Journal of Multivariate Analysis 181 (2021) 104680

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Scale matrix estimation of an elliptically symmetric distribution in high and low dimensions

Anis M. Haddouche [a,b,1], Dominique Fourdrinier [b,2,*], Fatiha Mezoued [a,2]

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

ARTICLE INFO

Article history:
Received 11 August 2019
Received in revised form 15 September 2020
Accepted 15 September 2020
Available online 23 September 2020

AMS 2010 subject classifications:
62H12
62F10
62C99

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

ABSTRACT

The problem of estimating the scale matrix Σ in a multivariate additive model, with elliptical noise, is considered from a decision-theoretic point of view. As the natural estimators of the form $\hat{\Sigma}_a = a\,S$ (where S is the sample covariance matrix and a is a positive constant) perform poorly, we propose estimators of the general form $\hat{\Sigma}_{a,G} = a\left(S + SS^{+}G(Z,S)\right)$, where S^{+} is the Moore–Penrose inverse of S and $G(Z,S)$ is a correction matrix. We provide conditions on $G(Z,S)$ such that $\hat{\Sigma}_{a,G}$ improves over $\hat{\Sigma}_a$ under the quadratic loss $L(\Sigma, \hat{\Sigma}) = \operatorname{tr}\left(\hat{\Sigma}\,\Sigma^{-1} - I_p\right)^2$. We adopt a unified approach to the two cases where S is invertible and S is singular. To this end, a new Stein–Haff type identity and calculus on eigenstructure for S are developed. Our theory is illustrated with a large class of estimators which are orthogonally invariant.

1. Introduction

Let Y be an observed $n \times p$ matrix under the multivariate additive model

$$Y = M + \mathcal{E}, \qquad \mathcal{E} \sim ES(0, \Sigma, I_n), \tag{1}$$

where M denotes the unknown $n \times p$ matrix of parameters, $\mathcal{E}$ is an $n \times p$ elliptically symmetric distributed noise with unknown covariance matrix proportional to $I_n \otimes \Sigma$ (see (9)). Here, Σ is a $p \times p$ invertible scale matrix and I_n is the n-dimensional identity matrix. The class of elliptically symmetric distributions encompasses a large number of important distributions such as the Gaussian, Cauchy, exponential, Student-t distributions and the Weibull distribution. Our main assumption is that the column space of M is known (or can be approximated by a known linear subspace). Also its known rank q satisfies

$$1 \le \operatorname{rank}(M) = q < p \wedge n\,, \tag{2}$$

where $p \wedge n = \min\{p, n\}$.

* Corresponding author.
E-mail address: Dominique.Fourdrinier@univ-rouen.fr (D. Fourdrinier).

1 Ph.D. Student.
2 Professor.

https://doi.org/10.1016/j.jmva.2020.104680
0047-259X/

This model has been addressed by Candès et al. [3] who assumed that, even if M is unknown, its rank is known or can be approximated by a low-rank matrix. See also Candès and Recht [2], Ji et al. [21] and Nguyen et al. [26] in an applied setting. Note that Canu and Fourdrinier [4] extended to an elliptical context the Gaussian approach adopted by Candès et al. [3]. While these authors were interested in estimating the mean matrix M, here, the parameter of interest is the scale matrix Σ, which coincides with the covariance matrix in the Gaussian setting.

Thanks to the low-rank assumption (2), the additive model (1) can be presented in a canonical form (Z, U) which separates information about the mean structure (a $q \times p$ matrix Z) and the information concerning the scale (a $p \times p$ matrix $S = U^\top U$, where U is a $(n-q) \times p$ matrix). In this canonical context (see Section 2.1), the usual estimators are represented by

$$\hat{\Sigma}_a = a\,S\,, \tag{3}$$

where S is the sample covariance matrix and a is a positive constant. As these natural estimators perform poorly, we consider alternative estimators under the quadratic loss function

$$L(\Sigma, \hat{\Sigma}) = \mathrm{tr}\big(\hat{\Sigma}\,\Sigma^{-1} - I_p\big)^2, \tag{4}$$

where $\hat{\Sigma}$ is an estimator of Σ and $\mathrm{tr}(A)$ denotes the trace of the matrix A.

In the literature, two cases are distinguished: the case where S is invertible ($p \le n-q$) and the case where S is singular ($p > n-q$). As pointed out by James and Stein [20], the estimators in (3) are inadmissible under a normal multivariate distribution, which naturally remains true in the general elliptical setting. Since then, numerous authors have suggested improved estimators which improve over the unbiased estimator $S/(n-q-1)$ and over the maximum likelihood estimator $S/(n-q)$. In the invertible case ($p \le n-q$), referred here as the low-dimensional setting, the literature includes, Stein [28,29], Efron and Morris [8], Haff [15], Takemura [30], Dey and Srinavasan [7], Sheena [27]. Also, Kubokawa and Srivastava [24] showed that, under the Stein loss function $\mathrm{tr}(\hat{\Sigma}\,\Sigma) - \log|\hat{\Sigma}\,\Sigma| - p$, the improved estimators proposed in the normal setting remain robust, in the sense that they are still improved estimators within the elliptical distributions class. However, to our knowledge, a similar extension under the quadratic loss (4), which is more difficult to handle than the Stein loss, has not yet been obtained. More recently, with the massive amount of high throughput data, much interest has turned to the non-invertible case ($p > n-q$), often referred to the high-dimensional setting. Many authors proposed improved estimators in the Gaussian setting such as, Chen et al. [5], Ikeda et al. [19], Tsukuma and Kubokawa [32], Tsukuma [31] and Konno [22], who extended the result due to Haff [15], in the invertible case, to the high-dimensional setting and proposed improved estimators under the quadratic loss (4). However, as in the invertible case ($p \le n-q$), no results have been yet established in the general elliptical framework under this loss.

Our main objective is the derivation of dominance results for alternative estimators of the form

$$\hat{\Sigma}_{a,G} = a\left(S + SS^{+}G(Z,S)\right), \tag{5}$$

over the usual estimators $\hat{\Sigma}_a = a\,S$ where a is a positive constant and $G(Z,S)$ is a $p \times p$ matrix function. The two main features of our approach is that we treat the general elliptically symmetric distributions context and we unify the two cases where S is singular and S is invertible. For that purpose, we denote by S^{+} the Moore–Penrose inverse when S is singular, and the regular inverse S^{-1} when it is invertible.

The remainder of this paper is organized as follows. The primary decision-theoretic results are presented in Section 2. We give sufficient conditions on the correction matrix function $G(Z,S)$ for which $\hat{\Sigma}_{a,G}$ improves on $\hat{\Sigma}_a$. To do this, we derive a new version of the so-called Stein–Haff identity for this setting, which is the basis on which the development of improved estimators depends. In Section 3, based on a new calculus on eigenstructure of S, we apply the results of Section 2 to the class of orthogonally invariant estimators. We also extend the estimator due to Haff [17] and Konno [22], in the low-dimensional setting and under the Gaussian assumption, to the class of elliptical symmetric distributions. In Section 4 examples illustrate the theory. In Section 5 we investigate the amount of improvement provided by the Konno estimator (see [22]) through numerical study. Finally, an Appendix contains technical results and the proofs of some of the findings in this paper.

2. Improved estimators

2.1. The canonical form of the model (1)

As mentioned in Section 1, we deal with the additive model (1) through its canonical form. To this end, we follow the lines of Canu and Fourdrinier [4]. For more details on the canonical model, see Fourdrinier et al. [13].

Thanks to the low-rank assumption (2), there exists Q_2, a semi-orthogonal $n \times m$ matrix, where $m = n-q$, such that

$$Q_2^\top M = 0_{m\times p}\,. \tag{6}$$

Complete Q_2 with Q_1 to form an $n \times n$ orthogonal matrix $Q = (Q_1 Q_2)$, so that we can write

$$Q^\top Y = \begin{pmatrix} Z \\ U \end{pmatrix} = \begin{pmatrix} Q_1^\top \\ Q_2^\top \end{pmatrix} Y = \begin{pmatrix} Q_1^\top \\ Q_2^\top \end{pmatrix} M + Q^\top \mathcal{E} = \begin{pmatrix} \theta \\ 0 \end{pmatrix} + Q^\top \mathcal{E},$$

with $\theta = Q_1^\top M$, thanks to (6). The canonical form of the additive model (1) is

$$\begin{pmatrix} Z \\ U \end{pmatrix} = \begin{pmatrix} \theta \\ 0 \end{pmatrix} + Q^\top \mathcal{E}, \tag{7}$$

where Z and U are respectively $q \times p$ and $m \times p$ matrices. In the Gaussian setting, this canonical form was recently considered by Tsukuma and Kubokawa [32] in order to estimate the covariance matrix. Now, if we assume that $\mathcal{E} = Y - M$ has a density with respect to the Lebesgue measure in $\mathbb{R}^{pn}$, it is necessarily of the form

$$\varepsilon \mapsto |\Sigma|^{-n/2} f\{\mathrm{tr}\left(\varepsilon\, \Sigma^{-1} \varepsilon^\top\right)\}, \tag{8}$$

for some function f (called the generating function). Note that it can be shown that the covariance of $\mathcal{E}$ equals

$$\frac{1}{p\,n} E\left[\mathrm{tr}\left(\mathcal{E}\, \Sigma^{-1}\, \mathcal{E}^\top\right)\right] I_n \otimes \Sigma\,, \tag{9}$$

where E denotes the expectation with respect to the density in (8) (see for instance Fang and Zang [9]). Note also that $Q^\top \mathcal{E}$ has density

$$\mathcal{E} \mapsto |\Sigma|^{-n/2} f\{\mathrm{tr}\left(Q^\top\, \mathcal{E}\, \Sigma^{-1}\, \mathcal{E}^\top\, Q\right)\} = |\Sigma|^{-n/2} f\{\mathrm{tr}\left(\mathcal{E}\, \Sigma^{-1}\, \mathcal{E}^\top\right)\}, \tag{10}$$

by orthogonality of Q. Thus the distribution of the noise is invariant under this orthogonal transformation. It follows that $(Z^\top U^\top)^\top = Q^\top Y$ has an elliptically symmetric distribution around the matrix $(\theta^\top 0^\top)^\top$. Hence, from (7) and (10), the joint density of Z and U is

$$(z, u) \mapsto |\Sigma|^{-n/2} f\left[\mathrm{tr}\left\{\begin{pmatrix} z-\theta \\ u \end{pmatrix} \Sigma^{-1} \begin{pmatrix} z-\theta \\ u \end{pmatrix}^\top\right\}\right] = |\Sigma|^{-n/2} f\left[\mathrm{tr}\{(z-\theta)\Sigma^{-1}(z-\theta)^\top\} + \mathrm{tr}\{\Sigma^{-1} u^\top u\}\right]. \tag{11}$$

As no unbiased estimator of the risk difference between $\hat{\Sigma}_{a,G}$ in (5) and $\hat{\Sigma}_a$ in (3) is available, we need to reduce the class of density (11). The development essentially incorporates an integration by parts involving densities obtained from the original density (11) via a one dimensional integral of the tail of the density in (11); it is related to a technique developed by Berger [1] in the spherically symmetric case. Thus, we define the densities

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

and

$$(z, u) \mapsto \frac{1}{K^{**}} |\Sigma|^{-n/2} F^{**}\left[\mathrm{tr}\{(z-\theta)\, \Sigma^{-1}\, (z-\theta)^\top\} + \mathrm{tr}\{\Sigma^{-1}\, u^\top u\}\right], \tag{13}$$

where, for all $t \in \mathbb{R}^+$,

$$F^*(t) = \frac{1}{2} \int_t^\infty f(v)\, dv \quad \text{and} \quad F^{**}(t) = \frac{1}{2} \int_t^\infty F^*(v)\, dv\,. \tag{14}$$

Here, the normalizing constants

$$K^* = \int_{\mathbb{R}^{pn}} |\Sigma|^{-n/2} F^*\left[\mathrm{tr}\{(z-\theta)\, \Sigma^{-1}\, (z-\theta)^\top\} + \mathrm{tr}\{\Sigma^{-1}\, u^\top u\}\right] dz\, du \tag{15}$$

and

$$K^{**} = \int_{\mathbb{R}^{pn}} |\Sigma|^{-n/2} F^{**}\left[\mathrm{tr}\{(z-\theta)\, \Sigma^{-1}\, (z-\theta)^\top\} + \mathrm{tr}\{\Sigma^{-1}\, u^\top u\}\right] dz\, du\,, \tag{16}$$

are assumed to be finite.

2.2. The main result

Instead of the reference estimators $\hat{\Sigma}_a = a\,S$ of Σ, we consider alternative estimators of Σ of the form $\hat{\Sigma}_{a,G} = a\left(S + SS^+G(Z,S)\right)$, where $a > 0$, S^+ denotes the Moore–Penrose inverse of S and $G(Z,S)$ is a $p \times p$ matrix function such that the correction factor $SS^+G(Z,S)$ is symmetric. The performance of estimators is evaluated through the risk function

$$R(\Sigma, \hat{\Sigma}) = E_{\theta,\Sigma}[L(\Sigma, \hat{\Sigma})]\,, \tag{17}$$

where $E_{\theta,\Sigma}$ denotes the expectation with respect to the density in (11) and $L(\Sigma, \hat{\Sigma})$ is given in (4).

As mentioned in Section 2.1, we reduce the distributional context in (11) to the subclass of densities such that, for all $t \in \mathbb{R}^+$,

$$c \le \frac{F^*(t)}{f(t)} \le b \tag{18}$$

where c and b are positive constants. These densities in (18) have been considered by Fourdrinier, Mezoued and Strawderman [11]. They contain the multivariate normal distribution and the variance mixture of normals given in Appendix A.1 (see [11] for further examples).

There exists $a_o > 0$ such that $\hat{\Sigma}_{a_o} = a_o S$ is optimal among the class of usual estimators $\hat{\Sigma}_a = a S$ (that is, the risk of $\hat{\Sigma}_{a_o}$ is less than or equal to the risk of $\hat{\Sigma}_a$ for any $a > 0$); this is

$$a_o = \frac{1}{K^{**}(p+m+1)}\,, \tag{19}$$

where K^{**} is given in (16). See Appendix A.3 for a proof. The improvement over the class of $a\,S$'s will be shown through the improvement of $\hat{\Sigma}_{a_o,G} = a_o S + a_o SS^{+} G(Z,S)$. We provide in the following theorem, which is the main result of this paper, improvement conditions of $a_o\big(S + SS^{+}G(Z,S)\big)$ over $a_o S$. See Appendix A.3 for a proof. Note that, we will use the Haff operator $\mathcal{D}_s\{\cdot\}$ whose generic element is

$$d^s_{ij} = \frac{1}{2}(1+\delta_{ij})\frac{\partial}{\partial S_{ij}}, \tag{20}$$

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

Theorem 2.1. *Consider a density as in (11) satisfying (18). Under the condition*

$$\operatorname{tr}\left[2\,SS^{+}\,\mathcal{D}_s\{SS^{+}\,G(Z,S)\} + (m-(p\wedge m)-1)\,S^{+}\,G(Z,S)\right] \geq 0\,, \tag{21}$$

the estimator $\hat{\Sigma}_{a_o,G} = a_o\left(S + SS^{+}G(Z,S)\right)$ *improves over* $\hat{\Sigma}_{a_o} = a_o S$ *if*

$$\operatorname{tr}\Bigg[2\,S^{+}S\,\mathcal{D}_s\{SS^{+}\,T^{*}\}^{\top} + (m-p\wedge m-1)\,S^{+}\,T^{*} \\ -2\,(p+m+1)\,\frac{c^2}{b^2}\,\left(2\,SS^{+}\,\mathcal{D}_s\{SS^{+}\,G(Z,S)\} + (m-(p\wedge m)-1)\,S^{+}\,G(Z,S)\right)\Bigg] \leq 0\,, \tag{22}$$

where

$$T^{*} = 4\,(S + SS^{+}\,G(Z,S))\,\mathcal{D}_s\{SS^{+}\,G(Z,S)\} + G\left(2\,m\,SS^{+} - (p-m+1)\,S^{+}\,G(Z,S)\right). \tag{23}$$

A consequence of the unified approach of the two cases, S invertible and S non-invertible, is that the improvement conditions of the alternative estimators $\hat{\Sigma}_{a,G}$ in (5) over $\hat{\Sigma}_a$ in (3) are conditions on $G(Z,S)$, which are similar in these two settings. This feature is illustrated by the fact that $\hat{\Sigma}_{a,G}$ can be written as

$$\hat{\Sigma}_{a,G} = SS^{+}\,a\left(S + G(Z,S)\right),$$

and hence, is the projection of $a\,(S + G(Z,S))$ onto the column space of S. It is worth noting the parallel with the situation considered by Chételat and Wells [6] who, estimating θ, underline the fact that their improved estimators apply shrinkage only on the component of Z in the subspace spanned by the column of S.

2.3. A new Stein–Haff type identity

Theorem 2.1 expresses conditions on $G := G(Z,S)$ in order that $\hat{\Sigma}_{a,G}$ improves on $\hat{\Sigma}_a$ under the quadratic risk (17), that is, such that the risk difference

$$\Delta(G) = R\left(\Sigma, a\left(S + SS^{+}\,G\right)\right) - R\left(\Sigma\,, a\,S\right) \tag{24}$$

is non-positive. We give in the following proposition conditions to ensure the finiteness of the risks difference (24). The proof is deferred to Appendix A.3.

Proposition 2.1. *Let* $\|A\|_F = \sqrt{\operatorname{tr}(A^{\top} A)}$ *be the Frobenius norm. Assume that the expectations* $E_{\theta,\Sigma}\left[\|\Sigma^{-1}S\|_F^2\right]$ *and* $E_{\theta,\Sigma}\left[\|\Sigma^{-1}SS^{+}G\|_F^2\right]$ *are finite. Then, for any* $a > 0$*, the risks of* $\hat{\Sigma}_a$ *and* $\hat{\Sigma}_{a,G}$ *are finite. In that case, the risk difference in (24) is also finite and can be written as*

$$\Delta(G) = a^2\,E_{\theta,\Sigma}\left[\operatorname{tr}\left(\Sigma^{-1}SS^{+}(2\,S+G)\,\Sigma^{-1}SS^{+}\,G\right)\right] - 2\,a\,E_{\theta,\Sigma}\left[\operatorname{tr}\left(\Sigma^{-1}\,SS^{+}\,G\right)\right]\,. \tag{25}$$

The dependence in (25) on the unknown parameter Σ^{-1} is problematic since it intervenes in the integrand terms. As a remedy, we provide in the following lemma a new version of the so-called Stein–Haff identity in the framework of elliptically symmetric distributions, which unifies the cases where S is singular and S is invertible. For this purpose, we define $E^{*}_{\theta,\Sigma}$ as the expectation with respect to the density (12).

Lemma 2.1. *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 $E_{\theta,\Sigma}\left[\left|\mathrm{tr}\left(\Sigma^{-1}S\,S^{+}\,G\right)\right|\right]<\infty$. Then we have*

$$E_{\theta,\Sigma}\left[\mathrm{tr}\left(\Sigma^{-1}\,SS^{+}\,G\right)\right]=K^{*}\,E^{*}_{\theta,\Sigma}\left[\mathrm{tr}\left(2\,SS^{+}\,\mathcal{D}_s\{SS^{+}G\}^{\top}+(m-(p\wedge m)-1)\,S^{+}\,G\right)\right]. \tag{26}$$

See Appendix A.3 for the proof. Note that, in the case where S is invertible ($p\le m$), since $S^{+}=S^{-1}$, Identity (26) corresponds to

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

which is the identity given by Kubokawa and Srivastava [24]. In the singular case, Identity (26) becomes

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

which is, to our knowledge, a new Stein–Haff type identity.

Although Lemma 2.1 allows to get rid of Σ^{-1} in the second term on the right-hand side of (25), thanks to the $E^{*}_{\theta,\Sigma}$-expectation, note that it appears twice in the first term. We can deal with this by applying Lemma 2.1 twice, which gives rise to the following corollary through the $E^{**}_{\theta,\Sigma}$-expectation with respect to the density (13).

Corollary 2.1. *Let $G(z,s)$ and $V(z,s)$ be $p\times p$ matrices function such that, for any fixed z, $G(z,s)$ and $V(z,s)$ are weakly differentiable with respect to s. With $V:=V(Z,S)$, assume that $SS^{+}V$ is symmetric and such that $E_{\theta,\Sigma}\left[\left|\mathrm{tr}\left(\Sigma^{-1}SS^{+}V\,\Sigma^{-1}SS^{+}G\right)\right|\right]<\infty$. Assume also that $E^{*}_{\theta,\Sigma}\left[\left|\mathrm{tr}\left(\Sigma^{-1}SS^{+}T^{*}\right)\right|\right]<\infty$. Then we have*

$$E_{\theta,\Sigma}\left[\mathrm{tr}\left(\Sigma^{-1}SS^{+}V\,\Sigma^{-1}SS^{+}G\right)\right]=K^{*}K^{**}E^{**}_{\theta,\Sigma}\left[\mathrm{tr}\left(2SS^{+}\mathcal{D}_s\{SS^{+}T^{*}\}^{\top}+(m-(p\wedge m)-1)S^{+}T^{*}\right)\right] \tag{27}$$

with

$$T^{*}=2\left[SS^{+}\,V\,\mathcal{D}_s\{SS^{+}\,G\}^{\top}+SS^{+}\,G\,\mathcal{D}_s\{SS^{+}\,V\}\right]-(p-m+1)\,G\,S^{+}\,V.$$

See Appendix A.3 for the proof. Identity (27) parallels (26) as the role of the density (11) (respectively, the role of the expectation $E_{\theta,\Sigma}$) is played by the density (12) (respectively, the role of the expectation $E^{*}_{\theta,\Sigma}$) so that we have, for any $p\times p$ matrix $H(z,s)$ weakly differentiable with respect to s such that $E^{*}_{\theta,\Sigma}\left[\left|\mathrm{tr}\left(\Sigma^{-1}SS^{+}\,H\right)\right|\right]<\infty$,

$$E^{*}_{\theta,\Sigma}\left[\mathrm{tr}\left(\Sigma^{-1}SS^{+}H\right)\right]=K^{**}E^{**}_{\theta,\Sigma}\left[\mathrm{tr}\left(2\,S^{+}S\,\mathcal{D}_s\{SS^{+}H\}^{\top}+(m-(p\wedge m)-1)S^{+}H\right)\right].$$

Now we can give an expression of the risk difference $\Delta(G)$ in (25) which does not involve the unknown parameter Σ^{-1} in the integrand term and which relies on Lemma 2.1 and Corollary 2.1. It is worth noticing that the conditions we will use to this end imply those expressed in Lemma 2.1 and Corollary 2.1.

Proposition 2.2. *Let $G(z,s)$ be $p\times p$ matrices function such that, for any fixed z, $G(z,s)$ is weakly differentiable with respect to s. Assume that the correction factor $SS^{+}G$ is symmetric. Assume also that $E_{\theta,\Sigma}\left[\|\Sigma^{-1}\,S\|_F^2\right]<\infty$, $E_{\theta,\Sigma}\left[\|\Sigma^{-1}SS^{+}G\|_F^2\right]<\infty$ and $E^{*}_{\theta,\Sigma}\left[\left|\mathrm{tr}\left(\Sigma^{-1}SS^{+}T^{*}\right)\right|\right]<\infty$. Then*

$$\begin{aligned}\Delta(G)=a^2\,K^{*}\,K^{**}\,E^{**}_{\theta,\Sigma}&\left[\mathrm{tr}\left(2\,SS^{+}\,\mathcal{D}_s\{SS^{+}\,T^{*}\}^{\top}+(m-(p\wedge m)-1)\,S^{+}\,T^{*}\right)\right]\\&-2\,a\,K^{*}\,E^{*}_{\theta,\Sigma}\left[\mathrm{tr}\left(2\,S^{+}S\,\mathcal{D}_s\{SS^{+}\,G\}+(m-(p\wedge m)-1)\,S^{+}G\right)\right]\end{aligned} \tag{28}$$

where

$$T^{*}=4\,(S+SS^{+}\,G)\,\mathcal{D}_s\{SS^{+}\,G\}+G\left(2\,m\,SS^{+}-(p-m+1)\,S^{+}\,G\right). \tag{29}$$

Proof. Recall that the conditions $E_{\theta,\Sigma}\left[\|\Sigma^{-1}\,S\|_F^2\right]<\infty$ and $E_{\theta,\Sigma}\left[\|\Sigma^{-1}SS^{+}G\|_F^2\right]<\infty$ guarantee the finiteness of the risk difference of $\hat{\Sigma}_a$ and $\hat{\Sigma}_{a,G}$. Note that the second finiteness risk condition implies the condition in Lemma 2.1 $E_{\theta,\Sigma}\left[\left|\mathrm{tr}\left(\Sigma^{-1}S\,S^{+}\,G\right)\right|\right]<\infty$ since, by the Cauchy–Schwarz inequality

$$E_{\theta,\Sigma}\left[\left|\mathrm{tr}\left(\Sigma^{-1}S\,S^{+}\,G\right)\right|\right]=E_{\theta,\Sigma}\left[\left|\mathrm{tr}\left(I_p\,\Sigma^{-1}S\,S^{+}\,G\right)\right|\right]\le p^{1/2}\,E_{\theta,\Sigma}\left[\|\mathrm{tr}\left(\Sigma^{-1}S\,S^{+}\,G\right)\|_F^2\right].$$

Similarly, note also that these conditions imply, with the condition $E^{*}_{\theta,\Sigma}\left[\left|\mathrm{tr}\left(\Sigma^{-1}SS^{+}T^{*}\right)\right|\right]<\infty$, the conditions in Corollary 2.1 with T^{*} in (29). In particular, for $V=2\,S+G$, we have $E_{\theta,\Sigma}\left[\left|\mathrm{tr}\left(\Sigma^{-1}SS^{+}V\,\Sigma^{-1}SS^{+}G\right)\right|\right]<\infty$.

Then, applying Lemma 2.1 to the second expectation in (25) and Corollary 2.1 to the first one with $V=2\,S+G$ gives (28), since

$$\begin{aligned}T^{*}&=2\left[SS^{+}\,(2\,S+G)\,\mathcal{D}_s\{SS^{+}\,G\}+SS^{+}\,G\,\mathcal{D}_s\{SS^{+}\,(2\,S+G)\}\right]-(p-m+1)\,G\,S^{+}\,(2\,S+G)\\&=4\,(S+SS^{+}\,G)\,\mathcal{D}_s\{SS^{+}\,G\}+4\,G\,SS^{+}\,\mathcal{D}_s\{S\}-(p-m+1)\,(2\,G\,SS^{+}+G\,S^{+}\,G)\\&=4\,(S+SS^{+}\,G)\,\mathcal{D}_s\{SS^{+}\,G\}+G\left(2\,m\,SS^{+}-(p-m+1)\,S^{+}\,G\right),\end{aligned}$$

where we used (A.63), the fact that $SS^{+}(I_p-SS^{+})=0_{p\times p}$ and the symmetry of $SS^{+}\,G$. □

As it is well known, the usual estimators of Σ are inappropriate in so far as the largest (smallest) eigenvalues of $\hat{\Sigma}_a$ tend to be larger (smaller) than the corresponding eigenvalues of Σ. This fact suggests that the eigenvalues of $\hat{\Sigma}_a$ be shrunk toward a central value, which gives rise to an orthogonally invariant estimator. Theorem 2.1 is well adapted to deal with this class of estimators.

3. Orthogonally invariant estimators

In this section, we propose competitive estimators of the scale matrix Σ through the eigenvalue decomposition of S. For that purpose, we set some notations. Let $\mathcal{O}_p$ be the group of $p \times p$ orthogonal matrices, and when $p \geq r$, let $\mathcal{L}_{p,r} = \{A \in \mathbb{R}^{p\times r} : A^\top A = I_r\}$ the Stiefel manifold of semi-orthogonal matrices. Note that $\mathcal{O}_p = \mathcal{L}_{p,p}$. Also define $\mathcal{D}_r$ as the set of $r \times r$ diagonal matrices $\mathrm{diag}(d_1, \ldots, d_r)$ such that $d_1 > \cdots > d_r > 0$. In this context, let

$$S = H_1 L H_1^\top$$

be the eigenvalue decomposition of S, where $H_1 \in \mathcal{L}_{p,p\wedge m}$ and $L \in \mathcal{D}_{p\wedge m}$.

We consider the subclass of alternatives estimators $\hat{\Sigma}_{a,G} = a\left(S + SS^+ G(Z, S)\right)$ in (5) that are orthogonally invariant: that is, such that

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

for some function $\Psi(L) \in \mathcal{D}_{p\wedge m}$ differentiable with respect to L. Note that $SS^+ G = G$, and hence, these estimators can be expressed as

$$\hat{\Sigma}_{a,\Psi} = a H_1 \left(L + \Psi(L)\right) H_1^\top . \tag{30}$$

In (30), the function $\Psi(L)$ modifies the eigenvalues in L. In the Gaussian context, Stein [29], Dey and Srinavasan [7], Haff [16], Kubokawa et al. [23], Ledoit and Wolf [25], Fisher and Sun [10] use such modifications. We give in the following corollary general conditions on $\Psi := \Psi(L)$, in the elliptically symmetric framework, for such estimators to improve over $\hat{\Sigma}_a$.

Corollary 3.1. *Consider a density as in (11) satisfying (18). Provided that*

$$\sum_{i=1}^{p\wedge m} \left\{ \left(p+m-2\,(p\wedge m)-1\right) \frac{\psi_i}{l_i} + 2\,\frac{\partial \psi_i}{\partial l_i} + \sum_{j\neq i}^{p\wedge m} \frac{\psi_i - \psi_j}{l_i - l_j} \right\} \geq 0\,, \tag{31}$$

the estimator $\hat{\Sigma}_{a_o,\Psi} = a_o H_1 \,\mathrm{diag}\big(l_1 + \psi_1, l_2 + \psi_2, \ldots, l_{p\wedge m} + \psi_{p\wedge m}\big) H_1^\top$ *improves over the optimal estimator* $\hat{\Sigma}_{a_o} = a_o H_1 \,\mathrm{diag}\big(l_1, l_2, \ldots, l_{p\wedge m}\big)\, H_1^\top$ *if*

$$\sum_{i=1}^{p\wedge m} \Bigg\{ (p+m-2\,(p\wedge m)-1) \frac{\phi_i}{l_i} + 2\,\frac{\partial \phi_i}{\partial l_i} + \sum_{j\neq i}^{p\wedge m} \frac{\phi_i - \phi_j}{l_i - l_j} -2\,(p+m+1)\frac{c^2}{b^2} \left(\left(p+m-2\,(p\wedge m)-1\right) \frac{\psi_i}{l_i} + 2\,\frac{\partial \psi_i}{\partial l_i} + \sum_{j\neq i}^{p\wedge m} \frac{\psi_i - \psi_j}{l_i - l_j} \right) \Bigg\} \leq 0\,, \tag{32}$$

where

$$\phi_i = 2\,(p+m-p\wedge m)\,\psi_i + (p+m-2\,(p\wedge m)-1)\,\frac{\psi_i^2}{l_i} + 4\,(\psi_i + l_i)\,\frac{\partial \psi_i}{\partial l_i} + 2\,(\psi_i + l_i) \sum_{j\neq i}^{p\wedge m} \frac{\psi_i - \psi_j}{l_i - l_j}\,. \tag{33}$$

Proof. Using the eigenvalue decomposition of S, it is clear that $S^+ = H_1 L^{-1} H_1^\top$, $SS^+ = H_1 H_1^\top$, $SS^+ G = G = H_1 \Psi H_1^\top$ and $S^+ G = H_1 L^{-1} \Psi H_1^\top$. In this setting, Condition (21) in Theorem 2.1 becomes

$$\mathrm{tr}\left[2\,H_1 H_1^\top \mathcal{D}_s\{H_1 \Psi H_1^\top\}(m-(p\wedge m)-1)\,H_1 L^{-1} \Psi H_1^\top\right] \geq 0.$$

Applying Identity (A.62) and using the fact that $H_1^\top H_1 = I_{p\wedge m}$ and $H_1^\top (I_p - H_1 H_1^\top) = 0_{(p\wedge m)\times p}$, this can be written as

$$\mathrm{tr}\left[2\,\Psi^{(1)} + (m-(p\wedge m)-1)\,L^{-1}\Psi\right] \geq 0 \tag{34}$$

where $\Psi^{(1)} \in \mathcal{D}_{p\wedge m}$ with

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

which is (31).

As for Condition (22) in Theorem 2.1, note that T^* in (23) is expressed as

$$T^* = 4\,H_1\,(L+\Psi)\,H_1^\top\,\mathcal{D}_s\{H_1\,\Psi\,H_1^\top\} + H_1\,\left(2\,m\,\Psi - (p-m+1)\,\Psi^2\,L^{-1}\right)\,H_1^\top .$$

Applying Identity (A.62) and using the facts that $H_1\,H_1^\top = I_{p\wedge m}$ and $H_1^\top(I_p - H_1 H_1^\top) = 0_{(p\wedge m)\times(p\wedge m)}$, it follows that

$$T^* = H_1\,\left(2\,m\,\Psi + 4\,(\Psi+L)\,\Psi^{(1)} - (p-m+1)\,\Psi^2\,L^{-1}\right)\,H_1^\top = H_1\,\Phi\,H_1^\top ,$$

where $\Phi \in \mathcal{D}_{p\wedge m}$ whose generic term is given in (33). Similarly to Condition (21), we apply Identity (A.62) with $G = T^*$ to obtain

$$\mathrm{tr}\left[\,2\,S^+S\,\mathcal{D}_s\{SS^+\,T^*\} + (m-(p\wedge m)-1)\,S^+\,T^*\,\right] = \mathrm{tr}\left[2\,\Phi^{(1)} + (m-(p\wedge m)-1)\,L^{-1}\,\Phi\right]\,, \tag{35}$$

where $\Phi^{(1)} \in \mathcal{D}_{p\wedge m}$ with

$$\phi_i^{(1)} = \frac{1}{2}\,(p-p\wedge m)\frac{\phi_i}{l_i} + \frac{\partial\phi_i}{\partial l_i} + \frac{1}{2}\sum_{j\neq i}^{p\wedge m}\frac{\phi_i-\phi_j}{l_i-l_j}.$$

Finally the desired result follows from (34) and (35). $\square$

4. Examples

4.1. Haff type estimator

The following proposition provides an example of function $\Psi(L)$ reducing to the estimator proposed by Haff [15] in the Gaussian context and in the case where S is invertible.

Proposition 4.1. *Consider a density as in (11) satisfying (18) and let p and m satisfy*

$$\frac{p+m-(p\wedge m)+2}{p+m+1} \le \frac{c^2}{b^2} \le \frac{2\,p+2\,m-5\,(p\wedge m)-3}{p+m+1}\,. \tag{36}$$

Let $\hat{\Sigma}_{a_o,\Psi}$ be an estimator of the form

$$\hat{\Sigma}_{a_o,\Psi} = a_o\,H_1\,\mathrm{diag}\big(l_1+\nu\,t(\nu), l_2+\nu\,t(\nu), \ldots, l_{p\wedge m}+\nu\,t(\nu)\big)H_1^\top, \tag{37}$$

with $\nu = 1/\mathrm{tr}(S^+)$ and $t(\cdot)$ a twice differentiable non-increasing convex function. Then $\hat{\Sigma}_{a_o,\Psi}$ improves upon $\hat{\Sigma}_{a_o}$ if

(i) $(p+m-2\,(p\wedge m)+1)\,t(\nu) + 2\,\nu\,t'(\nu) \ge 0$

(ii) $0 \le t(\nu) \le \dfrac{2\,(p+m-2\,(p\wedge m)-1)\,\big((p+m+1)\,c^2/b^2 - p - m + (p\wedge m) - 2\big)}{(p+m-2\,(p\wedge m)+1)(p+m-2\,(p\wedge m)+3)}$

(iii) $\left\{2\,(p+m-4\,(p\wedge m)+3)\,t(\nu) + 2\,\nu\,t'(\nu) + \left[2\,p+2\,m-5\,(p\wedge m)+5-(p+m+1)\dfrac{c^2}{b^2}\right]\right\}\,t'(\nu) + 2\,\left\{t(\nu)+(p\wedge m)^2\right\}\,\nu\,t''(\nu) \le 0\,.$

The proof of Proposition 4.1 is given in Appendix A.2. Although these conditions seem to be involved, a simple example is given by $t(\nu) = \alpha/(\beta+\nu)$, where $\alpha > 0$ and $\beta > 0$. Clearly $t'(\nu) = -1/(\beta+\nu)\,t(\nu) = -\alpha/(\beta+\nu)^2$ and $t''(\nu) = 2/(\beta+\nu)^2\,t(\nu) = 2\alpha/(\beta+\mu)^3$. Therefore, Condition (i) in Proposition 4.1 becomes

$$\left\{p+m-2\,(p\wedge m)+1-2\,\frac{\nu}{\beta+\nu}\right\}\,t(\nu) \ge 0,$$

which is satisfied if $p+m-2\,(p\wedge m) \ge 1$ since

$$\frac{\nu}{\beta+\nu} \le 1\,. \tag{38}$$

As for Condition (ii), note that $0 \le t(\nu) \le t(0) = \alpha/\beta$. Hence (ii) is satisfied if

$$\frac{\alpha}{\beta} \le t_1(p,m) = \frac{2\,(p+m-2\,(p\wedge m)-1)\,\big((p+m+1)(c^2/b^2) - p - m + p\wedge m - 2\big)}{(p+m-2\,(p\wedge m)+1)(p+m-2\,(p\wedge m)+3)}.$$

Now, Condition (iii) is equivalent to

$$\left\{6\,\frac{\nu}{\beta+\nu} - 2\,(p+m-4\,(p\wedge m)+3)\right\}\,\frac{1}{\beta+\nu}\,t^2(\nu) + \left\{4\,(p\wedge m)^2\,\frac{\nu}{\beta+\nu} - \left(2\,p+2\,m-5\,(p\wedge m)+5-(p+m+1)\frac{c^2}{b^2}\right)\right\}\,\frac{1}{\beta+\nu}\,t(\nu) \le 0\,.$$

According to (38), we have

$$\left\{-2\left(p+m-4(p\wedge m)\right)t(\nu)+4(p\wedge m)^2-\left(2p+2m-5(p\wedge m)+5-(p+m+1)\frac{c^2}{b^2}\right)\right\}\frac{1}{\beta+\nu}\,t(\nu)\le 0.$$

Now, using the fact that $t(\nu)\ge 0$, the last condition is satisfied if

$$t(\nu)\le t_2(p,m)=\frac{(p+m+1)(c^2/b^2)+4(p\wedge m)^2-2p-2m+5(p\wedge m)-5}{2(p+m-4(p\wedge m))}\,,$$

which holds if $\beta^{-1}\alpha\le t_2(p,m)$, since $t(\nu)\le t(0)=\alpha/\beta$.

Finally, choosing $\alpha/\beta\le t_1(p,m)\wedge t_2(p,m)$ implies that Conditions (*ii*) and (*iii*) in Proposition 4.1 are both satisfied.

As mentioned by Haff [15], even in the Gaussian framework and in the non-singular case, calculations under the quadratic loss in (4) are difficult. It is thus noteworthy that the above results hold for the considerably more general case of elliptically symmetric distributions, singular or non-singular S, and for a more general class of functions $t(\cdot)$.

4.2. The Konno estimator

The Konno estimator, proposed in [22] in the Gaussian setting, is a particular case of the estimator in (37); the function $t(\cdot)$ is constant. Thus it is of the form

$$\hat{\Sigma}_{KO}=a_o\left(S+\nu\,t\,H_1H_1^\top\right)=\frac{1}{K^{**}(p+m+1)}H_1\,\mathrm{diag}\big(l_1+\nu\,t,l_2+\nu\,t,\ldots,l_{p\wedge m}+\nu\,t\big)H_1^\top,\tag{39}$$

where $t>0$ and $\nu=1/\mathrm{tr}(S^+)$. As t is a constant, Condition (36) in Proposition 4.1 reduces to

$$\frac{p+m-(p\wedge m)+2}{p+m+1}\le\frac{c^2}{b^2},$$

since Condition (*iii*) is automatically satisfied (its left-hand side is zero, and hence, the second inequality in (36) does not intervene). Also Condition (*i*) is

$$p+m-2(p\wedge m)+1\ge 0\,.$$

Therefore the Konno estimator in (39) improves over the usual optimal estimator

$$\hat{\Sigma}_{a_o}=a_o\,S=\frac{1}{K^{**}(p+m+1)}H_1\,\mathrm{diag}\big(l_1,l_2,\ldots,l_{p\wedge m}\big)H_1^\top,$$

if

$$0\le t\le t_{\max}=\frac{2(p+m-2(p\wedge m)-1)\big((p+m+1)c^2/b^2-p-m+(p\wedge m)-2\big)}{(p+m-2(p\wedge m)+1)(p+m-2(p\wedge m)+3)}\,.\tag{40}$$

5. Numerical study

In this section, we report experiments designed to assess the behavior of the Konno's estimator in (39) with $t=t_{\max}$ in (40). We deal with two densities in the subclass (18). First, we consider the multivariate normal distribution for which $c=b=1$ and $K^*=K^{**}=1$. Secondly, we carry on with the variance mixture of normals with a beta distribution as mixing distribution. More specifically, the function $f(\cdot)$ in (11) has the form, for any $t\ge 0$,

$$f(t)=\int_0^1\frac{1}{(2\upsilon\pi)^{np/2}}\exp\left(\frac{-t}{2\upsilon}\right)h(\upsilon)\,d\upsilon\quad\text{where}\quad\forall\,\upsilon\in[0,1]\quad h(\upsilon)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\,\upsilon^{\alpha-1}(1-\upsilon)^{\beta-1}\,,\tag{41}$$

with $\alpha>np/2$ and $\beta>0$ and close to zero. It is seen in Appendix A.1 that $c=(\alpha-np/2)/(\alpha-np/2+\beta)$, $b=1$ and $K^{**}=\big(\alpha\beta+\alpha^2(\alpha+\beta+1)\big)/\big((\alpha+\beta)^2(\alpha+\beta+1)\big)$. As in our example, the corrected factor $SS^+G(Z,S)$ depends only on $S=U^\top U$ where U is $n\times p$ matrix, in the above conditions, the role of n is played by m.

We carry out simulations for the following structures of the invertible scale matrix Σ:

(i) I_p;
(ii) the $p\times p$ matrix where the (i,j)th element is $0.9^{|i-j|}$;
(iii) $\mathrm{diag}(100,100^{1-1/p},\ldots,100^{1-(p-1)/p})$.

Note that Case (ii) has an autoregressive structure with coefficient 0.9. Also, Case (iii) corresponds to a heteroscedastic scale matrix for which the diagonals elements are widely scattered: the largest diagonal element is about hundredfold of the smallest one. To assess how $\hat{\Sigma}_{KO}$ improves over $\hat{\Sigma}_{a_o}$, we compute, for each structure of Σ, the Percentage Reduction In Average Loss (PRIAL) defined as

$$\mathrm{PRIAL}(\hat{\Sigma}_{KO})=\frac{\text{average loss of }\hat{\Sigma}_{a_o}-\text{average loss of }\hat{\Sigma}_{KO}}{\text{average loss of }\hat{\Sigma}_{a_o}}.$$

Table 1
PRIAL's (%) of $\hat{\Sigma}_{KO}$ for the Gaussian and the beta mixture distributions. The non-invertible ($p > m$) and the invertible ($p \leq m$) cases are considered for the structures (i), (ii) and (iii) of Σ.

Σ	p	m	Gaussian	Mixture	m	p	Gaussian	Mixture
(i)	25	10	1.09	1.01	25	10	6.52	5.96
		15	1.88	1.79		15	5.10	4.84
		20	1.94	1.88		20	3.07	2.88
	100	40	0.39	0.37	100	40	2.45	2.29
		60	0.75	0.71		60	2.07	1.97
		80	1.04	1.00		80	1.62	1.56
(ii)	25	10	1.04	0.96	25	10	6.25	5.72
		15	1.84	1.75		15	4.99	4.75
		20	1.94	1.87		25	3.06	2.87
	100	40	0.39	0.36	100	40	2.42	2.26
		60	0.74	0.70		60	2.06	1.95
		80	1.03	0.99		80	1.62	1.56
(iii)	25	10	1.04	0.97	25	10	5.91	5.42
		15	1.81	1.73		15	4.86	4.61
		20	1.94	1.87		20	3.03	2.85
	100	40	0.39	0.36	100	40	2.37	2.22
		60	0.74	0.70		60	2.03	1.93
		80	1.03	0.99		80	1.61	1.55

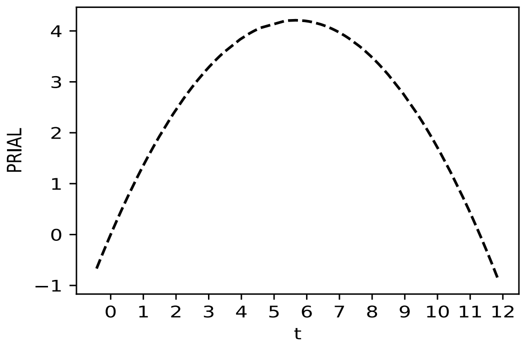

Fig. 1. Effect of t on the PRIAL for $(p, m) = (25, 10)$.

Table 1 shows the PRIAL's based on 1000 independent replications for some couples (p, m) where we deal both with the non-invertible case ($p > m$, second and third columns) and the invertible case ($p \leq m$, sixth and seventh columns). The first column displays the three structures (i), (ii) and (iii) of Σ. The fourth and eighth columns correspond to the simulations under the Gaussian distribution $\mathcal{N}(0, I_m \otimes \Sigma)$ while the fifth and the ninth columns correspond to the mixing distribution in (41) with $\alpha = mp/2 + 0.1$ and $\beta = 10^{-4}$.

It is observed that the PRIAL's are better when $\Sigma = I_p$ and coincide with the simulation results of Konno [22] for the Gaussian setting when $p > m$. For a fixed p, the PRIAL's increase with m and, inversely, decrease with p for a fixed m. Note that, the PRIAL's are better in the invertible case ($p \leq m$) with respect to the non-invertible case ($p > m$). Also, note that, for all considered structures of Σ, the PRIAL's are slightly better in the Gaussian setting.

Finally, we investigate in the Gaussian framework the effect of the constant t in (40) on the PRIAL. Fig. 1 provides the graphic of the PRIAL as a function of t when the scale matrix is $\Sigma = I_p$ and $(p, m) = (25, 10)$.

We observe that, when $0 < t \leq t_{max} \approx 0.875$, there is improvement of $\hat{\Sigma}_{KO}$ over $\hat{\Sigma}_{a_o}$, while, when $t < 0$, there is no improvement. It is worth noting that, when $t > t_{\max}$, there still exists improvement. The PRIAL increases till (approximately) 4.20% at $t \approx 5.66$ and then decreases till 0 at $t \approx 11.29$. Thus, it appears that the Konno class of improved estimators in (39) is larger than in our approach.

6. Conclusions and perspectives

In this paper, we address the problem of estimating the scale matrix Σ of an elliptically symmetric distribution belonging to a subclass which is reminiscent of the Berger class [1]. We derive dominance results for estimators of the form $a(S + SS^{+} G(Z, S))$ over the usual estimators $a\,S$ under quadratic loss. Thanks to a new Stein–Haff type identity, our approach unifies the cases where S is invertible and S is singular. Results on eigenstructure of S allow to provide general examples of improved orthogonally invariant estimators which extend the results in Haff [17] and Konno [22].

Note that, as Σ^{-1} naturally intervenes in the parameterization of the density (11), estimating Σ is often more difficult than estimating Σ^{-1}.

A natural extension of this work would be to determine examples of improved estimators which are not orthogonally invariant. We have already obtained such examples when S is invertible and we plan to tackle this question when S is singular. Also, calculations under the quadratic loss $L(\Sigma, \hat{\Sigma}) = \operatorname{tr}(\Sigma^{-1}\hat{\Sigma}\,\Sigma^{-1}\hat{\Sigma}) - 2\operatorname{tr}(\Sigma^{-1}\hat{\Sigma}) + p$ are difficult and give rise to complicated improvement conditions. This is due to the presence of the unknown parameter Σ^{-1} in $\operatorname{tr}(\Sigma^{-1}\hat{\Sigma}\,\Sigma^{-1}\hat{\Sigma})$, which requires a dual application of the Stein–Haff type identity (26). A more suitable loss function would be the data-based loss (also called intrinsic loss)

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

considered by Tsukuma and Kubokawa [32]. In fact, using this loss implies only one application of the Stein–Haff type identity to get rid of Σ^{-1}. This may lead to more simple improvement conditions.

CRediT authorship contribution statement

Anis M. Haddouche: Conceptualization, Methodology, Supervision, Visualization, Validation, Writing - review & editing, Writing - original draft, Software. **Dominique Fourdrinier:** Conceptualization, Methodology, Supervision, Visualization, Validation, Writing - review & editing, Writing - original draft, Software. **Fatiha Mezoued:** Conceptualization, Methodology, Supervision, Visualization, Validation, Writing - review & editing, Writing - original draft, Software.

Acknowledgments

The authors are grateful to the Editor-in-Chief and to two referees for their valuable comments and suggestions that strengthened the exposition and scope of this paper. All the authors are supported by the Franco-Algerian program PROFAS B+ (2016–2019). D. Fourdrinier is also partially supported by the Russian program #1.472.2016/1.4 Ministry of Education and Science of the Russian Federation.

Appendix

A.1. A subclass of elliptically symmetric distributions

Numerous densities in (11) satisfying Condition (18) are given in Fourdrinier, Mezoued and Strawderman [11]; they contain the multivariate normal density. As an example of interest for us, we consider the variance mixture of normal distributions where the mixing variable V has a beta distribution with parameters $\alpha > 0$ and $\beta > 0$. Thus, the generating function f in (11) has, for any $t \geq 0$, the form

$$f(t) = \int_0^1 \frac{1}{(2v\pi)^{np/2}} \exp\left(\frac{-t}{2v}\right) h(v)\,dv \tag{A.1}$$

where, for any $0 \leq v \leq 1$,

$$h(v) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\, v^{\alpha-1}(1-v)^{\beta-1}.$$

Then, the primitive F^* of f in (14) is, for any $t \geq 0$,

$$\begin{aligned} F^*(t) &= \frac{1}{2}\int_t^\infty \int_0^1 \frac{1}{(2v\pi)^{np/2}} \exp\left(\frac{-w}{2v}\right) h(v)\,dv\,dw = \frac{1}{2}\int_0^1 \frac{1}{(2v\pi)^{np/2}} \int_t^\infty \exp\left(\frac{-w}{2v}\right) dw\, h(v)\,dv \\ &= \int_0^1 \frac{v}{(2v\pi)^{np/2}} \exp\left(\frac{-t}{2v}\right) h(v)\,dv\,. \end{aligned} \tag{A.2}$$

Following the same lines, the primitive F^{**} of F^* in (14) is, for any $t \geq 0$,

$$F^{**}(t) = \int_0^1 \frac{v^2}{(2v\pi)^{np/2}} \exp\left(\frac{-t}{2v}\right) h(v)\,dv\,. \tag{A.3}$$

As for the normalizing constant K^* in (15), according to (A.2), we have

$$\begin{aligned} K^* &= \int_{\mathbb{R}^{pn}} \int_0^1 \frac{|\Sigma|^{-n/2}}{(2v\pi)^{np/2}}\, v \exp\left(\frac{-1}{2v}\left[\operatorname{tr}\{(z-\theta)\,\Sigma^{-1}(z-\theta)^\top\} + \operatorname{tr}\{\Sigma^{-1}u^\top u\}\right]\right) h(v)\,dv\;dz\,du \\ &= \int_0^1 v \int_{\mathbb{R}^{pn}} \frac{|\Sigma|^{-n/2}}{(2\,v\pi)^{np/2}} \exp\left(\frac{-1}{2v}\left[\operatorname{tr}\{(z-\theta)\,\Sigma^{-1}(z-\theta)^\top\} + \operatorname{tr}\{\Sigma^{-1}u^\top u\}\right]\right) dz\,du\,h(v)\,dv \\ &= \int_0^1 v\,h(v)\,dv = \frac{\alpha}{\alpha+\beta}\,. \end{aligned}$$

Similarly, according to (A.3), the normalizing constant K^{**} in (16) is

$$K^{**} = \int_0^1 v^2 \int_{\mathbb{R}^{pn}} \frac{|\Sigma|^{-n/2}}{(2v\pi)^{np/2}} \exp\left(\frac{-1}{2v}\operatorname{tr}\left[\operatorname{tr}\{(z-\theta)\,\Sigma^{-1}(z-\theta)^\top\} + \operatorname{tr}\{\Sigma^{-1}u^\top u\}\right]\right) dz\,du\,h(v)\,dv$$
$$= \int_0^1 v^2\,h(v)\,dv = \frac{\alpha\,\beta + \alpha^2(\alpha+\beta+1)}{(\alpha+\beta)^2\,(\alpha+\beta+1)}\,.$$

As we deal with a variance mixture of normals, the ratio $F^*(t)/f(t)$ is non-decreasing in t (see Fourdrinier, Strawderman and Wells [13]). Hence c and b in (18) are, $c = F^*(0)/f(0)$ and $b = \lim_{t\to\infty} F^*(t)/f(t)$. Also, from (A.1) and (A.2) we have,

$$\frac{F^*(t)}{f(t)} = \frac{\int_0^1 v^{\alpha-np/2}\,(1-v)^{\beta-1}\,\exp\left(\frac{-t}{2v}\right)\,dv}{\int_0^1 v^{\alpha-np/2-1}\,(1-v)^{\beta-1}\,\exp\left(\frac{-t}{2v}\right)\,dv} \le 1,$$

so that, $b = 1$. As for c, if $\alpha > np/2$, we have

$$c = \frac{F^*(0)}{f(0)} = \frac{\int_0^1 v^{\alpha-np/2}\,(1-v)^{\beta-1}\,dv}{\int_0^1 v^{\alpha-np/2-1}\,(1-v)^{\beta-1}\,dv} = \frac{\alpha-np/2}{\alpha-np/2+\beta},$$

as a ratio of beta functions.

While, in the multivariate normal case $c = b = 1$ since $F^{**} = F^* = f$, and hence, the expectations $E_{\theta,\Sigma}$, $E^*_{\theta,\Sigma}$, $E^{**}_{\theta,\Sigma}$ coincide, in the elliptical setting, these expectations are related since, for any integrable function $H(Z,U)$, we have

$$K^*\,E^*_{\theta,\Sigma}[H(Z,U)] = E_{\theta,\Sigma}[\varphi^*_{\theta,\Sigma}(Z,U)\,H(Z,U)] \tag{A.4}$$

and

$$K^{**}\,E^{**}_{\theta,\Sigma}[H(Z,U)] = K^*E^*_{\theta,\Sigma}[\varphi^{**}_{\theta,\Sigma}(Z,U)\,H(Z,U)] = E_{\theta,\Sigma}[\varphi^{**}_{\theta,\Sigma}(Z,U)\varphi^*_{\theta,\Sigma}(Z,U)\,H(Z,U)] \tag{A.5}$$

where, for any $z \in \mathbb{R}^{q\times p}$ and $u \in \mathbb{R}^{m\times p}$,

$$\varphi^*_{\theta,\Sigma}(z,u) = \frac{F^*(\nu)}{f(\nu)} \quad\text{and}\quad \varphi^{**}_{\theta,\Sigma}(z,u) = \frac{F^{**}(\nu)}{F^*(\nu)},\quad\text{with}\quad \nu = \operatorname{tr}\{(z-\theta)\,\Sigma^{-1}(z-\theta)^\top\} + \operatorname{tr}\{\Sigma^{-1}u^\top u\}. \tag{A.6}$$

The following lemma is helpful to deal with the dependence of the risk difference in (A.51) with respect to the unknown parameters θ and Σ in $\varphi^*_{\theta,\Sigma}(z,u)$ and $\varphi^{**}_{\theta,\Sigma}(z,u)$.

Lemma A.1. *In the context of (A.6), assume that there exist two positives constant c and b such that for all $t \in \mathbb{R}^+$*

$$c \le \frac{F^*(t)}{f(t)} \le b\,. \tag{A.7}$$

Then we have

$$\frac{1}{b^2} \le \frac{1}{K^*\varphi^{**}_{\theta,\Sigma}(z,u)} \le \frac{1}{c^2}\,, \tag{A.8}$$

where K^* is given in (15).

Proof. By definition of $F^{**}(\cdot)$ and $F^*(\cdot)$ in (14) and (14), we have for all $t \in \mathbb{R}^+$

$$\frac{F^*(t)}{F^{**}(t)} = \left(\int_t^\infty F^*(w)\,dw\right)^{-1} \int_t^\infty \frac{f(w)}{F^*(w)}\,F^*(w)\,dx = E_t\left[\frac{f(W)}{F^*(W)}\right],$$

where E_t is the expectation with respect to the density proportional to $F^*(w)\,\mathbb{1}_{[t,\infty[}(w)$. Hence, according to (A.6) and (A.7) with $t = \nu$, we have

$$\frac{1}{b} \le \frac{1}{\varphi^{**}_{\theta,\Sigma}(z,u)} = E_\nu\left[\frac{f(W)}{F^*(W)}\right] \le \frac{1}{c}\,. \tag{A.9}$$

Moreover, setting $H(Z,U) \equiv 1$ in (A.4), we have, according to (A.6) and by assumption (A.7), $c \le K^* \le b$. Therefore $1/b \le 1/K^* \le 1/c$ and (A.9) gives (A.8). □

A.2. Matrix calculations

In this appendix, for completeness, we give a proof of some known results relative to the Haff operator $\mathcal{D}_s\{\cdot\}$ defined in (20) and to the trace of S and S^+. We also give here the proof of Proposition 4.1.

Proposition A.1. *Let A, B and C be p × p matrix functions of S. Assuming that all partial derivatives and products exist as needed, we have*

$$\mathcal{D}_s\{A\,B\} = \mathcal{D}_s\{A\}\,B + \left(A^\top \mathcal{D}_S\right)^\top \{B\} \tag{A.10}$$

$$\mathrm{tr}\left[C\,(A^\top \mathcal{D}_s)^\top \{B\}\right] = \mathrm{tr}\left[A^\top\,(C\,\mathcal{D}_s)^\top\,\{B\}^\top\right] \tag{A.11}$$

Proof. The first identity was given by Haff [14,15], so that we only give the proof of (A.11). We have

$$\mathrm{tr}\left[C\,(A^\top \mathcal{D}_s)^\top \{B\}\right] = \sum_{i,j,k}^{p} C_{ij}(A^\top \mathcal{D}_s)^\top_{jk} B_{ki} = \sum_{i,j,k}^{p} C_{ij}(A^\top \mathcal{D}_s)_{kj} B^\top_{ik}.$$

Then, by symmetry of $\mathcal{D}_s\{\cdot\}$,

$$\mathrm{tr}\left[C\,(A^\top \mathcal{D}_s)^\top \{B\}\right] = \sum_{i,j,k,l}^{p} A^\top_{kl} C_{ij} d_{jl} B^\top_{ik} = \sum_{k,l,i}^{p} A^\top_{kl} (C\mathcal{D}_s)^\top_{li} B^\top_{ik} = \mathrm{tr}\left[A^\top\,(C\,\mathcal{D}_s)^\top\,\{B\}^\top\right]. \quad \square$$

The following lemmas rely on the eigenvalue decomposition of S recalled in the beginning of Section 3.

Lemma A.2. *Let $\rho_k = \mathrm{tr}(S^{+k})/\mathrm{tr}^k(S^+)$. Then, for $k \in \{2, 3, 4, \dots\}$,*

$$(p \wedge m)^{1-k} \le \rho_k \le \rho_{k-1} \le 1\,. \tag{A.12}$$

Proof. First, we show that ρ_k is non-increasing in k. As $S^+ = H_1 L^{-1} H_1^\top$, we have

$$\mathrm{tr}(S^{+k}) = \mathrm{tr}(L^{-k}) = \sum_{i=1}^{p\wedge m} \frac{1}{l_i^k} \le \sum_{i=1}^{p\wedge m} \frac{1}{l_i^{k-1}} \sum_{j=1}^{p\wedge m} \frac{1}{l_j} = \mathrm{tr}(S^{+(k-1)})\mathrm{tr}(S^+).$$

Hence

$$\rho_k = \frac{\mathrm{tr}(S^{+k})}{\mathrm{tr}^k(S^+)} \le \frac{\mathrm{tr}(S^{+(k-1)})}{\mathrm{tr}^{k-1}(S^+)} = \rho_{k-1}. \tag{A.13}$$

Secondly, setting $k = 2$ in (A.13) gives $\rho_2 \le 1$. Now, from the Hölder inequality, it follows that, for $k \in \{2, 3, 4, \dots\}$

$$\mathrm{tr}\left(L^{-1}\,I_{p\wedge m}\right) \le \left(\mathrm{tr}(L^{-k})\right)^{1/k} (p \wedge m)^{(k-1)/k}.$$

Then

$$\mathrm{tr}^k(L^{-1}) \le \mathrm{tr}(L^{-k})(p\wedge m)^{k-1}\,, \quad \text{and hence,} \quad (p\wedge m)^{1-k} \le \frac{\mathrm{tr}(L^{-k})}{\mathrm{tr}^k(L^{-1})} = \frac{\mathrm{tr}(S^{+k})}{\mathrm{tr}^k(S^+)} = \rho_k. \quad \square$$

Lemma A.3. *The following inequalities hold:*

$$-4\,(p\wedge m)\,\mathrm{tr}(S^{+3}) < \sum_{i=1}^{p\wedge m} \sum_{j\ne i}^{p\wedge m} \frac{l_i^{-2} - l_j^{-2}}{l_i - l_j} < 0, \tag{A.14}$$

and

$$-2\,(p\wedge m)\,\mathrm{tr}(S^{+2}) < \sum_{i=1}^{p\wedge m} \sum_{j\ne i}^{p\wedge m} \frac{l_i^{-1} - l_j^{-1}}{l_i - l_j} < 0\,. \tag{A.15}$$

Proof. As $S^+ = H_1\,L^{-1}\,H_1^\top = H_1\,\mathrm{diag}(l_1^{-1}, l_2^{-1}, \dots, l_{p\wedge m}^{-1})\,H_1^\top$, we have

$$\sum_{i=1}^{p\wedge m} \sum_{j\ne i}^{p\wedge m} \frac{l_i^{-2} - l_j^{-2}}{l_i - l_j} = 2\sum_{i=1}^{p\wedge m} \sum_{j> i}^{p\wedge m} \frac{l_i^{-2} - l_j^{-2}}{l_i - l_j} = -2\sum_{i=1}^{p\wedge m} \sum_{j> i}^{p\wedge m} \left[\frac{1}{l_i\,l_j^2} + \frac{1}{l_i^2\,l_j}\right],$$

which is non-positive. As for $j > i$, we have $l_i^{-1} < l_j^{-1}$ and $l_i^{-2} < l_j^{-2}$, then

$$\sum_{i=1}^{m} \sum_{j\ne i}^{p\wedge m} \frac{l_i^{-2} - l_j^{-2}}{l_i - l_j} > -4\sum_{i=1}^{p\wedge m} \sum_{j> i}^{p\wedge m} \frac{1}{l_j^3} = -4\sum_{i=1}^{p\wedge m} \left[\sum_{j=1}^{p\wedge m} \frac{1}{l_j^3} - \sum_{j\le i}^{p\wedge m} \frac{1}{l_j^3}\right].$$

Now, since $\sum_{i=1}^{p\wedge m}\sum_{j\le i}^{p\wedge m} l_j^{-3} > 0$, we have

$$\sum_{i=1}^{p\wedge m}\sum_{j\neq i}^{p\wedge m}\frac{l_i^{-2}-l_j^{-2}}{l_i-l_j} > -4\sum_{i=1}^{p\wedge m}\sum_{j=1}^{p\wedge m}\frac{1}{l_j^3} = -4\,(p\wedge m)\,\mathrm{tr}(S^{+3}),$$

which is (A.14).

Similarly, dealing with (A.15), we have

$$\sum_{i=1}^{p\wedge m}\sum_{j\neq i}^{p\wedge m}\frac{l_i^{-1}-l_j^{-1}}{l_i-l_j} = -2\sum_{i=1}^{p\wedge m}\sum_{j>i}^{p\wedge m}\frac{1}{l_i\,l_j} > -2\sum_{i=1}^{p\wedge m}\sum_{j>i}^{p\wedge m}\frac{1}{l_j^2} = -2\sum_{i=1}^{p\wedge m}\left[\sum_{j=1}^{p\wedge m}\frac{1}{l_j^2}-\sum_{j\le i}^{p\wedge m}\frac{1}{l_j^2}\right].$$

Hence, using the fact that $\sum_{i=1}^{p\wedge m}\sum_{j\le i}^{p\wedge m} l_j^{-2} > 0$ gives the desired result in (A.15). □

Proof of Proposition 4.1. The proof consists in proving Conditions (31) and (32) in Corollary 3.1 for $\Psi = \nu\, t(\nu)\, I_{p\wedge m}$, where $\nu = 1/\mathrm{tr}(S^+)$ and $t(\cdot)$ is a twice differentiable non-increasing convex function. Note that Condition (36) holds if and only if $p+m-(p\wedge m)+2 \le 2\,p+2\,m-5\,(p\wedge m)-3$ which is equivalent to

$$p+m \ge 4\,(p\wedge m)+5.$$

Let $\rho_k = \nu^k\,\mathrm{tr}(S^{+k})$. In this setting, Condition (31) is equivalent to

$$\sum_i^{p\wedge m}\left\{(p+m-2\,(p\wedge m)-1)\,\frac{\nu\, t(\nu)}{l_i}+2\,\frac{\partial\{\nu\, t(\nu)\}}{\partial l_i}\right\} \ge 0. \tag{A.16}$$

Using the fact that, for $r, k \in \{0, 1, 2, \dots\}$

$$\frac{\partial}{\partial l_i}\left\{\frac{\nu^k}{l_i^r}\right\} = k\,\frac{\nu^{k+1}}{l_i^{r+2}} - r\,\frac{\nu^k}{l_i^{r+1}} \quad\text{and}\quad \frac{\partial}{\partial l_i}\{t(\nu)\} = \frac{\nu^2}{l_i^2}\, t'(\nu)\,. \tag{A.17}$$

Condition (A.16) becomes

$$(p+m-2\,(p\wedge m)-1)\; t(\nu)+2\,\rho_2\, t(\nu)+2\,\rho_2\,\nu\, t'(\nu)\ge 0\,. \tag{A.18}$$

Thanks to the second Inequality in (A.12), this condition is satisfied if Condition (i) holds.

Dealing with Condition (32), we also use (A.17) to express (33) as

$$\phi_i = \left\{4\,\frac{\nu^3}{l_i^2}+(p+m-2\,(p\wedge m)-1)\,\frac{\nu^2}{l_i}\right\} t^2(\nu)+2\,\left\{2\,\frac{\nu^2}{l_i}+\nu\,(p+m-p\wedge m)\right\}\, t(\nu)+4\,\left\{\frac{\nu^3}{l_i^2}\, t(\nu)+\frac{\nu^2}{l_i}\right\}\,\nu\, t'(\nu).$$

Then, it follows that

$$\sum_i^{p\wedge m}\frac{\phi_i}{l_i} = \{4\,\rho_3+(p+m-2\,(p\wedge m)-1)\,\rho_2\}\; t^2(\nu)+2\;\{2\,\rho_2+(p+m-p\wedge m)\}\; t(\nu)+4\;\{\rho_3\, t(\nu)+\rho_2\}\;\nu\, t'(\nu) \tag{A.19}$$

and

$$\begin{aligned}\sum_i^{p\wedge m}\sum_{j\neq i}^{p\wedge m}\frac{\phi_i-\phi_j}{l_i-l_j} &= \left\{4\,\nu^3\sum_i^{p\wedge m}\sum_{j\neq i}^{p\wedge m}\frac{l_i^{-2}-l_j^{-2}}{l_i-l_j}+(p+m-2\,(p\wedge m)-1)\,\nu^2\sum_i^{p\wedge m}\sum_{j\neq i}^{p\wedge m}\frac{l_i^{-1}-l_j^{-1}}{l_i-l_j}\right\}\; t^2(\nu)\\ &\quad+4\,\nu^2\sum_i^{p\wedge m}\sum_{j\neq i}^{p\wedge m}\frac{l_i^{-1}-l_j^{-1}}{l_i-l_j}\, t(\nu)+4\,\left\{\nu^3\, t(\nu)\sum_i^{p\wedge m}\sum_{j\neq i}^{p\wedge m}\frac{l_i^{-2}-l_j^{-2}}{l_i-l_j}+\nu^2\sum_i^{p\wedge m}\sum_{j\neq i}^{p\wedge m}\frac{l_i^{-1}-l_j^{-1}}{l_i-l_j}\right\}\,\nu\, t'(\nu)\,.\end{aligned} \tag{A.20}$$

Moreover, using (A.17) repeatedly gives

$$\begin{aligned}\sum_i^{p\wedge m}\frac{\partial\phi_i}{\partial l_i} &= \{12\,\rho_4+2\,(p+m-2\,(p\wedge m)-5)\,\rho_3-(p+m-2\,(p\wedge m)-1)\,\rho_2\}\; t^2(\nu)\\ &\quad+2\;\{4\,\rho_3+(p+m-(p\wedge m)-2)\,\rho_2\}\; t(\nu)+4\;\{\rho_4\, t(\nu)+\rho_3\}\,\nu^2\, t''(\nu)\\ &\quad+2\,\big\{2\,\rho_4\,\nu\, t'(\nu)+12\rho_4\, t(\nu)+(p+m-2\,(p\wedge m)-5)\,\rho_3\, t(\nu)+16\rho_3\\ &\quad+(p+m-(p\wedge m)-2)\rho_2\big\}\;\nu t'(\nu).\end{aligned} \tag{A.21}$$

Substituting (A.19), (A.20), (A.21) and the left-hand side of (A.18) on the left-hand side of Condition (32), this last condition can be expressed as

$$H_1(z,u)+H_2(z,u)\le 0 \tag{A.22}$$

where

$$\begin{aligned} H_1(z,u) &= 2\,\{12\,\rho_4+2\,(p+m-2\,(p\wedge m)-5)\rho_3-(p+m-2\,(p\wedge m)-1)\,\rho_2\}\;t^2(\nu)\\ &+(p+m-2\,(p\wedge m)-1)\;\{(p+m-2\,(p\wedge m)-1)\,\rho_2+4\,\rho_3\}\;t^2(\nu)\\ &+2\,\left\{(p+m-2\,(p\wedge m)-1)\,\left(p+m-(p\wedge m)+2\,\rho_2-\frac{c^2}{b^2}(p+m+1)\right)\right\}\;t(\nu)\\ &+4\,\left\{4\,\rho_3+(p+m-(p\wedge m)-2)\,\rho_2-\frac{c^2}{b^2}(p+m+1)\,\rho_2\right\}\;t(\nu)\\ &\nu^2\,\left\{4\,\nu\sum_i^{p\wedge m}\sum_{j\neq i}^{p\wedge m}\frac{l_i^{-2}-l_j^{-2}}{l_i-l_j}+(p+m+2\,(p\wedge m)-1)\sum_i^{p\wedge m}\sum_{j\neq i}^{p\wedge m}\frac{l_i^{-1}-l_j^{-1}}{l_i-l_j}\right\}\;t^2(\nu) \end{aligned} \tag{A.23}$$

and

$$\begin{aligned} H_2(z,u) = &\left\{48\,\rho_4\,t(\nu)+8\,\rho_4\,\nu\,t'(\nu)+8\,(p+m-2\,(p\wedge m)-3)\,\rho_3\,t(\nu)+32\,\rho_3\right.\\ &+4(2p+2m-3\,(p\wedge m)-3)\,\rho_2-4\,\frac{c^2}{b^2}(p+m+1)\,\rho_2\\ &\left.+4\,\nu^3\,t(\nu)\sum_i^{p\wedge m}\sum_{j\neq i}^{p\wedge m}\frac{l_i^{-2}-l_j^{-2}}{l_i-l_j}+4\,\nu^2\sum_i^{p\wedge m}\sum_{j\neq i}^{p\wedge m}\frac{l_i^{-1}-l_j^{-1}}{l_i-l_j}\right\}\,\nu\,t'(\nu)+8\,\{\rho_4\,t(\nu)+\rho_3\}\;\nu^2 t''(\nu)\,. \end{aligned} \tag{A.24}$$

Successive upper bounds will be developed for $H_1(z,u)$ in (A.23) and $H_2(z,u)$ in (A.24). First, it follows from the second inequalities in (A.14) and (A.15) in Lemma A.3 that the last term on the right-hand side of (A.23) is non-positive. Then, using the fact that $\rho_4\le\rho_3\le\rho_2$ (see Lemma A.2), we have

$$\begin{aligned} H_1(z,u) &\le (p+m+2\,(p\wedge m)+3)(p+m-2\,(p\wedge m)+1)\,\rho_2\,t^2(\nu)\\ &+2\,(p+m-2\,(p\wedge m)-1)\left(p+m-(p\wedge m)+2\,\rho_2-(p+m+1)\frac{c^2}{b^2}\right)\;t(\nu)\\ &+4\,\rho_2\,\left(p+m-(p\wedge m)+2-(p+m+1)\frac{c^2}{b^2}\right)\;t(\nu)\,, \end{aligned}$$

since $t(\cdot)\ge 0$. Therefore, it follows from the first inequality in (36) that the third term on the right-hand side of the last inequality is non-positive. Then, since $\rho_2\le 1$, an upper bound for (A.23) is given by

$$\begin{aligned} H_1(z,u) &\le (p+m+2\,(p\wedge m)+3)(p+m-2\,(p\wedge m)+1)\,t^2(\nu)\\ &+2\,(p+m-2\,(p\wedge m)-1)\left(p+m-p\wedge m+2-(p+m+1)\frac{c^2}{b^2}\right)\;t(\nu)\,, \end{aligned}$$

which is non-positive if Condition (*ii*) holds.

As for the term $H_2(z,u)$ in (A.24), according to the first inequalities in (A.14) and (A.15) in Lemma A.3, the l_i's terms between brackets satisfy

$$4\,\nu^3\,t(\nu)\sum_i^{p\wedge m}\sum_{j\neq i}^{p\wedge m}\frac{l_i^{-2}-l_j^{-2}}{l_i-l_j}+4\,\nu^2\sum_i^{p\wedge m}\sum_{j\neq i}^{p\wedge m}\frac{l_i^{-1}-l_j^{-1}}{l_i-l_j}\ge -8\,\{2\,(p\wedge m)\,\rho_3\,t(\nu)+(p\wedge m)\,\rho_2\},$$

which implies that

$$\begin{aligned} H_2(z,u)\le &\left\{48\,\rho_4\,t(\nu)+8\,\rho_4\,\nu\,t'(\nu)+8\,(p+m-4\,(p\wedge m)-3)\,\rho_3\,t(\nu)+32\,\rho_3\right.\\ &\left.+4\,\left(2\,p+2\,m-5(p\wedge m)-3-(p+m+1)\frac{c^2}{b^2}\right)\,\rho_2\right\}\,\nu\,t'(\nu)+8\,\{\rho_4\,t(\nu)+\rho_3\}\;\nu^2\,t''(\nu)\,, \end{aligned}$$

since $t'(\cdot)\le 0$. Next, according to the second inequality in (36) and thanks to Lemma A.2, it is clear that

$$32\,\rho_3+4\,\left(2\,p+2\,m-5\,(p\wedge m)-3-(p+m+1)\,\frac{c^2}{b^2}\right)\,\rho_2\ge 4\,\left(2\,p+2\,m-4\,(p\wedge m)+5-(p+m+1)\,\frac{c^2}{b^2}\right)\,\rho_4.$$

Therefore, as $t'(\cdot)$ is non-positive, we have

$$H_2(z,u) \leq \Bigg\{48\,\rho_4\,t(\nu) + 8\,\rho_4\,\nu\,t'(\nu) + 8\,(p+m-4\,(p\wedge m)-3)\,\rho_3\,t(\nu) + 4\left(2\,p+2\,m-5(p\wedge m)-3-(p+m+1)\frac{c^2}{b^2}\right)\rho_4\Bigg\}\,\nu\,t'(\nu) + 8\,\{\rho_4\,t(\nu)+\rho_3\}\,\nu^2\,t''(\nu)\,.$$

Now, using the fact that $\rho_3 \geq \rho_4$ and factorizing $4\,\rho_4$ give that

$$H_2(z,u) \leq 4\,\rho_4\left\{2\,(p+m-4\,(p\wedge m)+3)\,t(\nu) + 2\,\nu\,t'(\nu) + \left(2\,p+2\,m-5\,(p\wedge m)+5-(p+m+1)\frac{c^2}{b^2}\right)\right\} \times \nu\,t'(\nu) + 8\,\rho_4\left\{t(\nu)+\frac{\rho_3}{\rho_4}\right\}\nu^2\,t''(\nu)\,. \tag{A.25}$$

Now, it follows from the Cauchy–Schwarz inequality that

$$\mathrm{tr}^2(S^{+3}) \leq \mathrm{tr}(S^{+4})\,\mathrm{tr}(S^{+2}) \leq \mathrm{tr}(S^{+4})\,\mathrm{tr}^2(S^{+}),$$

where for the second inequality we apply (A.12) in Lemma A.2 for $k=2$. Therefore, we have

$$\frac{\rho_3}{\rho_4} = \frac{\mathrm{tr}(S^{+})\,\mathrm{tr}^2(S^{+3})}{\mathrm{tr}(S^{+4})\,\mathrm{tr}(S^{+3})} \leq \frac{1}{\rho_3} \leq (p\wedge m)^2 \tag{A.26}$$

according again to (A.12) in Lemma A.2. Therefore, thanks to (A.26) and using the fact that $t''(\cdot) \geq 0$, the upper bound for $H_2(z,u)$ in (A.25) is non-positive if Condition (*iii*) in Proposition 4.1 holds. Since we have $H_1(z,u) \leq 0$, Condition (A.22) is satisfied, and hence, Condition (32) is satisfied as well, which is the sufficient domination condition of $\hat{\Sigma}_{a_o,\Psi}$ over $\hat{\Sigma}_{a_o}$ according to Corollary 3.1. □

A.3. Risk calculations and Stein–Haff type lemma

Here, we focus on risk calculations, optimal constant a_0 in (19) and proof of Theorem 2.1, and give the proof of Lemma 2.1 and Corollary 2.1.

Proof of Proposition 2.1. Let $\|A\|_F = \sqrt{\mathrm{tr}(A^\top A)}$ be the Frobenius norm associated to the inner product $\langle A, B\rangle = \mathrm{tr}(A^\top B)$ where A and B are $p\times p$ matrices. Then the Cauchy–Schwarz inequality expresses that

$$\mathrm{tr}^2\left(A^\top B\right) \leq \|A\|_F^2\,\|B\|_F^2\,. \tag{A.27}$$

First assume that

$$E_{\theta,\Sigma}\left[\|\Sigma^{-1}S\|_F^2\right] < \infty\,. \tag{A.28}$$

We will show that, for any $a>0$, the risk of $\hat{\Sigma}_a = a\,S$ is finite. Indeed the loss of $\hat{\Sigma}_a$ can be written as

$$L(\Sigma, a\,S) = \mathrm{tr}\left(a\,S\,\Sigma^{-1} - I_p\right)^2 = a^2\,\mathrm{tr}(\Sigma^{-1}S)^2 - 2\,a\,\mathrm{tr}(\Sigma^{-1}S) + p. \tag{A.29}$$

Then, noticing that $\mathrm{tr}^2(\Sigma^{-1}S) = \mathrm{tr}^2(\Sigma^{-1/2}S\,\Sigma^{-1/2})$, applying (A.27) with $A = I_p$ and $B = \Sigma^{-1/2}S\,\Sigma^{-1/2}$ gives

$$\mathrm{tr}^2(\Sigma^{-1}S) \leq p\,\|\Sigma^{-1/2}S\,\Sigma^{-1/2}\|_F^2 = p\,\mathrm{tr}\left(\Sigma^{-1}S\right)^2. \tag{A.30}$$

Now, with $A^\top = B = \Sigma^{-1}S$, applying (A.27) gives

$$\mathrm{tr}(\Sigma^{-1}S)^2 \leq \mathrm{tr}\left((\Sigma^{-1}S)^\top\,\Sigma^{-1}S\right) = \|\Sigma^{-1}S\|_F^2\,. \tag{A.31}$$

Taking expectation in (A.29), it follows from (A.31) that, according to (A.28), we have $E_{\theta,\Sigma}[\mathrm{tr}(\Sigma^{-1}S)^2] < \infty$, and hence according to (A.30), $E_{\theta,\Sigma}[\mathrm{tr}^2(\Sigma^{-1}S)] < \infty$, which implies that $E_{\theta,\Sigma}[\mathrm{tr}(\Sigma^{-1}S)] < \infty$. This is the announced result.

Secondly, we deal with the finiteness of the risk of $\hat{\Sigma}_{a,G} = a\,S + a\,SS^{+}$. Assume that

$$E_{\theta,\Sigma}\left[\|\Sigma^{-1}SS^{+}G\|_F^2\right] < \infty\,. \tag{A.32}$$

We will show that Conditions (A.28) and (A.32) insure that the risk of $\hat{\Sigma}_{a,G}$ is finite. Note that the loss of $\hat{\Sigma}_{a,G}$ is

$$L(\Sigma, \hat{\Sigma}_{a,G}) = L(\Sigma, a\,S) + a^2\,\mathrm{tr}(\Sigma^{-1}SS^{+}G)^2 + 2\,a\,\mathrm{tr}\left[\Sigma^{-1}SS^{+}G(a\,S\,\Sigma^{-1} - I_p)\right]. \tag{A.33}$$

As in (A.31), through the Cauchy–Schwarz inequality, we have

$$\mathrm{tr}(\Sigma^{-1}SS^{+}G)^2 \leq \|\Sigma^{-1}SS^{+}G\|_F^2,$$

and hence,

$$E_{\theta,\Sigma}\left[\mathrm{tr}(\Sigma^{-1}SS^{+}G)^2\right]\le E_{\theta,\Sigma}\left[\|\Sigma^{-1}SS^{+}G\|_F^2\right]<\infty \tag{A.34}$$

thanks to (A.32). Also by the Cauchy–Schwarz inequality associated to the inner product $\langle A,B\rangle = E_{\theta,\Sigma}\left[\mathrm{tr}\left(A^\top B\right)\right]$ it follows that

$$\left|E_{\theta,\Sigma}\left\{\mathrm{tr}\left[\Sigma^{-1}SS^{+}G\left(a\,S\,\Sigma^{-1}-I_p\right)\right]\right\}\right|\le\left(E_{\theta,\Sigma}\left[\|\Sigma^{-1}SS^{+}G\|_F^2\right]E_{\theta,\Sigma}\left[\|a\,S\,\Sigma^{-1}-I_p\|_F^2\right]\right)^{1/2}. \tag{A.35}$$

By assumption (A.34), the first term on the right-hand side of (A.35) is finite. Also it is clear that the second term is finite, expanding the squared Frobenius norm and using similar arguments than those used for the finiteness of the risk of $\hat{\Sigma}_a$. Finally, it follows that, taking expectation in (A.33), (A.28) and (A.32) are sufficient conditions for the risk $\hat{\Sigma}_{a,G}$ to be finite.

Lastly, under Conditions (A.28) and (A.32), the risk difference $\Delta(G)$ in (24) between $\hat{\Sigma}_{a,G}$ and $\hat{\Sigma}_a$ is finite and can be expressed as

$$\begin{aligned}\Delta(G)&=E_{\theta,\Sigma}\left[\mathrm{tr}\left(a(S+SS^{+}G)\,\Sigma^{-1}-I_p\right)^2\right]-E_{\theta,\Sigma}\left[\mathrm{tr}\left(a\,S\,\Sigma^{-1}-I_p\right)^2\right]\\&=a^2E_{\theta,\Sigma}\left[\mathrm{tr}\left(SS^{+}G\,\Sigma^{-1}SS^{+}G\Sigma^{-1}+2\,S\,\Sigma^{-1}SS^{+}G\,\Sigma^{-1}\right)\right]-2\,a\,E_{\theta,\Sigma}\left[\mathrm{tr}\left(SS^{+}G\,\Sigma^{-1}\right)\right]\\&=a^2E_{\theta,\Sigma}\left[\mathrm{tr}\left(\Sigma^{-1}SS^{+}(2\,S+G)\,\Sigma^{-1}SS^{+}G\right)\right]-2\,a\,E_{\theta,\Sigma}\left[\mathrm{tr}\left(\Sigma^{-1}SS^{+}G\right)\right].\quad\square\end{aligned}$$

In order to prove Lemma 2.1, we give, in the following lemma, the link between the differential expressions $\mathrm{tr}\left(U^\top\nabla_U\{G^\top S^+\}\right)$ and $\mathrm{tr}\left(SS^+\mathcal{D}_s\{SS^+G\}^\top\right)$, where $(\nabla_U)_{ij}=\partial/\partial U_{ij}$ for $i\in\{1,\dots,m\}$ and $j\in\{1,\dots,p\}$.

Lemma A.4. *For any $p\times p$ matrix function $G(z,s)$ weakly differentiable with respect to s, for any z, we have*

$$\mathrm{tr}\left(U^\top\nabla_U\{G^\top S^+\}\right)=\mathrm{tr}\left[2\,SS^+\mathcal{D}_s\{SS^+G\}^\top-((p\wedge m)+1)\,S^+G\right]. \tag{A.36}$$

Proof. Since $S^+=SS^+S^+$ we have

$$\mathrm{tr}\left(U^\top\nabla_U\{G^\top S^+\}\right)=\sum_{i,j,l}^{p}\sum_{k}^{m}U^\top_{ik}\frac{\partial}{\partial U_{kj}}\{(G^\top SS^+)_{jl}\,S^+_{li}\}=R_1+R_2$$

where

$$R_1=\sum_{i,j,l}^{p}\sum_{k}^{m}S^+_{li}\,U^\top_{ik}\frac{\partial(G^\top S\,S^+)_{jl}}{\partial U_{kj}}=\sum_{j,l}^{p}\sum_{k}^{m}(S^+U^\top)_{lk}\frac{\partial(G^\top SS^+)_{jl}}{\partial U_{kj}} \tag{A.37}$$

and

$$R_2=\sum_{i,j,l}^{p}\sum_{k}^{m}U^\top_{ik}\,(G^\top SS^+)_{jl}\frac{\partial S^+_{li}}{\partial U_{kj}}=\sum_{j}^{p}\sum_{k}^{m}\left(G^\top SS^+\frac{\partial S^+}{\partial U_{kj}}\,U^\top\right)_{jk}. \tag{A.38}$$

Dealing with R_1, we have, from the chain-rule,

$$\frac{\partial(G^\top SS^+)_{jl}}{\partial U_{kj}}=\sum_{r\ge q}^{p}\frac{\partial S_{rq}}{\partial U_{kj}}\,\frac{\partial(G^\top SS^+)_{jl}}{\partial S_{rq}}.$$

Note also that

$$\frac{\partial S_{qr}}{\partial U_{kj}}=\frac{\partial}{\partial U_{kj}}\sum_{o=1}^{m}U^\top_{qo}\,U_{or}=\sum_{o=1}^{p}\left(\frac{\partial U_{oq}}{\partial U_{kj}}\,U_{or}+U_{oq}\,\frac{\partial U_{or}}{\partial U_{kj}}\right)=U_{kr}\,\delta_{qj}+U_{kq}\,\delta_{rj}. \tag{A.39}$$

Then

$$\begin{aligned}\frac{\partial(G^\top SS^+)_{jl}}{\partial U_{kj}}&=\left(\sum_{r\le j}^{p}U_{kr}\,\frac{\partial}{\partial S_{rj}}+\sum_{q\ge j}^{p}U_{kq}\,\frac{\partial}{\partial S_{qj}}\right)(G^\top SS^+)_{jl}=\left(\sum_{r\le j}^{p}U_{kr}\,\frac{\partial}{\partial S_{rj}}+\sum_{r\ge j}^{p}U_{kr}\,\frac{\partial}{\partial S_{rj}}\right)(G^\top SS^+)_{jl}\\&=\left(\sum_{r}^{p}U_{kr}\,\frac{\partial}{\partial S_{rj}}+U_{kj}\,\frac{\partial}{\partial S_{jj}}\right)(G^\top SS^+)_{jl}=2\,\sum_{r}^{p}U_{kr}\,\frac{1}{2}(1+\delta_{rj})\frac{\partial}{\partial S_{rj}}\,(G^\top SS^+)_{jl}\\&=2\,\sum_{r}^{p}U_{kr}\,d^s_{rj}(G^\top SS^+)_{jl},\end{aligned}$$

by definition of the generic term d^s_{rj} of $\mathcal{D}_s\{\cdot\}$. Hence (A.37) becomes

$$R_1 = 2\sum_{l,j}^{p}\sum_{k}^{m}(S^+U^\top)_{lk}\sum_{r}^{p}U_{kr}\, d^s_{rj}(G^\top SS^+)_{jl} = 2\,\mathrm{tr}\left(SS^+\mathcal{D}_s\{SS^+G\}^\top\right). \tag{A.40}$$

As for R_2 in (A.38), using Proposition A.2 with $A = G^\top SS^+$ and $B = U^\top$, we have

$$R_2 = \sum_{j}^{p}\sum_{k}^{m}\left((G^\top SS^+)(I - SS^+)\right)_{jj}(U\,SS^+U^\top)_{kk} + \sum_{j}^{p}\sum_{k}^{m}\left((I - SS^+)U^\top\right)_{jk}\left(U\,(G^\top S^+S^+)^\top\right)_{kj}$$
$$- \sum_{j}^{p}\sum_{k}^{m}\left((G^\top S^+)_{jj}(U\,S^+U^\top)_{kk} + (S^+U^\top)_{jk}(U\,S^+G)_{kj}\right).$$

Since $SS^+(I_p - SS^+) = S\,(I_p - SS^+) = 0$, $(I_p - SS^+)\,U^\top = 0$, $\mathrm{tr}(SS^+) = p\wedge m$, and $S = U^\top U$. Thus R_2 simplifies in

$$R_2 = -(p\wedge m + 1)\,\mathrm{tr}\left(S^+G\right). \tag{A.41}$$

Combining (A.40) and (A.41) gives the desired result. □

Proof of Lemma 2.1. A Stein–Haff type identity, in the elliptical framework and for both singular and non-singular cases, was given by Fourdrinier Mezoued and Wells [12] as follows:

$$E_{\theta,\Sigma}\left[\mathrm{tr}\left(\Sigma^{-1}S\,H\right)\right] = K^*E^*_{\theta,\Sigma}\left[m\,\mathrm{tr}(H) + \mathrm{tr}\left(U^\top\nabla_U H^\top\right)\right].$$

These authors remark that an equivalent expression of this identity is

$$E_{\theta,\Sigma}\left[\mathrm{tr}\left(\Sigma^{-1}S\,H\right)\right] = K^*E^*_{\theta,\Sigma}\left[m\,\mathrm{tr}(SS^+H) + \mathrm{tr}\left(U^\top\nabla_U H^\top SS^+\right)\right],$$

since $S = SS^+S$. For $H = S^+G$ this identity becomes

$$E_{\theta,\Sigma}\left[\mathrm{tr}\left(\Sigma^{-1}SS^+G\right)\right] = K^*E^*_{\theta,\Sigma}\left[m\,\mathrm{tr}(S^+G) + \mathrm{tr}\left(U^\top\nabla_U G^\top S^+\right)\right]. \tag{A.42}$$

Therefore, we deduce the Stein–Haff type identity in (26) through the Haff operator $\mathcal{D}_s$ by replacing (A.36) in (A.42).

Proof of Corollary 2.1. We first apply Lemma 2.1 taking $V\Sigma^{-1}SS^+G$ instead of G. Thus

$$E_{\theta,\Sigma}\left[\mathrm{tr}\left(\Sigma^{-1}SS^+V\Sigma^{-1}SS^+G\right)\right] = K^*E^*_{\theta,\Sigma}\left[\mathrm{tr}\left[2SS^+\mathcal{D}_s\{G^\top SS^+\Sigma^{-1}V^\top SS^+\} + (m - (p\wedge m) - 1)\Sigma^{-1}SS^+GS^+V\right]\right]. \tag{A.43}$$

Secondly, according to (A.10), the first integrand term on the right-hand side of (A.43) can be rewritten as

$$\mathrm{tr}\left[SS^+\mathcal{D}_s\{G^\top SS^+\Sigma^{-1}V^\top SS^+\}\right] = \mathrm{tr}\left[\Sigma^{-1}V^\top SS^+\mathcal{D}_s\{G^\top SS^+\}\right]\mathrm{tr}\left[SS^+\left(\Sigma^{-1}SS^+G\mathcal{D}_s\right)^\top\{V^\top SS^+\}\right]. \tag{A.44}$$

Now, applying (A.11) to the second term on the right-hand side of (A.44) gives

$$\mathrm{tr}\left[SS^+\left(\Sigma^{-1}SS^+G\,\mathcal{D}_s\right)^\top\{V^\top SS^+\}\right] = \mathrm{tr}\left[\Sigma^{-1}SS^+G\,(SS^+\mathcal{D}_s)^\top\{SS^+V\}\right].$$

Furthermore, using the fact that $SS^+ = SS^+SS^+$, it is clear that from (A.10) in Proposition A.1 that

$$(SS^+\mathcal{D}_s)^\top\{SS^+V\} = \mathcal{D}_s\{SS^+V\} - \mathcal{D}_s\{SS^+\}SS^+V.$$

Hence, according to (A.64) and using the fact that $(I_p - SS^+)S = 0_{p\times p}$ we have

$$(SS^+\mathcal{D}_s)^\top\{SS^+V\} = \mathcal{D}_s\{SS^+V\} - \frac{1}{2}(p - p\wedge m)S^+V.$$

Therefore

$$\mathrm{tr}\left[SS^+\left(\Sigma^{-1}SS^+G\mathcal{D}_s\right)^\top\{V^\top S^+S\}\right] = \mathrm{tr}\left[\Sigma^{-1}SS^+G\left(\mathcal{D}_s\{SS^+V\} - \frac{1}{2}(p - p\wedge m)S^+V\right)\right]. \tag{A.45}$$

Now, combining (A.43), (A.44) and (A.45) and using the symmetry of SS^+V gives

$$E_{\theta,\Sigma}\left[\mathrm{tr}\left(\Sigma^{-1}SS^+V\Sigma^{-1}SS^+G\right)\right] = K^*E^*_{\theta,\Sigma}\left[\mathrm{tr}\left(\Sigma^{-1}SS^+T^*\right)\right], \tag{A.46}$$

where

$$T^* = 2\left[SS^+V\,\mathcal{D}_s\{SS^+G\}^\top + SS^+G\mathcal{D}_s\{SS^+V\}\right] - (p - m + 1)GS^+V.$$

Finally, applying Lemma 2.1 to (A.46) with T^* instead of G gives

$$E^*_{\theta,\Sigma}\left[\mathrm{tr}\left(\Sigma^{-1}SS^+T^*\right)\right] = K^{**}E^{**}_{\theta,\Sigma}\left[\mathrm{tr}\left(2S^+S\,\mathcal{D}_s\{SS^+T^*\}^\top - (m - (p\wedge m) - 1)S^+T^*\right)\right]. \quad \square$$

Determination of the optimal constant a_0. Let $\hat{\Sigma}_a = a\,S$, where a is a positive constant. Then, according to (17), the risk at Σ of $\hat{\Sigma}_a$ is

$$R(\Sigma,\,\hat{\Sigma}_a) = E_{\theta,\Sigma}\left[\operatorname{tr}\left(aS\Sigma^{-1} - I_p\right)^2\right] = a^2\,E_{\theta,\Sigma}\left[\operatorname{tr}\left(\Sigma^{-1}SS^{+}S\right)^2\right] - 2\,a\,E_{\theta,\Sigma}\left[\operatorname{tr}\left(\Sigma^{-1}SS^{+}S\right)\right] + p, \tag{A.47}$$

since $S = SS^{+}S$. Setting $G = V = S$ in Corollary 2.1 to deal with the first term on the right-hand side of (A.47), and in Lemma 2.1 to deal with the second term, we have

$$\begin{aligned} R(\Sigma,\,\hat{\Sigma}_a) = {} & a^2K^*K^{**}\,E^{**}_{\theta,\Sigma}\left[\operatorname{tr}\left[2\,S^{+}S\,\mathcal{D}_s\{SS^{+}T^*\}^{\top} + (m-(p\wedge m)-1)\,S^{+}T^*\right]\right] \\ & - 2\,a\,K^*E^{*}_{\theta,\Sigma}\left[\operatorname{tr}\left[2\,S^{+}S\,\mathcal{D}_s\{S\} + (m-(p\wedge m)-1)\,S^{+}S\right]\right] + p\,, \end{aligned} \tag{A.48}$$

where

$$T^* = 4\,S\,\mathcal{D}_s\{S\} - (p-m+1)\,S.$$

Furthermore, thanks to (A.63) and (A.64) and using the fact that $S(I_p - SS^{+}) = 0_{p\times p}$, the integrand terms on the right-hand side of (A.48) are evaluated as

$$\operatorname{tr}\left[2\,S^{+}S\,\mathcal{D}_s\{SS^{+}T^*\}^{\top} + (m-(p\wedge m)-1)\,S^{+}T^*\right] = (p+m+1)\,(p+m-p\wedge m)\operatorname{tr}(SS^{+}) \tag{A.49}$$

and

$$\operatorname{tr}\left(2\,S^{+}S\,\mathcal{D}_s\{S\} + (m-(p\wedge m)-1)\,S^{+}S\right) = (p+m-p\wedge m)\operatorname{tr}(S^{+}S) \tag{A.50}$$

Substituting (A.49) and (A.50) in (A.48) gives

$$R(\Sigma,\,\hat{\Sigma}_a) = a^2K^*K^{**}\,mp\,(p+m+1) - 2\,a\,K^*mp + p\,,$$

since $\operatorname{tr}(S^{+}S) = p\wedge m$. Clearly, $a_o = 1/K^{**}(p+m+1)$ gives rise to the optimal constant a under the risk (17). □

Proof of Theorem 2.1. Expressing the risk difference in (28) only through the $E_{\theta,\Sigma}$-expectation, thanks to (A.4) and (A.5), we have

$$\begin{aligned} \Delta(G) = a_o^2\,K^*E_{\theta,\Sigma}\Big[\,\varphi^*_{\theta,\Sigma}(Z,U)\varphi^{**}_{\theta,\Sigma}(Z,U)\,\Big(\operatorname{tr}\left[2\,S^{+}S\,\mathcal{D}_s\{SS^{+}T^*\}^{\top} - (m-(p\wedge m)-1)\,S^{+}T^*\right] \\ - \frac{2}{a_o\,K^*\,\varphi^{**}_{\theta,\Sigma}(Z,U)}\operatorname{tr}\left[2\,S^{+}S\,\mathcal{D}_s\{SS^{+}G\} - (m-(p\wedge m)-1)\,S^{+}G\right]\Big)\Big]\,, \end{aligned} \tag{A.51}$$

where a_o is given in (19). Using Inequality (A.8) in Lemma A.1 and also noting that $K^{**} \geq c^2$, it can be shown that $K^{**}/K^*\varphi^{**}_{\theta,\Sigma}(Z,U) \geq c^2/b^2$. Therefore, under Condition (21),

$$\begin{aligned} \Delta(G) \leq {} & a^2\,K^*E_{\theta,\Sigma}\Big[\,\varphi^*_{\theta,\Sigma}(Z,U)\,\varphi^{**}_{\theta,\Sigma}(Z,U)\Big(\operatorname{tr}\left[2\,S^{+}S\,\mathcal{D}_s\{SS^{+}T^*\}^{\top} - (m-(p\wedge m)-1)\,S^{+}T^*\right] \\ & - 2\,(p+m+1)\frac{c^2}{b^2}\operatorname{tr}\left[2\,S^{+}S\,\mathcal{D}_s\{SS^{+}G\} - (m-(p\wedge m)-1)\,S^{+}G\right]\Big)\Big]\,. \end{aligned}$$

Then, the estimator $\hat{\Sigma}_{a_o,G}$ dominates $\hat{\Sigma}_{a_o}$ if Inequality (22) holds, since $\varphi^*_{\theta,\Sigma}(Z,U)$ and $\varphi^{**}_{\theta,\Sigma}(Z,U)$ are non-negative functions. □

A.4. Differential expressions

The following proposition is used in the proof of Lemma 2.1 given in Appendix A.3.

Proposition A.2. *Recall that $S = U^{\top}U$ is a $p\times p$ symmetric matrix. For a non null integer N, let A be an $N\times p$ matrix and B a $p\times N$ matrix. Then*

$$\begin{aligned} \left(A\,\frac{\partial S^{+}}{\partial U_{kj}}\,B\right)_{li} = {} & \left(A\,(I_p - SS^{+})\right)_{lj}\,\left(US^{+}S^{+}B\right)_{ki} + \left((I_p - SS^{+})\,B\right)_{ji}\,\left(U\,S^{+}S^{+}A^{\top}\right)_{kl} - \left(A\,S^{+}\right)_{lj}\,\left(U\,S^{+}B\right)_{ki} \\ & - \left(S^{+}B\right)_{ji}\,\left(U\,S^{+}A^{\top}\right)_{kl}\,. \end{aligned}$$

Proof. Applying a well known result about the derivatives of the Moore–Penrose inverses S^{+} (see Harville [18]), we have for any $1\leq k\leq m$ and any $1\leq j\leq p$,

$$\frac{\partial S^{+}}{\partial U_{kj}} = -\,S^{+}\,\frac{\partial S}{\partial U_{kj}}\,S^{+} + (I_p - SS^{+})\,\frac{\partial S}{\partial U_{kj}}\,S^{+}S^{+} + S^{+}S^{+}\,\frac{\partial S}{\partial U_{kj}}\,(I_p - SS^{+}),$$

and then

$$\left(A\,\frac{\partial S^{+}}{\partial U_{kj}}\,B\right)_{li} = -\left(A\,S^{+}\,\frac{\partial S}{\partial U_{kj}}\,S^{+}\,B\right)_{li} + \left(A\,(I_p - SS^{+})\,\frac{\partial S}{\partial U_{kj}}\,S^{+}\,S^{+}\,B\right)_{li} + \left(A\,S^{+}S^{+}\,\frac{\partial S}{\partial U_{kj}}\,(I_p - SS^{+})\,B\right)_{li}. \tag{A.52}$$

Thanks to (A.39), the first term on the right-hand side of (A.52) becomes

$$\left(A\,S^{+}\,\frac{\partial S}{\partial U_{kj}}\,S^{+}\,B\right)_{li} = \sum_{q,r}^{p}(A\,S^{+})_{lq}\left(\frac{\partial S}{\partial U_{kj}}\right)_{qr}(S^{+}\,B)_{ri}$$
$$= \sum_{r}^{m}(A\,S^{+})_{lj}\,(S^{+}\,B)_{ri}\,U_{kr} + \sum_{q=1}^{m}(A\,S^{+})_{lq}\,(S^{+}B)_{ji}\,U_{kq} = (A\,S^{+})_{lj}\,(U\,S^{+}\,B)_{ki} + (S^{+}\,B)_{ji}\,(U\,S^{+}\,A^{\top})_{kl}. \tag{A.53}$$

Similarly, the two other terms on the right-hand side of (A.52) become respectively

$$\left(A\,(I_p - SS^{+})\,\frac{\partial S}{\partial U_{kj}}\,S^{+}S^{+}\,B\right)_{li} = \left(A\,(I_p - SS^{+})\right)_{lj}\,(U\,S^{+}S^{+}\,B)_{ki} + (S^{+}S^{+}\,B)_{ji}\,\left(U\,\left(I_p - SS^{+}\right)\,A^{\top}\right)_{kl} \tag{A.54}$$

and

$$\left(A\,S^{+}S^{+}\,\frac{\partial S}{\partial U_{kj}}\,(I_p - SS^{+})\,B\right)_{li} = (A\,S^{+}S^{+})_{lj}\,\left(U\,(I_p - SS^{+})\,B\right)_{ki} + \left((I_p - SS^{+})\,B\right)_{ji}\,(U\,S^{+}S^{+}\,A^{\top})_{kl}\,. \tag{A.55}$$

Hence, replacing (A.53), (A.54) and (A.55) in (A.52) and using the fact that $U(I_p - SS^{+}) = 0_{(p\wedge m)\times p}$, give the desired result. □

The following lemma evaluates the derivatives of eigenvalues and eigenvectors of S and is used in Lemma A.6.

Lemma A.5. *Let $1 \le i, j, a, r \le p$ and $1 \le k \le p \wedge m$. Then*

$$d^{s}_{ij} l_k = (H_1)_{ik}\,(H_1)_{jk} \tag{A.56}$$

and

$$(d^{s}_{ij} H_1)_{ak} = \frac{1}{2}\sum_{r\neq k}^{p\wedge m}\frac{(H_1)_{ar}}{l_k - l_r}\left[(H_1)_{ir}(H_1)_{jk} + (H_1)_{jr}(H_1)_{ik}\right] + \frac{1}{2\,l_k}\left[(I_p - H_1H_1^{\top})_{ai}(H_1)_{jk} + (I_p - H_1H_1^{\top})_{aj}(H_1)_{ik}\right]. \tag{A.57}$$

Proof. Recall that $S = H_1 L H_1^{\top}$ where $H_1 \in \mathcal{L}_{p\times(p\wedge m)}$ and $L \in \mathcal{D}_{(p\wedge m)\times(p\wedge m)}$. Take $H_2 \in \mathcal{L}_{p\times(p-(p\wedge m))}$ such that $H_2^{\top} H_1 = 0_{(p-p\wedge m)\times(p\wedge m)}$ to form $H = [H_1, H_2] \in \mathcal{O}_p$. Note that, in the invertible case, $H = H_1 \in \mathcal{O}_p$ and there is no H_2 to complete H_1.

For any differential operator of S, we have

$$\mathrm{d}S = (\mathrm{d}H_1)\,L\,H_1^{\top} + H_1\,(\mathrm{d}L)\,H_1^{\top} + H_1\,L\,(\mathrm{d}H_1^{\top}),$$

which yields

$$H^{\top}\,(\mathrm{d}S)\,H_1 = \begin{bmatrix} H_1^{\top}\,(\mathrm{d}S)\,H_1 \\ H_2^{\top}\,(\mathrm{d}S)\,H_1 \end{bmatrix} = \begin{bmatrix} H_1^{\top}\,(\mathrm{d}H_1)\,L + H_1^{\top}\,H_1\,(\mathrm{d}L) + H_1^{\top}\,H_1\,L\,(\mathrm{d}H_1^{\top})\,H_1 \\ H_2^{\top}\,(\mathrm{d}H_1)\,L + H_2^{\top}\,H_1(\mathrm{d}L) + H_2^{\top}\,H_1\,L\,(\mathrm{d}H_1^{\top})\,H_1 \end{bmatrix}.$$

The differential expression of $H_1^{\top} H_1 = I_{p\wedge m}$ gives that $(\mathrm{d}H_1^{\top})H_1 = -H_1^{\top}\,(\mathrm{d}H_1)$. Then

$$\begin{bmatrix} H_1^{\top}\,(\mathrm{d}S)\,H_1 \\ H_2^{\top}\,(\mathrm{d}S)\,H_1 \end{bmatrix} = \begin{bmatrix} H_1^{\top}\,(\mathrm{d}H_1)\,L - L\,H_1^{\top}\,(\mathrm{d}H_1) + (\mathrm{d}L) \\ H_2^{\top}\,(\mathrm{d}H_1)\,L \end{bmatrix}$$

since $H_2^{\top}\,H_1 = 0_{(p-p\wedge m)\times(p\wedge m)}$. Hence, for $1 \le k \le p\wedge m$

$$(\mathrm{d}L)_k = \left\{H_1^{\top}\,(\mathrm{d}S)\,H_1\right\}_{kk} - \{H_1^{\top}\,(\mathrm{d}H_1)\,L\}_{kk} + \{L\,H_1^{\top}\,(\mathrm{d}H_1)\}_{kk} = \left\{H_1^{\top}\,(\mathrm{d}S)\,H_1\right\}_{kk}, \tag{A.58}$$

since $\{L\,H_1^{\top}\,(\mathrm{d}H_1)\}_{kk} - \{H_1^{\top}\,(\mathrm{d}H_1)\,L\}_{kk} = 0$.

Now, for $r \neq k$, we have

$$\{H_1^{\top}\,(\mathrm{d}S)\,H_1\}_{rk} = \{H_1^{\top}\,(\mathrm{d}H_1)\}_{rk}\,l_k - l_r\,\{H_1^{\top}\,(\mathrm{d}H_1)\}_{rk} \quad \text{for } r, k \in \{1, \ldots, p\wedge m\}$$
$$\{H_2^{\top}\,(\mathrm{d}S)\,H_1\}_{rk} = \{H_2^{\top}\,(\mathrm{d}H_1)\}_{rk}\,l_k \quad \text{for } r = p\wedge m + 1, \ldots, p \text{ and } k \in \{1, \ldots, p\wedge m\}$$

Therefore,

$$(H_1^{\top}\,(\mathrm{d}H_1))_{rk} = \frac{1}{l_k - l_r}\,\{H_1^{\top}\,(\mathrm{d}S)\,H_1\}_{rk} \quad \text{for } r, k \in \{1, \ldots, p\wedge m\}$$
$$(H_2^{\top}\,(\mathrm{d}H_1))_{rk} = \frac{1}{l_k}\{H_2^{\top}\,(\mathrm{d}S)\,H_1\}_{rk} \quad \text{for } r \in \{p\wedge m + 1, \ldots, p\} \text{ and } k \in \{1, \ldots, p\wedge m\}. = \tag{A.59}$$

Now, if "d" is the Haff operator $\mathcal{D}_s$ in (20), we have $(d^s_{ij}S)_{cd} = \left(\delta_{ic}\delta_{jd} + \delta_{id}\delta_{jc}\right)/2$. Then

$$\{H_1^\top\,(d^s_{ij}S)\,H_1\}_{rk} \;=\; \frac{1}{2}\left[(H_1)_{ir}(H_1)_{jk} + (H_1)_{jr}(H_1)_{ik}\right] \quad \text{and} \quad \{H_2^\top\,(d^s_{ij}S)\,H_1\}_{rk} = \frac{1}{2}\left[(H_2)_{ir}(H_1)_{jk} + (H_2)_{jr}(H_1)_{ik}\right] \tag{A.60}$$

Hence, using (A.60), Equality (A.58) becomes $d^s_{ij}l_k = (H_1)_{ik}(H_1)_{jk}$, which is (A.56). Dealing with (A.57), note that, since $H\,H^\top = I_p$,

$$\left(d^s_{ij}H_1\right)_{ak} = \{HH^\top(d^s_{ij}\,H_1)\}_{ak} = \sum_{r\neq k}^{p\wedge m}(H_1)_{ar}\{H_1^\top(d^s_{ij}H_1)\}_{rk} + \sum_{r=(p\wedge m)+1}^{p}(H_2)_{ar}\{H_2^\top(d^s_{ij}H_1)\}_{rk}\,. \tag{A.61}$$

Combining (A.59), (A.60) we have, for $r\neq k$,

$$(H_1^\top\,(d^s_{ij}H_1))_{rk} = \frac{(H_1)_{ir}\,(H_1)_{jk} + (H_1)_{jr}(H_1)_{ik}}{2\,(l_k - l_r)} \quad \text{and} \quad (H_2^\top\,(d^s_{ij}H_1))_{rk} = \frac{(H_2)_{ir}\,(H_1)_{jk} + (H_2)_{jr}(H_1)_{ik}}{2\,l_k}.$$

Therefore (A.61) becomes

$$\left(d^s_{ij}H_1\right)_{ak} = \frac{1}{2}\sum_{r\neq k}^{p\wedge m}\frac{(H_1)_{ar}}{l_k - l_r}\left[(H_1)_{ir}(H_1)_{jk} + (H_1)_{jr}(H_1)_{ik}\right] + \frac{1}{2l_k}\left[(H_2\,H_2^\top)_{ai}(H_1)_{jk} + (H_2H_2^\top)_{aj}(H_1)_{ik}\right],$$

which gives the desired result in (A.57) since $H_2\,H_2^\top = I_p - H_1\,H_1^\top$. □

The following lemma is a generalization of the results of Haff [17] to both invertible and non-invertible cases.

Lemma A.6. *Under the notation of the proof of Lemma A.5, let $G = H_1\Psi H_1^\top$ where $\Psi\in\mathcal{D}_{p\wedge m}$. Then $SS^+G = G$ and*

$$\mathcal{D}_s\{SS^+\,G\} = \mathcal{D}_s\{H_1\Psi H_1^\top\} = H_1\Psi^{(1)}H_1^\top + \frac{1}{2}\mathrm{tr}(L^{-1}\Psi)(I_p - H_1H_1^\top)\,, \tag{A.62}$$

where $\Psi^{(1)}\in\mathcal{D}_{p\wedge m}$ with

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

Furthermore

$$\mathcal{D}_s\{S\} = \mathcal{D}_s\{H_1LH_1^\top\} = \frac{1}{2}(p+1)H_1H_1^\top + \frac{1}{2}(p\wedge m)(I_p - H_1H_1^\top)\,, \tag{A.63}$$

$$\mathcal{D}_s\{SS^+\} = \mathcal{D}_s\{H_1H_1^\top\} = \frac{1}{2}(p - p\wedge m)H_1\,L^{-1}\,H_1^\top + \frac{1}{2}\mathrm{tr}(L^{-1})(I_p - H_1H_1^\top) \tag{A.64}$$

and

$$(SS^+\,\mathcal{D}_s)^\top\{SS^+G\} = (H_1H_1^\top\mathcal{D}_s)^\top\{H_1\Psi H_1^\top\} = \mathcal{D}_s\{H_1\Psi H_1^\top\} - \frac{1}{2}(p-(p\wedge m))(H_1L^{-1}\Psi H_1^\top)\,. \tag{A.65}$$

Proof. The (i,a) elements of the matrix $\mathcal{D}_s\{H_1\Psi H_1^\top\}$ are expressed as

$$(\mathcal{D}_s\{H_1\Psi H_1^\top\})_{ia} = \sum_{j}^{p}\sum_{b}^{p\wedge m} d^s_{ij}\{(H_1)_{jb}\psi_b(H_1)_{ab}\} = A_{ia} + B_{ia} + C_{ia},$$

where

$$A_{ia} = \sum_{j}^{p}\sum_{b}^{p\wedge m}\psi_b(H_1)_{ab}(d^s_{ij}H_1)_{jb}\,,\quad B_{ia} = \sum_{j}^{p}\sum_{b}^{p\wedge m}\psi_b(H_1)_{jb}(d^s_{ij}H_1)_{ab} \quad\text{and}\quad C_{ia} = \sum_{j}^{p}\sum_{b}^{p\wedge m}(H_1)_{jb}(H_1)_{ab}(d^s_{ij}\psi)_b\,. \tag{A.66}$$

First, we deal with A_{ia} and B_{ia}. Thanks to (A.57), we obtain

$$A_{ia} = \frac{1}{2}\sum_{j}^{p}\sum_{b}^{p\wedge m}\sum_{r\neq b}^{p\wedge m}\frac{\psi_b\,(H_1)_{ab}(H_1)_{jr}}{l_b - l_r}\left[(H_1)_{ir}(H_1)_{jb} + (H_1)_{jr}(H_1)_{ib}\right]$$
$$+ \frac{1}{2}\sum_{j}^{p}\sum_{b}^{p\wedge m}\frac{\psi_b\,(H_1)_{ab}}{l_b}\left[(I_p - H_1H_1^\top)_{ji}(H_1)_{jb} + (I_p - H_1H_1^\top)_{jj}(H_1)_{ib}\right]$$

and

$$B_{ia} = \frac{1}{2}\sum_{j}^{p}\sum_{b}^{p\wedge m}\sum_{r\neq b}^{p\wedge m}\frac{\psi_b\,(H_1)_{jb}(H_1)_{ar}}{l_b - l_r}\left[(H_1)_{ir}(H_1)_{jb} + (H_1)_{jr}(H_1)_{ib}\right]$$
$$+\frac{1}{2}\sum_{j}^{p}\sum_{b}^{p\wedge m}\frac{\psi_b\,(H_1)_{jb}}{l_b}\left[(I_p - H_1H_1^\top)_{ai}(H_1)_{jb} + (I_p - H_1H_1^\top)_{aj}(H_1)_{ik}\right].$$

Summing on j, we have

$$A_{ia} = \frac{1}{2}\sum_{b}^{p\wedge m}\sum_{r\neq b}^{p\wedge m}\frac{\psi_b(H_1)_{ab}}{l_b - l_r}\left[(H_1)_{ir}(H_1^\top H_1)_{rb} + (H_1^\top H_1)_{rr}(H_1)_{ib}\right]$$
$$+\frac{1}{2}\sum_{b}^{p\wedge m}\frac{\psi_b\,(H_1)_{ab}}{l_b}\left[(H_1^\top(I_p - H_1H_1^\top))_{bi} + \mathrm{tr}(I_p - H_1H_1^\top)(H_1)_{ib}\right]$$

and

$$B_{ia} = \frac{1}{2}\sum_{b}^{p\wedge m}\sum_{r\neq b}^{p\wedge m}\frac{\psi_b(H_1)_{ar}}{l_b - l_r}\left[(H_1)_{ir}(H_1^\top H_1)_{bb} + (H_1^\top H_1)_{br}(H_1)_{ib}\right]$$
$$+\frac{1}{2}\sum_{b}^{p\wedge m}\frac{\psi_b}{l_b}\left[(I_p - H_1H_1^\top)_{ai}(H_1^\top H_1)_{bb} + ((I_p - H_1H_1^\top)H_1)_{ab}(H_1)_{ik}\right].$$

Now, using the fact that $(H_1\,H_1^\top)_{rb} = \delta_{rb}$, we have

$$A_{ia} = \frac{1}{2}\sum_{b}^{p\wedge m}\sum_{r\neq b}^{p\wedge m}\frac{\psi_b(H_1)_{ab}}{l_b - l_r}(H_1)_{ib} + \frac{1}{2}\sum_{b}^{p\wedge m}\frac{\psi_b\,(H_1)_{ab}}{l_b}\left[(H_1^\top(I_p - H_1H_1^\top))_{bi} + \mathrm{tr}(I_p - H_1H_1^\top)(H_1)_{ib}\right]$$

and

$$B_{ia} = \frac{1}{2}\sum_{b}^{p\wedge m}\sum_{r\neq b}^{p\wedge m}\frac{\psi_b(H_1)_{ar}}{l_b - l_r}(H_1)_{ir} + \frac{1}{2}\sum_{b}^{p\wedge m}\frac{\psi_b}{l_b}\left[(I_p - H_1H_1^\top)_{ai} + ((I_p - H_1H_1^\top)H_1)_{ab}(H_1)_{ik}\right].$$

Since $H_1^\top H_1 = I_{p\wedge m}$, we have $H_1^\top(I_p - H_1H_1^\top) = \left((I_p - H_1H_1^\top)H_1\right)^\top = 0_{(p\wedge m)\times p}$ and $\mathrm{tr}(I_p - H_1H_1^\top) = p - (p\wedge m)$. Hence

$$A_{ia} = \frac{1}{2}\sum_{b}^{p\wedge m}(H_1)_{ib}\sum_{r\neq b}^{p\wedge m}\frac{\psi_b}{l_b - l_r}(H_1^\top)_{ba} + \frac{1}{2}(p - (p\wedge m))\sum_{b}^{p\wedge m}(H_1)_{ib}\frac{\psi_b}{l_b}(H_1^\top)_{ba} \tag{A.67}$$

and

$$B_{ia} = \frac{1}{2}\sum_{b}^{p\wedge m}\sum_{r\neq b}^{p\wedge m}\frac{\psi_b(H_1)_{ar}(H_1)_{ir}}{l_b - l_r} + \frac{1}{2}\mathrm{tr}(L^{-1}\Psi)(I_p - H_1H_1^\top)_{ia}.$$

The expression of B_{ia} can be specified noticing that

$$\sum_{b}^{p\wedge m}\sum_{r\neq b}^{p\wedge m}\frac{\psi_b(H_1)_{ar}(H_1)_{ir}}{l_b - l_r} = -\sum_{b}^{p\wedge m}\sum_{r\neq b}^{p\wedge m}\frac{\psi_r(H_1)_{ab}(H_1)_{ib}}{l_b - l_r},$$

That is,

$$B_{ia} = -\frac{1}{2}\sum_{b}^{p\wedge m}(H_1)_{ib}\sum_{b\neq r}^{p\wedge m}\frac{\psi_r}{l_b - l_r}(H_1^\top)_{ba} + \frac{1}{2}\mathrm{tr}(L^{-1}\Psi)(I_p - H_1H_1^\top)_{ia}. \tag{A.68}$$

Dealing with C_{ia} in (A.66), we have from the chain-rule

$$C_{ia} = \sum_{j}^{p}\sum_{b}^{p\wedge m}(H_1)_{jb}(H_1)_{ab}\sum_{q}^{p\wedge m}\frac{\partial\psi_b}{\partial l_q}d_{ij}^{s}l_q.$$

Then, using (A.56), we have

$$C_{ia} = \sum_{b}^{p\wedge m} (H_1)_{ab} \sum_{q}^{p\wedge m} \frac{\partial \psi_b}{\partial l_q} (H_1)_{iq} (H_1 H_1^\top)_{qb}\,, \quad \text{and hence,} \quad C_{ia} = \sum_{b}^{p\wedge m} (H_1)_{ib} \frac{\partial \psi_b}{\partial l_b} (H_1^\top)_{ba}\,, \tag{A.69}$$

since $H_1^\top H_1 = I_{p\wedge m}$. Finally, gathering expressions (A.67), (A.68) and (A.69) gives the result in (A.62).

Now, in order to prove (A.63) (respectively (A.64)) we apply (A.62) for $\Psi = L$ (respectively $\Psi = I_{p\wedge m}$), which gives the desired result since $\Psi^{(1)} = I_{p\wedge m}(p+1)/2$ (respectively $\Psi^{(1)} = (p - p\wedge m)L^{-1}$). As for (A.65), we have from (A.10)

$$\mathcal{D}_s\{H_1 \Psi H_1\} = \mathcal{D}_s\{H_1 H_1^\top H_1 \Psi H_1\} = \mathcal{D}_s\{H_1 H_1^\top\} H_1 \Psi H_1 + (H_1 H_1^\top \mathcal{D}_s)^\top \{H_1 \Psi H_1^\top\}.$$

Then, according to (A.64), we have

$$(H_1 H_1^\top \mathcal{D}_s)^\top \{H_1 \Psi H_1^\top\} = \mathcal{D}_s\{H_1 \Psi H_1\} - \frac{1}{2}(p - p\wedge m) H_1 L^{-1} H_1^\top. \quad \square$$

References

[1] J. Berger, Minimax estimation of location vectors for a wide class of densities, Ann. Statist. 3 (1975) 1318–1328.
[2] E. Candès, B. Recht, Exact matrix completion via convex optimization, Found. Comput. Math. 9 (6) (2009) 717.
[3] E. Candès, C. Sing-Long, J.D. Trzasko, Unbiased risk estimates for singular value thresholding and spectral estimators, IEEE Trans. Signal Process. 61 (2013) 4643–4657.
[4] S. Canu, D. Fourdrinier, Unbiased risk estimates for matrix estimation in the elliptical case, J. Multivariate Anal. 158 (2017) 60–72.
[5] Y. Chen, A. Wiesel, A. Hero, Shrinkage estimation of high dimensional covariance matrices, in: 2009 IEEE International Conference on Acoustics, Speech and Signal Process., 2009, pp. 2937–2940.
[6] D. Chételat, M. Wells, Improved multivariate normal mean estimation with unknown covariance when p is greater than n, Ann. Statist. 40 (6) (2012) 3137–3160.
[7] D. Dey, C. Srinivasan, Estimation of a covariance matrix under Stein's loss, Ann. Statist. 13 (1985) 1581–1591.
[8] B. Efron, C. Morris, Multivariate empirical Bayes and estimation of covariance matrices, Ann. Statist. 4 (1976) 22–32.
[9] K. Fang, Y. Zhang, Generalized Multivariate Analysis, Science Press, Springer-Verlag, Beijing, 1990.
[10] T. Fisher, X. Sun, Improved Stein-type shrinkage estimators for the high-dimensional multivariate normal covariance matrix, Comput. Statist. Data Anal. 55 (2011) 1909–1918.
[11] D. Fourdrinier, F. Mezoued, W. Strawderman, Bayes Minimax estimators of a location vector for densities in the Berger class, Electron. J. Statist. 6 (2012) 783–809.
[12] D. Fourdrinier, F. Mezoued, M. Wells, Estimation of the inverse scatter matrix of an elliptically symmetric distribution, J. Multivariate Anal. 143 (2016) 32–55.
[13] D. Fourdrinier, W. Strawderman, M. Wells, Shrinkage Estimation, Springer, 2018.
[14] L. Haff, An identity for the Wishart distribution with applications, J. Multivariate Anal. 9 (1979) 531–544.
[15] L. Haff, Empirical Bayes estimation of the multivariate normal covariance matrix, Ann. Statist. 8 (1980) 586–597.
[16] L. Haff, On linear log-odds and estimation of discriminant coefficients, Commun. Statist. Theory Methods 15 (1986) 2131–2144.
[17] L. Haff, The variational form of certain Bayes estimators, Ann. Statist. 19 (1991) 1163–1190.
[18] D. Harville, Matrix Algebra From a Statistician's Perspective, Springer, 2008.
[19] Y. Ikeda, T. Kubokawa, M. Srivastava, Comparison of linear shrinkage estimators of a large covariance matrix in normal and non-normal distributions, Comput. Statist. Data Anal. 95 (2016) 95–108.
[20] W. James, C. Stein, Estimation with quadratic loss, in: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics, University of California Press, Berkeley, California, 1961, pp. 361–379.
[21] H. Ji, C. Liu, Z. Shen, Y. Xu, Robust video denoising using low rank matrix completion, in: 2010 IEEE Computer Society Conference on Computerv. and Pattern Recognition, IEEE, 2010, pp. 1791–1798.
[22] Y. Konno, Shrinkage estimators for large covariance matrices in multivariate real and complex normal distributions under an invariant quadratic loss, J. Multivariate Anal. 100 (2009) 2237–2253.
[23] T. Kubokawa, C. Robert, A. Saleh, Empirical Bayes estimation of the covariance matrix of a normal distribution with unknown mean under an entropy loss, Sankhyā A (1992) 402–410.
[24] T. Kubokawa, M. Srivastava, Robust improvement in estimation of a covariance matrix in an elliptically contoured distribution, Ann. Statist. 27 (1999) 600–609.
[25] O. Ledoit, M. Wolf, A well-conditioned estimator for large-dimensional covariance matrices, J. Multivariate Anal. 88 (2004) 365–411.
[26] H. Nguyen, X. Peng, M. Do, Z. Liang, Spatiotemporal denoising of MR spectroscopic imaging data by low-rank approximations, 2011.
[27] Y. Sheena, Unbiased Estimator of risk for an orthogonally invariant estimator of a covariance matrix, J. Jpn. Statist. Soc. 25 (1995) 35–48.
[28] C. Stein, Estimation of a covariance matrix, in: Rietz Lecture, 39th Annual Meeting of the IMS, Atlanta, Georgia, 1975.
[29] C. Stein, Lectures on the theory of estimation of many parameters, J. Sov. Math. 34 (1986) 1373–1403.
[30] A. Takemura, An orthogonally invariant minimax estimator of the covariance matrix of a multivariate normal population, Tsukuba J. Math. 8 (1984) 367–376.
[31] H. Tsukuma, Estimation of a high-dimensional covariance matrix with the Stein loss, J. Multivariate Anal. 148 (2016) 1–17.
[32] H. Tsukuma, T. Kubokawa, Unified improvements in estimation of a normal covariance matrix in high and low dimensions, J. Multivariate Anal. 143 (2016) 233–248.