*, 17:91–102, 1989.

- [14] T. Kubokawa and M. S. Srivastava. Estimation of the precision matrix of a singular Wishart distribution and its application in high-dimensional data. *Journal of Multivariate Analysis*, 99:1906–1928, 2008.
- [15] S. F. V. Nielsen, K. H. Madsen, M. N. Schmidt, and M. Mørup. Region-based change detection for polarimetric SAR images using Wishart mixture models. In *Proceeding of the 2017 International Workshop on Pattern Recognition in Neuroimaging*, 2016.
- [16] C. Stein. Estimation of a covariance matrix. In *Rietz Lecture, 39th Annual Meeting of the IMS*. Atlanta, Georgia, 1975.
- [17] H. Tsukuma and Y. Konno. On improved estimation of normal precision and discriminant coefficients. *Journal of Multivariate Analysis*, 97(1):1477–1500, 2006.
- [18] W. Yang, X. Yang, T. Yan, H. Song, and G.S Xia. Region-based change detection for polarimetric sar images using wishart mixture models. In *IEEE Transactions on Geoscience and Remote Sensing*, 2016.
- [19] X. Zhou, X. Sun, and J. Wang. Estimation of the multivariate normal precision matrix under the entropy loss. *Annals of the Institute of Statistical Mathematics*, 53(4):760–768, 2001.