Scale matrix estimation of an elliptically symmetric distributions

Anis M. HADDOUCHE[(1)], Dominique Fourdrinier[(2)] and Fatiha Mezoued[(3)]

21 January 2021

(1) INSA de Rouen Normandie, France
(2) Université de Rouen Normandie, France
(3) École Nationale Supérieure de Statistique et d'Économie Appliquée, Algérie

Introduction

The model

Consider the following additive model

$$Y = M + \mathcal{E}, \qquad \mathcal{E} \sim E\left(0_{mp}, I_m \otimes \Sigma\right) \tag{1.1}$$

where

- Y is an observed $m \times p$ matrix
- M is an unknown matrix of parameters

$$rank\,(M) = q < m \wedge p. \tag{1.2}$$

- $\mathcal{E}$ is an elliptically symmetric noise with unknown scale matrix (invertible) Σ.

We assume that $\mathcal{E}$ has a density w.r.t the Lebesgue measure in $\mathbb{R}^{pm}$ of the form

$$\varepsilon \mapsto |\Sigma|^{-m/2} f\{\mathrm{tr}\left(\varepsilon\,\Sigma^{-1}\varepsilon^{\mathrm{T}}\right)\},$$

for some function f (called the generating function).

The model

The Gaussian case $\Sigma = \sigma^2 \, I_p$ (σ^2 known)

[1] *E. J. Candès, C. A. Sing-Long and J. D. Trzasko. Unbiased risk estimates for singular value thresholding and spectral estimators.* IEEE T. Signal Process*, 61 : 4643–4657, 2013.*

Extension to the elliptical case (Σ known or unknown)

[2] *S. Canu and D. Fourdrinier. Unbiased risk estimates for matrix estimation in the elliptical case.* Journal of multivariate analysis *158 : 60–72, 2017.*

Applied setting

[10] *H. Ji, C. Liu, Z. Shen, and Y. Xu, Robust video denoising using low rank matrix completion.* 2010 IEEE Computer society conference on computer vision and pattern recognition*, 1791–1798, 2010.*

The canonical form

Thanks to (1.2), there exists a $m \times (m-q)$ semi-orthogonal matrix Q_2 such that

$$Q_2^{\mathrm{T}} M = 0\,.$$

Complete Q_2 with Q_1 to form an orthogonal matrix $Q = (Q_1 Q_2)$. Then

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

The density of $Q^{\mathrm{T}} \mathcal{E}$ is the same as that of $\mathcal{E}$. It follows that the density of $(Z^{\mathrm{T}} U^{\mathrm{T}})^{\mathrm{T}} = Q^{\mathrm{T}} Y$ is

$$(z,u) \mapsto |\Sigma|^{-m/2} f\left[\mathrm{tr}\{ (z-\theta)\Sigma^{-1}(z-\theta)^{\mathrm{T}} \} + \mathrm{tr}\{\Sigma^{-1} u^{\mathrm{T}} u\} \right]. \quad (1.3)$$

Why the canonical form ?

Recall that

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

- Inference on the $m \times p$ matrix M is reduced to inference on the $q \times p$ matrix θ ;

- Inference on the $p \times p$ matrix Σ relies on the $(m - q) \times p$ matrix U ;

- Both have low dimension than the $m \times p$ observed matrix Y.

Note that

$$S = U^{\mathrm{T}}\, U$$

is a sufficient statistic for Σ and may serve as an estimate of Σ.

Note also that S is invertible when $p \leq m - q$ and is non–invertible when $p > m - q$.

Related expectations

Let

$$F^*(t) = \frac{1}{2}\int_t^\infty f(\nu)\,d\nu \quad \text{and} \quad F^{**}(t) = \frac{1}{2}\int_t^\infty F^*(\nu)\,d\nu.$$

Bellow $E_{\theta,\Sigma}$ will be the expectation w.r.t (1.3), $E^*_{\theta,\Sigma}$ the expectation w.r.t

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

and $E^{**}_{\theta,\Sigma}$ the expectation w.r.t

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

where K^* and K^{**} are normalizing constants.

Subclass of Berger densities

We consider the subclass of densities such that

$$0 < c \leq \frac{F^*(t)}{f(t)} \leq b\,.$$

Examples :

- The logistic type distribution

$$f(t) \propto \frac{\exp\left(-\beta t - \gamma\right)}{\left(1 + \exp\left(-\beta\, t - \gamma\right)\right)^2}$$

where $\beta > 0$ and $\gamma > 0$. Here

$$c = \frac{1}{2\beta} \qquad b = \frac{(1 + e^{-\gamma})}{2\beta}.$$

- The Gaussian distribution : $c = b = 1$ since $F^* = f$.

[5] *D. Fourdrinier, F. Mezoued, and W. E. Strawderman. Bayes minimax estimators of a location vector for densities in the Berger class.* Electronic Journal of Statistics. *6 :783–809, 2012.*

Stein Phenomenon

In the following, we set $n = m - q$.

In the Gaussian setting, James and Stein

[9] *W. James and C. Stein, Estimation with Quadratic Loss.* Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, *1961.*

show that the usual estimators of the form

$$\hat{\Sigma}_a = a\,S \qquad \text{where} \qquad a > 0\,,$$

are inadmissible in

- the high dimensional setting ($p > n$)
- the low dimensional setting ($p \leq n$) for $p \approx n$

This phenomenon extends to the elliptical case.

Our objective

In a unified approach ($p > n$ et $p \leq n$), we aim to improve

$$\hat{\Sigma}_a = a\,S \qquad \text{where} \qquad a > 0\,,$$

by alternative (general) estimators of the form

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

where S^{+} is the Moore-Penrose inverse of S and $SS^{+}G(Z,S)$ is a correction matrix.

Evaluation of theses estimators will be made

- under the quadratic loss

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

- under the data–based loss

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

Previous works

The Gaussian case :

Konno (S non-invertible, quadratic loss)

[11] *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*.

Tsukuma and Kubokawa (A unified approach, Stein–loss)

[14] *H. Tsukuma and T. Kubokawa, Unified improvements in estimation of a normal covariance matrix in high and low dimensions.* Journal of Multivariate Analysis*, 143 : 233–248, 2016*.

The elliptical case :

Kubokawa and Srivastava (S invertible, Stein-loss)

[12] *T. Kubokawa and M. S. Srivastava, Robust improvement in estimation of a covariance matrix in an elliptically contoured distribution,* The Annals of Statistics*, 27 : 600–609, 1999*.

General estimators

Estimateurs alternatifs

Consider the quadratic risk

$$R(\hat{\Sigma}, \Sigma) = E_{\theta,\Sigma}[\mathrm{tr}(\Sigma^{-1}\hat{\Sigma} - I_p)^2].$$

When $\hat{\Sigma} = \hat{\Sigma}_a = a\,S$, the best constant a is given by

$$a_o = \frac{1}{K^{**}(n+p+1)}\,.$$

Consider alternative estimators of the form

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

where $SS^+G(Z,S)$ is a symmetric correction matrix.

The estimators $\hat{\Sigma}_{a_o,G}$ improves over $\hat{\Sigma}_{a_o}$ as soon as

$$\Delta(G) = R(\hat{\Sigma}_{a_o,G}, \Sigma) - R(\hat{\Sigma}_{a_o}, \Sigma) \leq 0$$

for all Σ, with strict inequality for some Σ.

Finiteness of the risk difference

Proposition 1

Assume that $E_{\theta,\Sigma}\left[\mathrm{tr}\left(\Sigma^{-1}\,S\right)^2\right] < \infty$ *and* $E_{\theta,\Sigma}\left[\mathrm{tr}(\Sigma^{-1}SS^{+}G)^2\right] < \infty$. *Then*

$$\Delta(G) = a_o^2\,E_{\theta,\Sigma}\left[\mathrm{tr}\left(\Sigma^{-1}SS^{+}(2S+G)\Sigma^{-1}SS^{+}G\right)\right] - 2\,a_o\,E_{\theta,\Sigma}\left[\mathrm{tr}\left(\Sigma^{-1}SS^{+}G\right)\right] < \infty.$$

Under which conditions on G this risk difference is non–positive ?

Our approach

Replacing the integrand term of $\Delta(G)$ by a random matrix $\delta(G)$, which does not depends on Σ^{-1} such that

$$\Delta(G) \leq E_{\theta,\Sigma}\big[\delta(G)\big].$$

A sufficient condition for $\Delta(G)$ to be non–positive is that $\delta(G)$ is non–positive.

To this end, we rely on the following Stein–Haff type identity.

A new Stein-Haff type identity

Lemma 1

Let $V(z,s)$ 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 $E_{\theta,\Sigma}\left[\left|\mathrm{tr}\left(\Sigma^{-1} S\, S^{+}\, V\right)\right|\right] < \infty$. Then we have

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

Here, $\mathcal{D}_S$ is the Haff (differential) operator with generic terms

$$d_{ij} = \frac{1}{2}(1+\delta_{ij})\frac{\partial}{\partial S_{ij}}$$

with $\delta_{ij} = 1$ where $i = j$ et $\delta_{ij} = 0$ where $i \neq j$.

Related identity

When S is invertible ($p \leq n$), since $S^{+} = S^{-1}$, this identity corresponds to

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

which is the same identity given by Kubokawa and Srivastava.

[12] *T. Kubokawa and M. S. Srivastava, Robust improvement in estimation of a covariance matrix in an elliptically contoured distribution,* The Annals of Statistics*, 27 : 600–609, 1999*.

When S is non–invertible ($p < n$), it becomes

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

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

Corollary 1

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

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

with

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

Application of the Stein-Haff type identity

Recall that

$$\Delta(G) = a_o^2\, E_{\theta,\Sigma}[\,\mathrm{tr}\,(\Sigma^{-1} SS^+ (2S + G)\Sigma^{-1} SS^+ G)\,] \\ - 2\, a_o\, E_{\theta,\Sigma}[\mathrm{tr}\,(\Sigma^{-1} SS^+ G)\,].$$

In order to get rid of Σ^{-1} in the integrand term, we apply

- Lemma 1 to $E_{\theta,\Sigma}\left[\mathrm{tr}\left(\Sigma^{-1} SS^+ G\right)\right]$,
- Corollary 1 to $E_{\theta,\Sigma}\left[\mathrm{tr}\left(\Sigma^{-1} SS^+ (2S + G)\Sigma^{-1} SS^+ G\right)\right]$.

Then, we have

$$\Delta(G) = a_o^2 K^* K^{**} E^{**}_{\theta,\Sigma}[\mathrm{tr}\left(2SS^+ \mathcal{D}_s\{SS^+ T^*\}^{\mathrm{T}} + (n - (p \wedge n) - 1) S^+ T^*\right)] \\ - 2 a_o K^* E^*_{\theta,\Sigma}[\mathrm{tr}\left(2SS^+ \mathcal{D}_s\{SS^+ G\} + (n - (p \wedge n) - 1) S^+ G\right)]$$

where

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

An expression of $\Delta(G)$

In order to have an homogeneous expression of $\Delta(G)$ in term of $E_{\theta,\Sigma}$-expectation, we use the following identities.

For any integrable function $H(Z, U)$, we have

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

and

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

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

$$\varphi^*_{\theta,\Sigma}(z,u) = \frac{F^*(\nu)}{f(\nu)} \quad \text{and} \quad \varphi^{**}_{\theta,\Sigma}(z,u) = \frac{F^{**}(\nu)}{F^*(\nu)},$$

with

$$\nu = \mathrm{tr}\{(z-\theta)\,\Sigma^{-1}\,(z-\theta)^{\mathrm{T}}\} + \mathrm{tr}\{\Sigma^{-1}\,u^{\mathrm{T}}\,u\}.$$

Dominance results

Theorem 1

Consider a density as in (1.3) s.t $c \leq F^*(t)/f(t) \leq b$. *Under the condition*

$$\mathrm{tr}\left[\, 2\, S^+ S\, \mathcal{D}_S\{SS^+ G\} - (n - (p \wedge n) - 1)\, S^+ G \right] \geq 0\,,$$

the estimators $\hat{\Sigma}_{a_o,G}$ *improves over* $\hat{\Sigma}_{a_o}$ *as soon as*

$$\mathrm{tr}\Big[2S^+ S\, \mathcal{D}_S\{SS^+ T^*\}^\top - S^+ T^* - 2(p+n+1)\frac{c^2}{b^2}\left(2S^+ S\, \mathcal{D}_S\{SS^+ G\} - (n-(p\wedge n)-1)S^+ G\right)\Big] \leq 0\,,$$

where

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

Limitations of the quadratic loss

- Difficult to handle on since we apply the Haff $\mathcal{D}_S$ twice.
- Imposes strong conditions on SS^+G, which makes difficult to derive improved estimators.

From where theses difficulties come from ?

Limitations of the quadratic loss

The quadratic loss can be rewritten as

two Σ^{-1}

$$L(\Sigma, \hat{\Sigma}) = \mathrm{tr}\left((\Sigma^{-1} \hat{\Sigma} - I_p)^2 \right) = \mathrm{tr}(\Sigma^{-1} \hat{\Sigma} \Sigma^{-1} \hat{\Sigma}) - 2\, \mathrm{tr}(\Sigma^{-1} \hat{\Sigma}) + p\,.$$

Requires a twice application of the Stein-Haff type identity.

How to remedy ?

We introduce the data, which give rise to the data–based loss

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

One Σ^{-1}

$$= \mathrm{tr}(\Sigma^{-1} \hat{\Sigma}\, S^+ \hat{\Sigma}) - 2\, \mathrm{tr}\,(S^+ \hat{\Sigma}) + \mathrm{tr}\,(S^+ \Sigma)\,.$$

Alternative estimators

Consider the data–based risk

$$R_S(\hat{\Sigma}, \Sigma) = E_{\theta,\Sigma}[\mathrm{tr}(S^+ \, \Sigma \, (\Sigma^{-1} \, \hat{\Sigma} - I_p)^2)].$$

When $\hat{\Sigma} = \hat{\Sigma}_a = a\, S$, the best constant a is

$$a_o = \frac{1}{K^* \, (p \vee n)} \, .$$

Consider alternative estimators of the form

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

where $SS^+ G(Z,S)$ is a correction matrix which is not necessarily symmetric.

The estimators $\hat{\Sigma}_{a_o,G}$ improves over $\hat{\Sigma}_{a_o}$ as soon as

$$\Delta(G) = R_S(\hat{\Sigma}_{a_o,G}, \Sigma) - R_S(\hat{\Sigma}_{a_o}, \Sigma) \leq 0,$$

for all Σ, with strict inequality for some Σ.

Risk difference finiteness

Proposition 2

Assume that $E_{\theta,\Sigma}\left[\|S^{+}G\|_F^2\right]$, $E_{\theta,\Sigma}\left[\|\Sigma^{-1}SS^{+}G\|_F^2\right]$, $E_{\theta,\Sigma}\left[\mathrm{tr}(\Sigma\, S^{+})\right]$ *and* $E_{\theta,\Sigma}\left[\mathrm{tr}(\Sigma^{-1}\, S)\right]$ *are finite.*
Then

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

Conditions d'amélioration

Theorem 2

Consider a density as in (1.3) s.t $c \leq F^*(t)/f(t) \leq b$. *Under the condition*

$$\mathrm{tr}(S^+ G) \geq 0 \,,$$

the estimators $\hat{\Sigma}_{a_o,G}$ *improves over* $\hat{\Sigma}_{a_o}$ *as soon as*

$$\mathrm{tr}\Big[(n-(p\wedge n)-1)(S^+GSS^+ + (S^+G)^2) + \alpha\, S^+G \\ +2\, SS^+\, \mathcal{D}_S\{SS^+G + SS^+GS^+G + SS^+\, G\, SS^+\}^T\Big] \leq 0\,,$$

where

$$\alpha = (n-(p\wedge n) - 2\,\frac{c}{b}\,(p\vee n) - 1)\,.$$

Orthogonally invariant estimators

The eigenvalue decomposition of S

Let the eigenvalue decomposition of S

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

where

$H_1 \in \mathbb{R}^{p\times(p\wedge n)}$ is a semi-orthogonal matrix

and

$$L = \mathrm{diag}(l_1, \ldots, l_{(p\wedge n)}) \quad \text{with} \quad l_1 > l_2 > \cdots > l_{(p\wedge n)} > 0\,.$$

The orthogonally invariant estimators are of the form

$$\hat{\Sigma} = H_1 \Phi(L) H_1^T$$

where

$$\Phi(L) = \mathrm{diag}(\phi_1(L), \ldots, \phi_{(p\wedge n)}(L)) \text{ with } \phi_1(L) > \cdots > \phi_{(p\wedge n)}(L).$$

The Haff type estimators

Recall that

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

Let

$$G(Z,S) = \nu \, t(\nu) H_1 \, H_1^{\mathrm{T}}$$

where

$$\nu = 1/\mathrm{tr}(S^{+})$$

and $t(\cdot)$ a twice differentiable non–increasing convex function.

The Haff type estimators are of the form

$$\hat{\Sigma}_{Haff} = a_o \, H_1 \, \mathrm{diag}\left(l_1 + \nu \, t(\nu), l_2 + \nu \, t(\nu), \ldots, l_{p \wedge n} + \nu \, t(\nu) \right) H_1^{\mathrm{T}} \, .$$

[8] *L.R. Haff, Empirical Bayes estimation of the multivariate normal covariance matrix,* The Annals of Statistics*, 8 : 586–597, 1980.*

Proposition 3

Consider a density as in (1.3) s.t $c \leq F^*(t)/f(t) \leq b$ *and*

$$\frac{p+n-(p\wedge n)+2}{p+n+1} \leq \frac{c^2}{b^2} \leq \frac{2\,p+2\,n-5\,(p\wedge n)-3}{p+n+1}\,.$$

Then $\hat{\Sigma}_{Haff}$ *improves over* $\hat{\Sigma}_{a_o}$ *as soon as*

(*i*) $(p+n-2\,(p\wedge n)+1)\,t(\nu)+2\,\nu\,t'(\nu)\geq 0\,,$

(*ii*) $0 \leq t(\nu) \leq \dfrac{2\,(p+n-2\,(p\wedge n)-1)\left((p+n+1)\frac{c^2}{b^2}-p-n+p\wedge n-2\right)}{(p+n-2\,(p\wedge n)+1)(p+n-2\,(p\wedge n)+3)}$

(*iii*) $$\{2\,(p+n-4\,(p\wedge n)+3)\,t(\nu)+2\,\nu\,t'(\nu) + \left[2\,p+2\,n-5\,(p\wedge n)+5-(p+n+1)\frac{c^2}{b^2}\right]\Big\}\,t'(\nu) + 2\,\left\{\,t(\nu)+(p\wedge n)^2\right\}\,\nu\,t''(\nu)\leq 0\,.$$

The Haff type estimators under a data–based loss

Proposition 4

Consider a density as in (1.3) s.t $c \leq F^*(t)/f(t) \leq b$ *and*

$$\frac{c}{b} \geq \frac{(p \vee n) - (p \wedge n) + 1}{(p \vee n)}.$$

Then $\hat{\Sigma}_{Haff}$ *improves over* $\hat{\Sigma}_{a_o}$

$$0 \leq t(\nu) \leq 2\,\frac{((p \wedge n) - 1) + (p \vee n)(c/b - 1)}{(p \vee n) - (p \wedge n) + 1}\,.$$

The Konno estimators

Recall that

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

Let

$$G(Z,S) = \nu\, t\, H_1\, H_1^{\mathrm{T}}$$

where

$$\nu = 1/\mathrm{tr}(S^{+})$$

and

$$t \text{ is a positive constant,}$$

The Konno estimators are of the form

$$\hat{\Sigma}_{Kon.} = a_o\, H_1\, \mathrm{diag}\big(l_1 + \nu\, t, l_2 + \nu\, t, \ldots, l_{p \wedge n} + \nu\, t\big) H_1^{\mathrm{T}}\ .$$

[11] *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*.

The Konno estimators under the quadratic loss

Proposition 5

Consider a density as in (1.3) s.t $c \leq F^*(t)/f(t) \leq b$ *and*

$$\frac{p+n-(p\wedge n)+2}{p+n+1} \leq \frac{c^2}{b^2} \leq \frac{2\,p+2\,n-5\,(p\wedge n)-3}{p+n+1}\,.$$

Then $\hat{\Sigma}_{Kon.}$ *improves over* $\hat{\Sigma}_{a_o}$ *as soon as*

$$0 \leq t \leq \frac{2\,(p+n-2\,(p\wedge n)-1)\,((p+n+1)c^2/b^2-p-n+p\wedge n-2\,)}{(p+n-2\,(p\wedge n)+1)(p+n-2\,(p\wedge n)+3)}$$

The Konno's estimators under the data–based loss

Proposition 6

Consider a density as in (1.3).

Then $\hat{\Sigma}_{Kon.}$ *improves over* $\hat{\Sigma}_{a_o}$ *as soon as*

$$0 \leq t \leq \frac{2\,((p \wedge n) - 1)}{(p \vee n) - (p \wedge n) + 1}\,.$$

This result is no longer restricted to the Burger subclass s.t. $c \leq F^{*}(t)/f(t) \leq b$ and then is valuable for all the classe of e.s.d

Estimateurs de type Efron et Morris

Efron and Morris type estimators

Recall that

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

Let $Q \in \mathbb{R}^{(p\wedge n)\times q}$ be a matrix of constants such that

$$Rang(Q) = q \leq p \wedge n.$$

Let the orthogonal projection matrix

$$Q_o = Q\,(Q^{\mathrm{T}}\,Q)^{-1}\,Q^{\mathrm{T}}.$$

When

$$G(Z,S) = \frac{t}{\mathrm{tr}(L^{-1}\,Q_o)}\,H_1\,Q_o\,H_1^{\mathrm{T}}, \qquad t > 0.$$

The Efron and Morris type estimators are of the form

$$\hat{\Sigma}_{EM} = a_o\left(S + \frac{t}{\mathrm{tr}(L^{-1}\,Q_o)}\,H_1\,Q_o\,H_1^{\mathrm{T}}\right)$$

Efron and Morris type estimators under the data–based loss

Proposition 7

Consider a density as in (1.3).

Then $\hat{\Sigma}_{EM}$ *improves on* $\hat{\Sigma}_{a_o}$ *as soon as*

$$0 \leq t \leq \frac{2\,((p \wedge n) - 1)}{(p \vee n) - (p \wedge n) + 1}\,.$$

Numerical study

We evaluate the percentage of improvement of $\hat{\Sigma}_{Kon.}$ w.r.t $\hat{\Sigma}_{a_o}$.

The Percentage Relative Improvement in Average Loss. (PRIAL) is defined as

$$PRIAL = \frac{\text{average loss of } \hat{\Sigma}_{a_o} - \text{average loss of } \hat{\Sigma}_{Kon.}}{\text{average loss of } \hat{\Sigma}_{a_o}} \times 100\,.$$

Note that :

$PRIAL_Q$: the percentage of improvement under the quadratic loss

$PRIAL_S$: the percentage of improvement under the data–based loss

Numerical study

1000 independent Monte–Carlo replications for some couples (p, n) and for

$$(\Sigma)_{ij} = 0.9^{|i-j|}.$$

p	n	$PRIAL_Q\%$	$PRIAL_S\%$
20	4	0.91	15.00
20	8	2.56	18.56
20	12	5.05	25.56
20	16	4.17	47.034
100	20	0.40	3.39
100	40	1.24	4.19
100	60	1.56	5.76
100	80	1.531	10.42

Some conclusions and perspectives

Conclusions

- Estimation of the scale matrix under quadratic loss and data–based loss for a wide class of e. s. d.
- Unified approach that can deal with both invertible and non–invertible S.
- A new and more general Stein-Haff identity for the high-dimensional e. s. d. setting.
- A robust improved estimators under the data–based loss.

Achievements :

- Published : [7] *A. M. Haddouche, D. Fourdrinier ans F. Mezoued, Scale matrix estimation of an elliptically symmetric distribution in high and low dimensions,* Journal of Multivariate Analysis *181 : 104680, 2021.*
- Submitted to Statistics and Probability letters : [6] *A. M. Haddouche, D. Fourdrinier and F. Mezoued, Covariance matrix estimation under a data–based loss.*

Perspectives

Consider

$$Y = M + \mathcal{E}, \qquad \mathcal{E} \sim ES\,(0_{mp}, I_m \otimes \Sigma)$$

where Σ is of Rank = $r < p$.

The noise $\mathcal{E}$ does not have a density with respect to the Lebesgue measure.

[4] *J.A Díaz-García, V. Leiva-Sánchez and M. Galea, Singular elliptical distribution : density and applications,* Communications in Statistics–theory and methods. *5 : 665–681, 2002.*

Singular elliptical symmetric distributions

Estimate Σ under :

$$L(\hat{\Sigma}, \Sigma) = \mathrm{tr}\left(\hat{\Sigma}\Sigma^{+} - \Sigma\Sigma^{+}\right)^{2}$$
$$L_S(\hat{\Sigma}, \Sigma) = \mathrm{tr}\left(S^{+}\Sigma\left(\hat{\Sigma}\Sigma^{+} - \Sigma\Sigma^{+}\right)^{2}\right)$$

Estimate Σ^{+} under

$$L(\hat{\Sigma}^{+}, \Sigma^{+}) = \| \hat{\Sigma}^{+} - \Sigma^{+} \|_F^2 \; .$$

The Gaussian case

Orthogonally invariant estimators

[3] *D. Chételat, M. T. Wells, Improved second order estimation in the singular multivariate normal model,* Journal of Multivariate Analysis , *147 :11 – 19, 2016.*

A Stein–Haff type identity

$$E\left[\operatorname{tr}\left(\Sigma^{+}H_1\Phi(L)H_1^{\top}\right)\right] = E\left[\sum_{i=1}^{n\wedge r}\left\{(|n-r|-1)\frac{\phi_i}{l_i} + 2\frac{\partial\phi_i}{\partial l_i} + 2\sum_{j>i}\frac{\phi_i-\phi_j}{l_i-l_j}\right\}\right]$$

where $Rang(S) = n \wedge r$.

Generalization to the elliptical singular case

- How to derive a new Stein-Haff type identity for orthogonally invariant estimators ?

- How to derive a Stein-Haff type identity of the form

$$E\left[\mathrm{tr}\left(\Sigma^{+}SS^{+}G(Z,S)\right)\right]=?$$

Predictive density estimation

Predictive density estimation under the Wasserstein loss for location and scale–location distribution :

[13] *T. Matsuda and W. E. Strawderman, Predictive density estimation under the Wasserstein loss,* Journal of Statistical Planning and Inference*, 210 : 53 – 63, 2021.*

The authors showed that the plug-in densities is a complete class.

A duality with the point estimation problem is showed.

Results on point estimation are transportable to the predictive density estimation under the Wasserstein loss function.

Predictive density estimation

An Extension to the case of an unknown covariance matrix is an interesting future problem (which include the elliptical case)

The problem reduces to point estimation of the covariance matrix under the following Wasserstein loss

$$L(\hat{\Sigma}, \Sigma) = \mathrm{tr}\left(\Sigma + \hat{\Sigma} - 2\,(\Sigma^{1/2}\,\hat{\Sigma}\,\Sigma^{1/2})^{1/2}\right)$$

Using, the usual key tool, Stein–Haff type identity seems not possible since the loss does not depends on Σ^{-1}.

Thank you for your attention !

For references and other details, I can be reached at

mohamed.haddouche@insa-rouen.fr

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

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

D. Chételat and M.T. Wells.
Improved second order estimation in the singular multivariate normal model.
Journal of Multivariate Analysis, 147(C), 2016.

J.A. Díaz-García, V. Leiva-Sánchez, and M. Galea.
Singular elliptical distribution : Density and apllications.
Communications in Statistics - Theory and Methods, 31(5) :665–681, 2002.

D. Fourdrinier, F. Mezoued, and W.E. Strawderman.

Bayes minimax estimators of a location vector for densities in the Berger class.
Electronic Journal of Statistics, 6 :783–809, 2012.

Anis M Haddouche, Dominique Fourdrinier, and Fatiha Mezoued.
Covariance matrix estimation under data-based loss.
arXiv preprint arXiv :2012.11920, 2020.

Anis M. Haddouche, Dominique Fourdrinier, and Fatiha Mezoued.
Scale matrix estimation of an elliptically symmetric distribution in high and low dimensions.
Journal of Multivariate Analysis, 181 :104680, 2021.

L.R. Haff.
Empirical bayes estimation of the multivariate normal covariance matrix.
The Annals of Statistics, 8(3) :586–597, 1980.

W. James and 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*, pages 361–379, Berkeley, Calif., 1961. University of California Press.

H. Ji, C. Liu, Z. Shen, and Y. Xu.
Robust video denoising using low rank matrix completion.
In *2010 IEEE Computer Society Conference on Computerv. and Pattern Recognition*, pages 1791–1798. IEEE, 2010.

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(10) :2237–2253, 2009.

T. Kubokawa and M.S. Srivastava.
Robust improvement in estimation of a covariance matrix in an elliptically contoured distribution.
The Annals of Statistics, 27(2) :600–609, 1999.

Takeru Matsuda and William E. Strawderman.

Predictive density estimation under the wasserstein loss.
Journal of Statistical Planning and Inference, 210 :53 – 63, 2021.

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