



HAL
open science

Estimation d'une matrice d'échelle.

Mohamed Anis Haddouche

► **To cite this version:**

Mohamed Anis Haddouche. Estimation d'une matrice d'échelle.. Analyse numérique [math.NA]. Normandie Université; École nationale supérieure de statistique et d'économie appliquée (Alger), 2019. Français. NNT : 2019NORMR058 . tel-02376077

HAL Id: tel-02376077

<https://theses.hal.science/tel-02376077>

Submitted on 22 Nov 2019

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

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



Normandie Université

THÈSE EN CO - TUTELLE INTERNATIONALE

Pour obtenir le diplôme de doctorat

Spécialité Mathématiques

Préparée au sein de l'Université de Rouen Normandie
et de l'ENSSEA

Estimation d'une matrice d'échelle.

Mohamed Anis HADDOUCHE

Thèse soutenue publiquement le 31 Octobre 2019
devant le jury composé de

Mme. Hafida GUERBYENNE	Professeure / Enseignante-Chercheuse / USTHB	Rapporteuse
Mme. Nadja EL SAADI	Professeure / Enseignante-Chercheuse / ENSSEA	Examinatrice
Mme. Fatiha MEZOUE	Professeure / Enseignante-Chercheuse / ENSSEA	Directrice
M. William E. STRAWDERMAN	Professeur / Enseignant-Chercheur / Rutgers University	Rapporteur
M. Dominique FOURDRINIER	Professeur / Enseignant-Chercheur / Université de Rouen Normandie	Directeur
M. Fayçal HAMD	Professeur / Enseignant-Chercheur / USTHB	Examineur

Thèse dirigée par

M. Dominique FOURDRINIER

M. Fatiha MEZOUE

Université de Rouen Normandie

ENSSEA



À la mémoire de Fatiha, Sid Ali et Abd El Rahman.

*Only as a warrior can one
withstand the path of knowledge.
A warrior cannot complain or
regret anything. His life is an
endless challenge, and
challenges cannot possibly be
good or bad. Challenges are
simply challenges.*

Carlos Castaneda

REMERCIEMENTS

Cette thèse est loin d'être un travail solitaire. Je n'aurais jamais pu le réaliser sans le soutien d'un grand nombre de personnes que je souhaiterais remercier ici.

En premier lieu, je remercie mes deux directeurs de thèse, Fatiha MEZOUEE et Dominique FOURDRINIER pour la confiance qu'ils m'ont accordé en acceptant d'encadrer ce travail doctoral et pour m'avoir guidé, encouragé et conseillé tout au long de cette thèse. J'aimerais également leur dire que j'ai été extrêmement sensible à leurs qualités humaines d'écoute et de compréhension. Je ne sais pas comment exprimer ma gratitude à ces deux personnes autrement qu'en leur promettant d'agir comme eux avec des étudiants dans ma situation si un jour l'occasion m'en est donnée.

Ensuite, je remercie Hafida GUERBYENNE et William E. STRAWDERMAN pour avoir accepté de relire ce mémoire et d'en être rapporteurs. La version finale de ce manuscrit a bénéficié de leur lecture très attentive et de leurs remarques précieuses. Je tiens aussi à remercier les membres du jury de ma soutenance : EL SAADI Nadja, SAIDI Ghania et HAMDI Fayçal. Je les remercie pour leur présence, pour leur lecture attentive de ma thèse et pour les remarques qu'ils m'adresseront afin d'améliorer mon manuscrit.

J'adresse aussi de sincères remerciements à toutes les personnes qui ont contribué à la création et au bon fonctionnement du programme, Algéro-Français, de bourse en doctorat Profas B+, sans qui ce travail n'aurait pas été possible. Grâce à ce programme, j'ai bénéficié d'une allocation de recherche de 14 mois qui m'a permis de travailler sereinement avec mon directeur de thèse Dominique FOURDRINIER, lors de mes différents séjours au laboratoire LITIS.

Je souhaiterais également remercier mes amis et mes collègues de l'université de Rouen Normandie et de L'ENSSEA pour la bonne ambiance de travail et pour les nombreux bons moments passés ensemble. Entre autres, Abraham (l'homme fort!), Djamila, Renata, Karim, Léo (l'Ardechois), Léo (le Chinois), Fabrice, Rachid, Wassim, Yann et Youcef.

Je me tourne maintenant vers ma famille et mes proches que j'ai quelque peu délaissés ces dernières années pour achever cette thèse : ma chère maman Sofia, mon oncle Liés et sa femme Anissa, ma tante Remla, mes frères et sœurs : Omar, Zola, Meriem, Salim, Sid Ahmed et Nanou, ainsi que Adel (mon meilleur ami de l'hémisphère nord!). Je les remercie chaleureusement pour leur soutien inconditionnel et leurs encouragements.

Enfin, mes plus profonds remerciements vont à Sihem, merci pour ton attention, ton soutien quotidien et ton enthousiasme contagieux.

ESTIMATION D'UNE MATRICE D'ÉCHELLE

RÉSUMÉ

Beaucoup de résultats sur l'estimation d'une matrice d'échelle en analyse multidimensionnelle sont obtenus sous l'hypothèse de normalité (condition sous laquelle il s'agit de la matrice de covariance). Or il s'avère que, dans des domaines tels que la gestion de portefeuille en finance, cette hypothèse n'est pas très appropriée. Dans ce cas, la famille des distributions à symétrie elliptique, qui contient la distribution gaussienne, est une alternative intéressante.

Nous considérons dans cette thèse le problème d'estimation de la matrice d'échelle Σ du modèle additif $Y_{p \times m} = M + \mathcal{E}$, d'un point de vue de la théorie de la décision. Ici, p représente le nombre de variables, m le nombre d'observations, M une matrice de paramètres inconnus de rang $q < p$ et \mathcal{E} un bruit aléatoire de distribution à symétrie elliptique, avec une matrice de covariance proportionnelle à $I_m \otimes \Sigma$. Ce problème d'estimation est abordé sous la représentation canonique de ce modèle où la matrice d'observation Y est décomposée en deux matrices, à savoir, $Z_{q \times p}$ qui résume l'information contenue dans M et une matrice $U_{n \times p}$, où $n = m - q$, qui résume l'information suffisante pour l'estimation de Σ .

Comme les estimateurs naturels de la forme $\hat{\Sigma}_a = aS$ (où $S = U^T U$ et a est une constante positive) ne sont pas de bons estimateurs lorsque le nombre de variables p et le rapport p/n sont grands, nous proposons des estimateurs alternatifs de la forme $\hat{\Sigma}_{a,G} = a(S + S S^+ G(Z, S))$ où S^+ est l'inverse de Moore-Penrose de S (qui coïncide avec l'inverse S^{-1} lorsque S est inversible). Nous fournissons des conditions sur la matrice de correction $S S^+ G(Z, S)$ telles que $\hat{\Sigma}_{a,G}$ améliore $\hat{\Sigma}_a$ sous le coût quadratique $L(\Sigma, \hat{\Sigma}) = \text{tr}(\hat{\Sigma} \Sigma^{-1} - I_p)^2$ et sous une modification de ce dernier, à savoir le coût basé sur les données $L_S(\Sigma, \hat{\Sigma}) = \text{tr}(S^+ \Sigma (\hat{\Sigma} \Sigma^{-1} - I_p)^2)$. Nous adoptons une approche unifiée des deux cas où S est inversible et S est non inversible. À cette fin, une nouvelle identité de type Stein-Haff et un nouveau calcul sur la décomposition en valeurs propres de S sont développés. Notre théorie est illustrée par une grande classe d'estimateurs orthogonalement invariants et par un ensemble de simulations.

Mots clés : distribution à symétrie elliptique, coût quadratique, coût basé sur les données, estimateurs orthogonalement invariants, Stein-Haff identité, matrice de covariance, matrice d'échelle.

SCALE MATRIX ESTIMATION

ABSTRACT

Numerous results on the estimation of a scale matrix in multivariate analysis are obtained under Gaussian assumption (condition under which it is the covariance matrix). However in such areas as Portfolio management in finance, this assumption is not well adapted. Thus, the family of elliptical symmetric distribution, which contains the Gaussian distribution, is an interesting alternative.

In this thesis, we consider the problem of estimating the scale matrix Σ of the additif model $Y_{p \times m} = M + \mathcal{E}$, under theoretical decision point of view. Here, p is the number of variables, m is the number of observations, M is a matrix of unknown parameters with rank $q < p$ and \mathcal{E} is a random noise, whose distribution is elliptically symmetric with covariance matrix proportional to $I_m \otimes \Sigma$. It is more convenient to deal with the canonical forme of this model where Y is decomposed in two matrices, namely, $Z_{q \times p}$ which summarizes the information contained in M , and $U_{n \times p}$, where $n = m - q$, which summarizes the information sufficient to estimate Σ .

As the natural estimators of the form $\hat{\Sigma}_a = a S$ (where $S = U^T U$ and a is a positive constant) perform poorly when the dimension of variables p and the ratio p/n are large, we propose estimators of the form $\hat{\Sigma}_{a,G} = a(S + S S^+ G(Z, S))$ where S^+ is the Moore-Penrose inverse of S (which coincides with S^{-1} when S is invertible). We provide conditions on the correction matrix $S S^+ G(Z, S)$ such that $\hat{\Sigma}_{a,G}$ improves over $\hat{\Sigma}_a$ under the quadratic loss $L(\Sigma, \hat{\Sigma}) = \text{tr}(\hat{\Sigma} \Sigma^{-1} - I_p)^2$ and under the data based loss $L_S(\Sigma, \hat{\Sigma}) = \text{tr}(S^+ \Sigma (\hat{\Sigma} \Sigma^{-1} - I_p)^2)$. We adopt a unified approach of the two cases where S is invertible and S is non-invertible. To this end, a new Stein-Haff type identity and calculus on eigenstructure for S are developed. Our theory is illustrated with the large class of orthogonally invariant estimators and with simulations.

Key words : elliptically symmetric distributions, quadratic loss, data based loss, orthogonally invariant estimators, Stein-Haff type identity, covariance matrix, scale matrix.

SYMBOLES

\mathbb{R}	L'ensemble des nombres réels.
\hat{A}	L'estimateur du paramètre inconnu A .
I_m	La matrice identité de dimension $m \times m$.
A^T	La transposée de la matrice A .
A^+	L'inverse de Moore-Penrose de A .
$\text{tr}(A)$	La trace de la matrice A .
$\ A\ $	La norme euclidienne du vecteur A .
$\ A\ _F$	La norme de Frobenius de la matrice A .
$ A $	Le déterminant de la matrice A .
$\text{rang}(A)$	Le rang de la matrice A .
\otimes	Le produit de Kronecker.
\sim	Signifie <i>distribué selon</i> .
$E_{\theta, \Sigma}$	L'espérance d'une variable distribuée selon une densité $f(\cdot)$ paramétrée par θ et Σ .
δ_{ij}	Le symbole de Kronecker où $\delta_{ij} = 1$ si $i = j$ et $\delta_{ij} = 0$ si $i \neq j$.
$a \wedge b$	Le minimum entre les deux entiers a et b .
$a \vee b$	Le maximum entre les deux entiers a et b .
f'	La dérivée de la fonction f .
$\stackrel{d}{=}$	L'égalité en distribution.
\mathcal{O}_p	L'ensemble des matrices orthogonales $\{A \in \mathbb{R}^{p \times p} : A^T A = A A^T = I_n\}$.
$\mathcal{L}_{p,r}$	L'ensemble de Stiefel $\{A \in \mathbb{R}^{p \times r} : A^T A = I_r\}$.
\mathcal{D}_p	L'ensemble des matrices diagonales $\text{diag}(d_1, \dots, d_p)$ où $d_1 >, \dots, > d_p$.
$\Gamma(\cdot)$	La fonction Gamma sur l'ensemble \mathbb{R}_+^* .

LISTE DES FIGURES

1.1	Estimation des valeurs propres de $\Sigma = I_p$ par $\hat{\Sigma}_{1/n-1}$	4
5.1	Frontière efficiente.	58
5.2	Loi uniforme sur la sphère unité.	59
5.3	Densités de la loi normale pour Σ proportionnel à I_2	60
5.4	Contours de la loi normale avec des covariances positives.	61
5.5	Contours de la loi normale avec des covariances négatives.	61
5.6	Effet de t sur le PRIAL lorsque $\Sigma = I_p$ pour $(p, n) = (120, 100)$	64
5.7	Effet de t sur le PRIAL lorsque $(\Sigma)_{ij} = 0.9^{ i-j }$ pour $(p, n) = (120, 100)$	64

LISTE DES TABLEAUX

2.1	Exemples de densités à symétrie sphérique.	14
5.1	Distribution du carre du rayon pour quelques lois sphériques	60
5.2	Résultat de 1000 simulations de type Monte Carlo pour $p/n = 2$, $\Sigma = I_p$	63
5.3	Résultat de 1000 simulations de type Monte Carlo pour $p/n = 2$, $(\Sigma)_{ij} = 0.9^{ i-j }$	63
5.4	Résultats pour deux valeurs de p fixées, la valeur de n variant, $\Sigma = I_p$	63
5.5	Résultats pour deux valeurs de p fixées, la valeur de n variant, $(\Sigma)_{ij} = 0.9^{ i-j }$	63

TABLE DES MATIÈRES

Remerciement	i
Résumé	ii
Abstract	iii
Symboles	iv
Liste des figures	v
Liste des tableaux	v
Table des matières	vi
1 Introduction	1
1.1 Matrice de covariance	1
1.2 Estimation de la matrice d'échelle	2
1.3 État de l'art	4
1.4 Objectif et structure de la thèse	8
2 Le cadre des lois à symétrie elliptique	11
2.1 Introduction	11
2.2 Lois à symétrie sphérique	12
2.3 Lois à symétrie elliptique	14
2.4 Lois à symétrie elliptique vectorisée	17
2.5 Le modèle additif	18
2.6 Une sous-classe de lois à symétrie elliptique de type Berger	21
2.7 Une identité de type Stein-Haff	24
2.8 Synthèse	27
3 Estimation de la matrice d'échelle sous un coût quadratique	28
3.1 Introduction	28
3.2 Estimateurs de forme générale	29
3.3 Estimateurs orthogonalement invariants	36
3.4 Synthèse	43
4 Estimation d'une matrice d'échelle sous un coût basé sur les données	44
4.1 Introduction	44
4.2 Estimateurs de forme générale	45

4.3	Estimateurs orthogonalement invariants	50
4.4	Approche spécifique pour l'estimateur de Konno	54
4.5	Synthèse	55
5	Application	56
5.1	La théorie moderne du portefeuille	56
5.2	Simulation des lois à symétrie elliptique	59
5.3	Étude numérique	62
6	Conclusion	65
	Appendices	67
A	Définitions et calculs matriciels	68
B	Expressions différentielles	72
C	Article soumis	79
	Bibliographie	109

INTRODUCTION

Sommaire

1.1	Matrice de covariance	1
1.2	Estimation de la matrice d'échelle	2
1.3	État de l'art	4
1.3.1	Le contexte en petite dimension	4
1.3.2	Le contexte en grande dimension	6
1.3.3	Le contexte unifié	8
1.4	Objectif et structure de la thèse	8

1.1 Matrice de covariance

En statistique et en théorie des probabilités, la covariance entre deux variables aléatoires réelles est un nombre permettant de quantifier leurs écarts conjoints par rapport à leurs espérances respectives. C'est une extension de la notion de variance initialement introduite par Fisher [Fis19] qui la définit comme une mesure de la variabilité d'un phénomène observé. Le concept de covariance se généralise naturellement à plusieurs variables (vecteur aléatoire) par la matrice de covariance. Celle-ci permet de quantifier la variation¹ de chaque variable par rapport à chacune des autres. Elle joue un rôle important dans de nombreuses procédures d'estimation en statistique comme :

- La gestion de portefeuille dans les marchés boursiers (*cf.* Markowitz [Mar52]) et la mesure du risque (*cf.* Jorion [Jor96]) qui sont des applications en finance où la matrice de covariance capture la dépendance entre les actifs boursiers.
- L'analyse en composantes principales (ACP) et ses extensions (*cf.* Jolliffe [Jol02]) qui sont des techniques populaires de réduction des dimensions.

1. Il est intéressant de remarquer que, pour deux v.a. réelles et définies sur \mathbb{R} , leur covariance est positive dès qu'elles sont monotones dans le même sens, et négative, dès qu'elles sont monotone en sens contraire.

- La méthode des moments généralisée (MMG) qui est une méthode permettant d'estimer les paramètres dans un modèle statistique (*cf.* Hansen [Han82]) où une estimation précise de la matrice de covariance augmente l'efficacité de la MMG.
- L'analyse discriminante linéaire (ADL) qui fait partie des techniques d'analyse discriminante prédictive (*cf.* Fisher [Fis36]) et permet d'expliquer et de prédire l'appartenance d'un individu à une classe prédéfinie à partir de ses caractéristiques mesurées à l'aide de variables prédictives.

Beaucoup de résultats sur l'estimation de la matrice de covariance en analyse multidimensionnelle sont obtenus sous l'hypothèse de normalité. Bien que la loi normale est l'une des lois de probabilité la plus adapté pour modéliser des phénomènes naturels issus de plusieurs événements aléatoires, il s'avère que, dans de nombreux domaines comme la finance ou le traitement de signal, l'hypothèse de normalité n'est pas très adaptée. En gestion de portefeuille (*cf.* Markowitz [Mar52]) par exemple où les rendements des actifs sont souvent modélisés par la loi normale, les valeurs extrêmes des rendements sont mal captées par celle-ci en raison de ses queues légères. Ainsi la famille des distributions à symétrie elliptique (*cf.* chapitre 2), qui contient la distribution gaussienne, s'avère une alternative intéressante puisqu'elle offre la possibilité de construire des distributions possédant des queues lourdes qui captent mieux les valeurs extrêmes.

Le contexte de base de cette thèse est le suivant. Soit un vecteur aléatoire $U \in \mathbb{R}^p$ composé de p variable U_1, \dots, U_p , de distribution à symétrie elliptique, ou on note

$$U \sim ES(\theta, \Sigma),$$

de moyenne $\theta = E_{\theta, \Sigma}(U)$, où $E_{\theta, \Sigma}(\cdot)$ désigne l'espérance par rapport à la loi de U , et de matrice de covariance

$$Cov(U) = \frac{1}{p} E_{\theta, \Sigma} [(U - \theta)^T \Sigma^{-1} (U - \theta)] \Sigma,$$

pour une certaine matrice, dite d'échelle, Σ qu'on cherche à estimer. On voit que Σ n'est pas nécessairement la matrice de covariance de U sauf dans le cas où la loi est gaussienne (*cf.* théorème 2.3.2).

Certains aspects de Σ jouent un rôle majeur dans l'estimation de celle-ci, comme ses valeurs propres λ_i et ses vecteurs propres v_i qui interviennent dans de nombreuses applications comme l'analyse en composantes principales (ACP). Ils sont définis par l'équation

$$\Sigma v_i = \lambda_i v_i, \quad \text{tels que } \|v_i\|^2 = 1.$$

Ainsi, si toutes ses valeurs propres sont positives, la matrice de covariance est définie positive et donc inversible. Dans ce travail, nous supposons que Σ est inversible.

1.2 Estimation de la matrice d'échelle

Disposant d'un n -échantillon U_1, \dots, U_n d'un vecteur aléatoire $U \sim ES(\theta, \Sigma)$, les estimateurs (*cf.* définition A.2) usuels de la matrice d'échelle Σ sont de la forme

$$\hat{\Sigma}_a = a \sum_{i=1}^n (U_i - \hat{\theta}) (U_i - \hat{\theta})^T = a S, \quad (1.2.1)$$

où a est une constante positive et $\hat{\theta} = (1/n) \sum_{i=1}^n U_i$ est la moyenne empirique de l'échantillon. Ce qui motive le choix de cette classe est que, sous l'hypothèse de normalité du vecteur U , l'estimateur du maximum de vraisemblance correspond à $\hat{\Sigma}_{1/n}$ et l'estimateur sans biais (*cf.*

définition A.3) correspond à $\hat{\Sigma}_{1/n-1}$. Se pose alors la question du choix d'un estimateurs dans la classe (1.2.1), *i.e.* de la constante a .

En théorie de la décision, tout estimateur $\hat{\Sigma}$ de Σ est évalué au travers d'une fonction de coût $L(\Sigma, \hat{\Sigma})$ qui représente le coût (positif) encouru par l'estimation $\hat{\Sigma}(U_1, \dots, U_n)$ alors que Σ est la matrice d'échelle inconnue. Bien entendu, on souhaite que ce coût soit aussi petit que possible mais son principal inconvénient est qu'il dépend de l'observation (U_1, \dots, U_n) ; il peut donc différer pour d'autres observations (U'_1, \dots, U'_n) . Pour remédier à cette contrainte, une approche classique, dans la théorie de la décision, consiste à considérer le coût *moyen*, *i.e.*, la fonction de risque

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

qui, en d'autres termes, est l'espérance du coût par rapport à la loi de U . La comparaison entre estimateurs se fait alors dans le cadre suivant. On dit qu'un estimateur $\hat{\Sigma}_1$ est meilleur que (ou domine) $\hat{\Sigma}_2$ si, pour tout Σ ,

$$R(\Sigma, \hat{\Sigma}_1) \leq R(\Sigma, \hat{\Sigma}_2)$$

avec une égalité pour au moins un Σ . Un estimateur $\hat{\Sigma}$ est alors dit admissible s'il n'existe pas d'estimateurs qui le domine, et inadmissible dans le cas contraire.

Dans notre contexte, les coûts classiques sont la fonction de coût quadratique (que nous considérons dans le chapitre 3)

$$L_1(\Sigma, \hat{\Sigma}) = \text{tr}(\hat{\Sigma}\Sigma^{-1} - I_p)^2 \quad (1.2.2)$$

et aussi le coût de Stein (aussi connu comme le coût entropique ou coût de Kullback-Leibler)

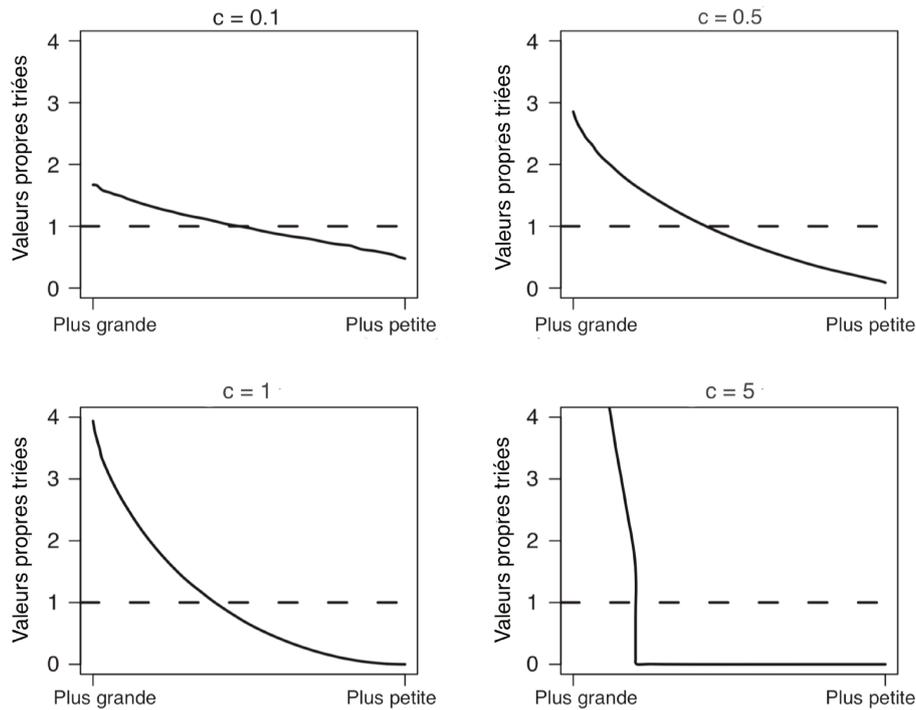
$$L_2(\Sigma, \hat{\Sigma}) = \text{tr}(\hat{\Sigma}\Sigma^{-1}) - \log |\hat{\Sigma}\Sigma^{-1}| - p. \quad (1.2.3)$$

Remarquons que ceux-ci ont toutes les propriétés attrayantes des fonctions de coûts. En effet, ils sont convexes (au sens où, pour tout Σ fixé, les fonctions $\hat{\Sigma} \mapsto L_i(\Sigma, \hat{\Sigma})$, où $i = 1, 2$, sont convexes) et ils sont, positifs, égaux à 0 lorsque $\hat{\Sigma} = \Sigma$ (quand on a pris la bonne décision d'estimation). De plus, ils sont invariants par transformation linéaire, *i.e.*, pour toutes matrices non singulière H de dimension $p \times p$,

$$L_i(H\Sigma H^T, H\hat{\Sigma}H^T) = L_i(\Sigma, \hat{\Sigma}).$$

L'estimation de Σ est caractérisée par le nombre de variables p et la taille de l'échantillon n . Dans le cas gaussien, James et Stein [JS61] montrent que pour les coûts $L_1(\Sigma, \hat{\Sigma})$ et $L_2(\Sigma, \hat{\Sigma})$, les estimateurs usuels de la forme $\hat{\Sigma}_a$ dans (1.2.1) sont inadmissibles pour Σ lorsque p et $c = p/n$ sont grands. Cette inadmissibilité est à mettre en parallèle avec celle de $\hat{\theta}$ dans (1.2.1) quand $p \geq 3$ découverte par Stein [Ste56] dans le contexte de l'estimation du vecteur moyenne θ d'une distribution normale (phénomène de Stein). Notons que ce phénomène se généralise, naturellement, au cas des distributions à symétrie elliptique.

Ainsi, quant à $\hat{\Sigma}_{1/n-1}$, s'il s'agit d'un estimateur de référence (*cf.* Anderson [And03]) car sans biais (*cf.* définition A.3) et consistant (*cf.* définition A.4), il sur-estime (sous-estime) les plus grandes (plus petites) valeurs propres de Σ (*cf.* Pourahmadi [Pou13]). Il est donc naturel de chercher à l'améliorer en terme de risque. Ce phénomène est visible dans la Figure 1.1 où on cherche à estimer $\Sigma = I_p$. La courbe en trait plein représente la distribution des valeurs propres de $\hat{\Sigma}_{1/n-1}$ triées par ordre décroissant. La ligne droite en 1 représente les valeurs propres de Σ . Les différents graphiques correspondent aux quatre valeurs indiquées de c . Pour $c \leq 1$, on remarque que le phénomène s'accroît lorsque p grandit. L'estimation est de plus en plus

FIGURE 1.1 – Estimation des valeurs propres de $\Sigma = I_p$ par $\hat{\Sigma}_{1/n-1}$.

mauvaise lorsque $c > 1$. Dans ces deux cas, les valeurs propres de la matrice d'échelle empirique sont trop dispersées car la plupart d'entre elles sont soit trop grandes, soit proches de zéro. Ainsi $\hat{\Sigma}_{1/n-1}$ est-il un estimateur mal conditionné. De plus il n'est pas inversible quand $c > 1$ (ou $p > n$), puisque $p - n$ valeurs propres sont égales à 0, alors que Σ est inversible.

1.3 État de l'art

Depuis le travail pionnier de James et Stein [JS61] une vaste littérature s'en est suivie et des estimateurs améliorés ont été proposés sous différents coûts. Nous allons passer en revue les travaux les plus pertinents à notre sens en les regroupant en fonction des dimensions p et n . Ainsi nous parlerons du *contexte en petite dimension* lorsque $p \leq n$ et du *contexte en grande dimension* lorsque $p > n$. Notons que dans tous les travaux que nous allons citer, U est supposé suivre une distribution gaussienne de matrice d'échelle (covariance) Σ sauf si on mentionne le contraire.

1.3.1 Le contexte en petite dimension

James et Stein [JS61] ont montré que sous $L_2(\Sigma, \hat{\Sigma})$ dans (1.2.3), les estimateurs de la forme $\hat{\Sigma}_a$ sont inadmissibles et peuvent être améliorés par un estimateur de la forme

$$\hat{\Sigma}^{JS} = T D T^T \quad (1.3.1)$$

où, $D = \text{diag}(d_1, \dots, d_p)$, $d_i = (n + p + 1 - 2i)^{-1}$ et la matrice triangulaire supérieure T vérifie $T^T T = S$. Un résultat similaire est obtenu sous le coût $L_1(\Sigma, \hat{\Sigma})$ dans (1.2.2), mais les auteurs

ont eu des difficultés à donner une expression explicite pour les d_i . Plus tard, Stein [Ste75], [Ste77], [Ste86] observe que l'estimateur du maximum de vraisemblance $\hat{\Sigma}_{1/n}$ est biaisé vers le haut pour les plus grandes valeurs propres de Σ et vers le bas pour les plus petites. Il suggère d'utiliser des estimations améliorées des valeurs propres au sein d'estimateurs orthogonalement invariants (*cf.* chapitre 3) pour réduire la dispersion des valeurs propres. Sa démarche repose sur la décomposition en valeurs propres de S , à savoir

$$S = HLH^T$$

où, $L = \text{diag}(l_1, \dots, l_p)$ est la matrice des valeurs propres avec $l_1 > \dots > l_p > 0$ et H est la matrice orthogonale des vecteurs propres associés. Il propose une classe d'estimateurs alternatifs qui sont orthogonalement invariants *i.e* de la forme

$$\hat{\Sigma} = H\Phi(L)H^T \quad (1.3.2)$$

où $\Phi(L) = \text{diag}(\phi_1(L), \dots, \phi_p(L))$ avec $\phi_1(L) > \dots > \phi_p(L)$. Il montre que

$$\hat{R}_2(\hat{\Sigma}) = \sum_{i=1}^p \left\{ \left(n - p + 1 + 2l_i \sum_{j \neq i}^p \frac{1}{l_i - l_j} \right) \psi_i(L) + 2l_i \frac{\partial \psi_i(L)}{\partial l_i} - \log \psi_i(L) \right\} - \alpha_i$$

où pour $i = 1, \dots, p$,

$$\psi_i(L) = \phi_i(L)/l_i, \quad \alpha_i = \sum_{i=1}^p \frac{\Gamma\{(n-i+1)/2\}}{\Gamma\{(n-i+1)/2\}} + p \log 2 + p.$$

est un estimateur sans biais du risque *i.e* tel que,

$$R_2(\hat{\Sigma}, \Sigma) = E_{\theta, \Sigma}[\hat{R}_2(\hat{\Sigma})].$$

Alors, l'estimateur de Stein est obtenu en minimisant $\hat{R}_2(\hat{\Sigma})$ par rapport à $\psi_i(L)$. Il est donné par

$$\hat{\Sigma}^{St} = H\Phi^{St}(L)H^T \quad (1.3.3)$$

où, pour $i = 1, \dots, p$

$$\phi_i^{St}(L) = \frac{l_i}{\beta_i(L)}, \quad \beta_i(L) = n - p + 1 + 2l_i \sum_{j \neq i}^p \frac{1}{l_i - l_j}.$$

Ainsi, l'estimateur de Stein combine l'information rapportée par l'estimation de la i 'ème valeur propre de Σ par l_i avec une information supplémentaire donnée par les l_j pour $j \neq i$. Plus spécifiquement, il diminue les valeurs $a l_i$ pour i petit et les augmente pour i grand. Ainsi $\phi_1^{St}(L) \leq a l_1$ et $\phi_p^{St}(L) \geq a l_p$. Ce type d'estimateur est connue dans la littérature sous le nom de *shrinkage estimators* : il réduit l'étendue des estimations $a l_i$. Notons que, comme l'estimateur de maximum de vraisemblance correspond à $\phi_i^{Mv}(L) = l_i/n$, alors quand n est grand $\phi_i^{St}(L)$ est proche $\phi_i^{Mv}(L)$ et donc l'amélioration que propose l'estimateur de Stein par rapport à celui du maximum de vraisemblance est très modeste.

Bien qu'il s'agisse d'un estimateur amélioré, l'estimateur de Stein présente un comportement indésirable. En effet, certaines valeurs de $\phi_i^{St}(L)$ peuvent être négatives et ne pas préserver l'ordonnancement $\phi_1^{St}(L) > \dots > \phi_p^{St}(L)$. Pour y remédier, Stein suggère une procédure dite

d'isotonisation (*isotonizing procedure*) qui est expliquée en détail par Lin et Perlman [LP85]. Ces derniers montrent par des simulations de type Monte Carlo que, non seulement l'estimateur de Stein $\hat{\Sigma}^{St}$ isotonisé domine $\hat{\Sigma}_{1/n}$ mais aussi l'estimateur de type James-Stein $\hat{\Sigma}^{JS}$ dans (1.3.1). Plus tard, Dey et Srinivasan [DS85] proposent aussi un estimateur orthogonalement invariant de la forme

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

où, pour $i = 1, \dots, p$

$$\phi_i^{DS}(L) = \frac{l_i}{n + p + 1 - 2i},$$

et montrent que ce dernier améliore les estimateurs de la forme $\hat{\Sigma}_a$ sous $L_2(\Sigma, \hat{\Sigma})$.

Haff [Haf91] a obtenu un estimateur similaire à celui de Stein au travers d'une approche bayésienne avec la contrainte de maintenir l'ordre des estimateurs des valeurs propres. Notons que l'estimateur de Stein dans (1.3.3) est obtenu sous le coût de Stein $L_2(\Sigma, \Sigma)$ dans (1.2.3), et que les deux approches (celles de Haff et de Stein) sont difficiles à appliquer sous le coût quadratique $L_1(\Sigma, \Sigma)$ dans (1.2.2). De ce fait, il est courant de prendre des estimateurs améliorés sous $L_2(\Sigma, \hat{\Sigma})$ et les redimensionner pour $L_1(\Sigma, \hat{\Sigma})$ (cf. Haff [Haf91], Lin et Perlman [LP85]) au sens où si $\hat{\Sigma}$ est obtenu sous $L_2(\Sigma, \hat{\Sigma})$, alors on considère $n\hat{\Sigma}/(p+n+1)$ sous $L_1(\Sigma, \hat{\Sigma})$. Pour une discussion plus approfondie sur l'estimation bayésienne de la matrice de covariance, voir Dempster, Rubin et Tsutakawa [DRT81], Kubokawa, Robert et Saleh [KRS92], Berger et Yan [YB94], Berger, Strawderman et Tang [BST05]. Notons que, dans cette approche, il est nécessaire de choisir une loi à priori pour Σ ; celle qui est le plus souvent utilisée dans le cas gaussien est la distribution de Wishart inversée. Pour une discussion détaillée sur le choix des lois a priori dans ce contexte voir Alvarez, Niemi et Simpson [ANS14], Yang et Berger [YB18].

Haff [Haf80] a introduit une classe d'estimateurs inspirée de l'approche bayésienne empirique. Sa démarche est basée sur le choix d'une distribution Wishart comme loi a priori pour Σ^{-1} . Son estimateur est de la forme

$$\hat{\Sigma}^{Hf} = a[S + \nu t(\nu)C] \tag{1.3.4}$$

où, $\nu = 1/\text{tr}(S^{-1})$, C est une matrice définie positive arbitraire et $t(\nu)$ est une fonction décroissante positive et absolument continue. La motivation bayésienne de son approche sert à discuter les choix de C et la fonction $t(\nu)$. Haff montre que sous $L_2(\Sigma, \hat{\Sigma})$ dans (1.2.3), les estimateurs de la forme $\hat{\Sigma}^{Hf}$ améliorent les estimateurs usuels $\hat{\Sigma}_a$ lorsque $0 \leq t(\nu) \leq 2(p-1)/n$. Sous le coût $L_1(\Sigma, \hat{\Sigma})$ dans (1.2.2), il obtient un résultat similaire mais pour une classe d'estimateur réduite car il fixe $t(\nu)$ à une constante γ et obtient une amélioration pour $0 \leq \gamma \leq 2(p-1)/(n-p+3)$. Nous considérons aussi cette classe d'estimateurs dans le chapitre 3 et le chapitre 4 dans un cadre distributionnel plus général que le cas gaussien, le cas elliptique. Nous donnerons aussi un résultat plus général (une plus grande classe d'estimateurs) dans une approche unifiée des cas où $p > n$ et où $p \leq n$.

1.3.2 Le contexte en grande dimension

L'estimation de la matrice de covariance lorsque p est plus grand que n intervient dans de nombreuses applications comme dans le traitement de signal, en génomique ou dans les mathématiques financières (cf. Pourahmadi [Pou13] chapitre 2 pour des exemples détaillés). Rappelons que, dans ce cas, les estimateurs usuels de la forme $\hat{\Sigma}_a = aS$ sont mal conditionnés et non inversible (cf. la fin de la section 1.2). Ainsi, la construction d'estimateurs à la fois mieux conditionnés

et inversibles semble nécessaire. À cette fin, de nombreux estimateurs de type *shrinkage* ont été considérés dans la littérature. Celui de Ledoit et Wolf [LW04] est une combinaison convexe de $\hat{\Sigma}_{1/(n-1)}$ et d'une matrice définie positive T . Il est de la forme

$$\hat{\Sigma}^{\text{LW}} = (1 - \lambda) \hat{\Sigma}_{1/(n-1)} + \lambda T, \quad (1.3.5)$$

où $\lambda \in [0, 1]$ est l'intensité avec laquelle on réduit (*one shrinks*) l'estimateur empirique et on le rapproche de la matrice cible T , supposée proportionnelle à l'identité, soit $T = \nu I_p$, où $\nu > 0$. Ainsi, $\lambda = 0$ n'implique aucune contribution de T , tandis que $\lambda = 1$ n'implique aucune contribution de $\hat{\Sigma}_{1/(n-1)}$. Les valeurs intermédiaires pour λ révèlent la contribution simultanée de $\hat{\Sigma}_{1/(n-1)}$ et T . Le couple (λ, ν) optimal qui minimise le risque associé au coût $\|\hat{\Sigma}^{\text{LW}} - \Sigma\|_F^2$ est donné par

$$\lambda = \frac{E[\|\hat{\Sigma}_{1/(n-1)} - \Sigma\|_F^2]}{E[\|\hat{\Sigma}_{1/(n-1)} - T\|_F^2]} \quad \text{et} \quad \nu = \frac{\text{tr}(\Sigma)}{p},$$

ce qui ne conduit pas à un estimateur puisque dépendant de Σ . Un remède consiste à estimer λ et ν . Notons que, dans leur approche, Ledoit et Wolf ne supposent aucune hypothèse distributionnelle. L'intensité optimale λ suggère de combien nous devons *shrinker* $\hat{\Sigma}_{1/(n-1)}$ vers la matrice cible T . Ledoit et Wolf estiment ν et λ par des estimateurs consistants mais biaisés. Chen, Wiesel et Hero [CWH09] améliorent l'estimateur de Ledoit et Wolf dans le cas gaussien au travers du théorème de Rao-Blackwell. Schäfer et Strimmer [SS05] proposent des estimateurs sans biais mais pas consistants pour ν et λ . Fisher et Sun [FS11] montrent que, dans le cas gaussien, leur estimateur améliore ceux cités précédemment. De plus, ils considèrent trois matrices cible différentes T dans (1.3.5), à savoir, νI_p , $\text{diag}(\hat{\Sigma}_{1/(n-1)})$ et I_p . De manière similaire, un *shrinkage estimator* de la forme (1.3.5) avec une matrice cible $T = \text{diag}(\hat{\Sigma}_{1/(n-1)})$ est étendu par Ikeda, Kubokawa et Srivastava [IKS16] au cas où aucune distribution n'est spécifiée.

Konno [Kon09] étend l'estimateur de Haff (le cas particulier lorsque $t(\nu)$ est constant) dans (1.3.4) au contexte de grande dimension. Son estimateur est de la forme

$$\hat{\Sigma}^K = a [S + \nu t S S^+], \quad (1.3.6)$$

où $\nu = 1/\text{tr}(S^+)$, t est une constante positive, S^+ est l'inverse de Moore-Penrose (cf. définition A.1) de S . En fait, il construit son estimateur en prenant dans (1.3.4) la fonction $t(\nu) = t$ et $C = S S^+$. De ce fait, son estimateur est orthogonalement invariant (de la forme (1.3.2)) puisqu'en utilisant la décomposition en valeurs propres de S (dans le cas non inversible où $p > n$); à savoir

$$S = H_1 L H_1^T$$

où $L = \text{diag}(l_1, \dots, l_n)$ soit la matrice des valeurs propres non nulles où $l_1 > \dots > l_n > 0$ et H_1 soit la matrice semi-orthogonale des vecteurs propres associés, son estimateur peut s'écrire sous la forme

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

où, $\Phi^K(L) = \text{diag}(\phi_1^K(L), \dots, \phi_n^K(L))$ avec $\phi_i^K(L) = a(l_i + \nu t)$. De ce fait, il ne corrige que les n valeurs propres positives de S puisque les autres, les $(p - n)$ restantes, sont nulles. Konno montre que, sous le coût quadratique $L_1(\Sigma, \hat{\Sigma})$, son estimateur améliore les estimateurs usuels $\hat{\Sigma}_a$ pour $0 < t \leq 2(n-1)(p-n-1)/(p-n+1)(p-n+3)$. Nous verrons dans le chapitre 3 et le chapitre 4, que cet estimateur peut être étendu au cas elliptique.

1.3.3 Le contexte unifié

Une approche unifiée ($p > n$ ou $p \leq n$) pour l'estimation de la matrice d'échelle dans le cadre gaussien est proposée par Tsukuma et Kubokawa [TK16]. Dans le cas non inversible ($p > n$), le coût de Stein $L_2(\Sigma, \hat{\Sigma})$ n'est pas défini car la matrice $\hat{\Sigma}\Sigma^{-1}$ a des valeurs propres nulles. Par conséquent, une alternative raisonnable dans une telle approche est le coût de Stein modifié

$$L_2^M(\Sigma, \hat{\Sigma}) = \text{tr}(\Lambda(\hat{\Sigma}\Sigma^{-1})) - \log |\Lambda(\hat{\Sigma}\Sigma^{-1})| - p \wedge n$$

tel que, $\Lambda(\hat{\Sigma}\Sigma^{-1})$ est la matrice diagonale des $p \wedge n$ valeurs propres strictement positives de $\hat{\Sigma}\Sigma^{-1}$. Notons que, dans le contexte en petite dimension ($p \leq n$), on a $L_2^M(\Sigma, \hat{\Sigma}) = L_2(\Sigma, \hat{\Sigma})$ puisque $\hat{\Sigma}\Sigma^{-1}$ admet p valeurs propres strictement positives. Sous un coût similaire, Tsukuma et Kubokawa montrent que l'estimateur (une adaptation au cas unifié) de James-Stein $\hat{\Sigma}^{JS}$ dans (1.3.1) améliore les estimateurs usuels en améliorant l'optimal sous ce coût, soit $\hat{\Sigma}_{1/p \vee n}$. De plus, ils montrent aussi que $\hat{\Sigma}^{JS}$ est dominé par un estimateur tronqué en utilisant les informations disponibles dans la matrice moyenne de l'échantillon. Les propriétés de dominance des estimateurs tronqués par rapport aux estimateurs non tronqués ont été étudiées par Kubokawa et Srivastava [KS03], Kubokawa et Tsai [KT06].

Chételat et Wells [CW16] s'intéressent au cas où la matrice de covariance Σ est non inversible ; $\text{rang}(\Sigma) = r < n \wedge p$. Ils montrent que, pour $1 < r < n - 4$, l'estimateur de type Konno dans (1.3.6)

$$\hat{\Sigma}^K = a [S + \nu t S S^+]$$

améliore les estimateurs usuels pour $0 < t \leq r - 1/n - r + 2$, sous le coût quadratique modifié

$$L_1^M(\Sigma, \hat{\Sigma}) = \text{tr}(\hat{\Sigma}\Sigma^+ - I_p)^2$$

où Σ^+ est l'inverse de Moore-Penrose de Σ . Notons que cet estimateur est aussi orthogonalement invariant puisqu'il peut être écrit sous forme (1.3.2).

1.4 Objectif et structure de la thèse

La plupart des travaux existants dans la littérature, qui abordent l'estimation de la matrice d'échelle, ont été faits dans le cadre gaussien (il s'agit donc de la matrice de covariance). Néanmoins, il existe quelques articles traitant de l'estimation de la matrice d'échelle des distributions à symétrie elliptique. Kubokawa et Srivastava [KS99] montrent que, dans le contexte en petite dimension, les résultats de dominance obtenus dans le cas gaussien par Stein [Ste75], [Ste77], [Ste86] restent valides dans le cas elliptique. En effet, ils montrent que, sous le coût de Stein $L_2(\Sigma, \hat{\Sigma})$ dans (1.2.3), l'estimateur $\hat{\Sigma}^{JS}$ dans (1.3.1) améliore les estimateurs usuels de la forme $\hat{\Sigma}_a$. Plus tard, Konno [Kon07] étend ce résultat au cas des distributions à symétrie elliptique complexes. Cependant, des résultats d'améliorations relative au coût quadratique $L_1(\Sigma, \hat{\Sigma})$ dans (1.2.2) n'ont pas encore été établis à notre connaissance.

L'objectif de cette thèse est d'améliorer, sous un coût quadratique et sous une modification de celui-ci, les estimateurs usuels dans (1.2.1) de la matrice d'échelle des distributions à symétrie elliptique du point de vue théorie de la décision. Nous proposons une approche unifiée où nous traitons à la fois le contexte en petite dimension et le contexte en grande dimension. Le problème est formalisé comme suit.

Soit Y une matrice $m \times p$ observée avec $p < m$. Nous considérons le modèle additif

$$Y = M + \mathcal{E}, \quad \mathcal{E} \sim ES(0, I_m \otimes \Sigma), \quad (1.4.1)$$

où M est une matrice de paramètres inconnus, \mathcal{E} un bruit de distribution à symétrie elliptique, avec une matrice de covariance proportionnelle à $I_m \otimes \Sigma$ telle que Σ soit une matrice $p \times p$ inversible que l'on cherche à estimer. On suppose que M est une matrice de rang inférieur, à savoir,

$$\text{rank}(M) = q < p.$$

En pratique, cette hypothèse sur le rang est souvent réalisée. Par exemple, dans le contexte du débruitage vidéo, les colonnes de M peuvent être des images individuelles dans une séquence vidéo, qui sont généralement très corrélées (*cf.* Ji et al. [JLSX10a]). En traitement d'image, où l'intérêt est d'estimer l'image bruitée M , ce modèle a été considéré par Canu et Fourdrinier [CF17] et représente une extension du modèle gaussien de Candès et al. [CST13], où $\mathcal{E} \sim \mathcal{N}(0, I_m \otimes \Sigma)$ avec une matrice de covariance connue. Plus généralement, le modèle (1.4.1) est une représentation alternative du modèle de régression multivarié $Y = X\beta + \mathcal{E}$ où X est un $m \times q$ matrice de constantes connues de rang $q < p$ et β est une matrice $q \times p$ de paramètres inconnus. Bien que tout estimateur de Σ soit fonction de la matrice d'observation Y (comme les estimateurs usuels $aY^T Y$), cette hypothèse sur le rang permet de réduire essentiellement les informations dans Y (nécessaire pour l'estimation de Σ) en une matrice U de plus petite dimension, soit une matrice $n \times p$ où $n = m - q$, afin que $aU^T U$ estime Σ . Plus précisément, l'objectif de cette thèse est d'améliorer les estimateurs usuels

$$\hat{\Sigma}_a = aS = aU^T U$$

où $a > 0$, par des estimateurs de la forme

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

où $SS^+ G(Z, S)$ est une matrice de correction, dans une approche unifiée ($p \leq n$ ou $p > n$). La qualité des estimateurs est évaluée sous le coût quadratique

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

et sous une modification de ce dernier, soit le coût basé sur les données

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

À cette fin, dans le [chapitre 2](#), nous donnons quelques définitions et propriétés des distributions à symétrie elliptique (et sphérique). Nous donnons aussi la représentation canonique du modèle additif (1.4.1) qui est utile pour réduire la dimension des observations puisque $aU^T U$ va estimer Σ au lieu de aYY^T . Par la suite, la difficulté des calculs va nous conduire à définir une classe de lois particulière dite de type Berger (*cf.* [section 2.6](#)). Enfin, le résultat principal de ce chapitre est l'identité de type Stein-Haff développée dans la [section 2.7](#). Celle-ci est un outil majeur pour la construction d'estimateurs améliorés de Σ . À contrario des identités existantes dans la littérature, nous verrons que celle-ci traite à la fois le cas où S est inversible (contexte en petite dimension) et le cas où est non inversible (contexte en grande dimension) et aussi, elle est donnée dans le cadre général des distributions à symétrie elliptique.

Dans le [chapitre 3](#), sous ce coût quadratique, nous construisons des estimateurs améliorés de la forme général $a(S + SS^+ G(Z, S))$ par rapport aux estimateurs habituels aS , dans la sous-classe de lois de type Berger. Les conditions d'amélioration sont données dans le [théorème 3.2.1](#). Par la suite, les résultats sur la structure en valeurs propre de S dans le [lemme B.1](#) nous permettront de

fournir des exemples généraux d'estimateurs améliorés orthogonalement invariants de la forme (1.3.2). Ensuite nous donnons des estimateurs améliorés de type Haff dans (1.3.4), un exemple pour ces derniers et nous traitons le cas particulier de l'estimateur de Konno dans (1.3.6).

Dans le chapitre 4, sous ce coût basé sur les données, nous construisons, aussi, des estimateurs améliorés de la forme $a(S + SS^+ G(S))$ par rapport aux estimateurs habituels aS , où nous considérons, pour l'estimateur de type Konno (1.3.6), une plus large classe de lois que celle dans le chapitre 3. Nous verrons, aussi, que notre théorie, développée dans la sous-classe de type Berger dans la section 2.6, s'applique à la classe des estimateurs orthogonalement invariants de la forme (1.3.2) et permet de construire une plus grande classe d'estimateur que celle dans le chapitre précédent. Ensuite nous fournissons, aussi, une adaptation de l'estimateur de Haff dans (1.3.4) et le cas particulier correspondant à l'estimateur de Konno dans (1.3.6) au cas elliptique.

Par la suite, dans le chapitre 5, nous expliquons par un exemple en finance l'utilité d'estimer la matrice d'échelle. Nous montrons aussi comment simuler des distributions à symétrie elliptique. Ensuite, nous discutons par des simulations la performance de l'estimateur de type Konno pour deux structures différentes de la matrice d'échelle.

Enfin, nous concluons dans le chapitre 6 par une synthèse des résultats présentés avant d'aborder des nouvelles problématiques, possibilités et pistes, que ce travail nous suggère.

LE CADRE DES LOIS À SYMÉTRIE ELLIPTIQUE

Sommaire

2.1	Introduction	11
2.2	Lois à symétrie sphérique	12
2.3	Lois à symétrie elliptique	14
2.4	Lois à symétrie elliptique vectorisée	17
2.5	Le modèle additif	18
2.5.1	Représentation canonique	19
2.5.2	Densité	20
2.6	Une sous-classe de lois à symétrie elliptique de type Berger	21
2.7	Une identité de type Stein-Haff	24
2.8	Synthèse	27

2.1 Introduction

Les distributions à symétrie elliptique sont souvent considérées dans de nombreux domaines comme la finance, la théorie de la communication ou le traitement de signal. Plusieurs exemples pratiques sont donnés dans Chmielewski [Chm81]. En particulier, dans la théorie de la gestion de portefeuille, développée par Markowitz [Mar52], des études empiriques continuent de démontrer l'inadéquation de l'hypothèse de normalité pour la modélisation des rendements des actifs, car les valeurs extrêmes des rendements sont mal captées par celle-ci en raison de ses queues légères. Ainsi, la famille des distributions à symétrie elliptique, qui contient la distribution gaussienne, s'avère une alternative intéressante puisqu'elle offre la possibilité de construire des distributions possédant des queues lourdes qui captent mieux ces valeurs extrêmes (*cf.* Gupta, Varga et Bodnar [GVB13]).

Dans ce chapitre, nous commençons d'abord par définir les distributions à symétrie sphérique. Celles-ci peuvent être un moyen pour définir les distributions à symétrie elliptique. Nous donnerons ensuite la représentation canonique du modèle additif (1.4.1) et nous définissons une sous-classe de lois à symétrie elliptique de type Berger et nous donnons des exemples de densités

dans cette sous-classe. Enfin, nous développons une nouvelle identité de type Stein-Haff, qui est, dans notre approche, un outil essentiel pour la construction d'estimateurs améliorés de la matrice d'échelle Σ .

2.2 Lois à symétrie sphérique

Il existe plusieurs manières différentes pour définir les distributions à symétrie sphérique. En effet, ceci peut être fait au travers des fonctions caractéristiques, mais cette approche n'est pas intuitive contrairement à celle que nous donnons dans la [définition 2.2.3](#) (tirée de Fourdrinier, Strawderman et Wells [[FSW18](#)]). Nous renvoyons le lecteur intéressé à Kelker [[Kel70](#)] pour une revue historique des distributions à symétrie sphérique et à Fang, Kotz et Ng [[FKN90](#)] pour une présentation plus complète. La définition que nous donnons est basée sur la notion de transformation orthogonale suivante.

Définition 2.2.1 (Transformation orthogonale)

Soit $x \in \mathbb{R}^p$ et $y \in \mathbb{R}^p$. Une transformation H est orthogonale si elle préserve le produit scalaire classique $\langle x, y \rangle = x^T y$, i.e.

$$\langle Hx, Hy \rangle = x^T H^T H y = \langle x, y \rangle,$$

c'est-à-dire $H \in \mathcal{O}_p$.

Cette définition nous servira à donner une propriété importante des vecteurs aléatoires qui est l'invariance orthogonale.

Définition 2.2.2 (Invariance orthogonale d'un vecteur aléatoire)

On dit qu'un vecteur aléatoire $X \in \mathbb{R}^p$ est invariant par transformation orthogonale H si

$$HX \stackrel{d}{=} X.$$

Cette propriété d'invariance orthogonale est utilisée ici pour définir les distributions à symétrie sphérique.

Définition 2.2.3 (Distribution à symétrie sphérique)

Le vecteur aléatoire $X \in \mathbb{R}^p$ a une distribution à symétrie sphérique autour de $\theta \in \mathbb{R}^p$, si le vecteur aléatoire $X - \theta$ est invariant pour toute transformation orthogonale H . On note alors

$$X \sim SS(\theta).$$

On peut aussi définir les lois à symétrie sphérique par leurs représentations stochastiques.

Définition 2.2.4 (Représentation stochastique)

Un vecteur aléatoire $X \in \mathbb{R}^p$ a une distribution à symétrie sphérique autour de θ si et seulement si

$$X \stackrel{d}{=} \theta + RU, \quad R = \|X - \theta\| \geq 0, \quad U = \frac{X - \theta}{\|X - \theta\|} \quad (2.2.1)$$

où, R est le rayon de X et U , sa direction, suit une distribution uniforme sur

$$\mathcal{S}_1 = \{x \in \mathbb{R}^p; \|x - \theta\|^2 = 1\}$$

la sphère de rayon unitaire dans \mathbb{R}^p . Les variables aléatoires U et R sont indépendantes.

Nous pouvons voir dans cette représentation que le vecteur aléatoire $U \in \mathbb{R}^p$ suit la même distribution, uniforme, quelle que soit la distribution de X . Cette propriété est utilisée pour simuler des distributions sphériques (cf. section 5.2). Quant à la distribution du rayon R , celle-ci est liée à la densité de X , si elle existe, comme on peut le voir dans le théorème suivant (cf. Fourdrinier, Strawderman et Wells [FSW18], chapitre 1).

Théorème 2.2.1

Soit $X \in \mathbb{R}^p$ un vecteur aléatoire de distribution à symétrie sphérique autour de $\theta \in \mathbb{R}^p$. Alors les deux affirmations suivantes sont équivalentes.

- (1) $R = \|X - \theta\|$ a une densité $h(\cdot)$ par rapport à la mesure de Lebesgue dans \mathbb{R}_+ .
- (2) X a une densité $f(\cdot)$ par rapport à la mesure de Lebesgue dans \mathbb{R}^p .

De plus, si (1) ou (2) est vérifiée, alors il existe une fonction $g(\cdot)$ de \mathbb{R}_+ à \mathbb{R}_+ , tel que

$$f(x) = g(\|x - \theta\|^2) \quad \text{presque partout.}$$

et

$$h(r) = \frac{2\pi^{p/2}}{\Gamma(\frac{p}{2})} r^{p-1} g(r^2) \quad \text{presque partout.}$$

La fonction $g(\cdot)$ est appelée fonction générative et $h(\cdot)$ densité radiale.

La forme quadratique $\|x - \theta\|^2$ dans la densité $f(x)$ est une mesure statistique de distance appelée distance euclidienne. Elle mesure l'écart des variables par rapport à la moyenne. En outre, le nom de distribution sphérique vient du fait que

$$\|x - \theta\|^2 = (x - \theta)^T(x - \theta) = r^2$$

est l'équation de la sphère centrée en θ et de rayon r .

Nous allons voir maintenant les conditions d'existence de la moyenne et de la matrice de covariance des distributions à symétrie sphérique (cf. Fourdrinier, Strawderman et Wells [FSW18], chapitre 1).

Théorème 2.2.2

Soit $X \in \mathbb{R}^p$ un vecteur aléatoire de distribution à symétrie sphérique autour de $\theta \in \mathbb{R}^p$. Alors, la moyenne de X existe si et seulement si la moyenne de $R = \|X - \theta\|$ existe. Dans ce cas, on note

$$E(X) = \theta.$$

La matrice de covariance de X existe si et seulement si $E(R^2)$ est finie. Dans ce cas on note

$$\text{Cov}(X) = \frac{E(R^2)}{p} I_p.$$

Parmi les lois à symétrie sphérique, on peut citer la distribution gaussienne, la distribution de Student, la distribution de Kotz, la distribution de puissance exponentielle et la distribution logistique sphérique, etc. Le tableau 2.1, tirée de Fang, Kotz et Ng [FKN90], résume les densités de ces exemples où $\text{Cov}(X) = \sigma^2 I_p$.

Loi	Densité	Paramètres
Normal, $\mathcal{N}_p(\theta, \sigma^2)$	$\frac{1}{(2\pi\sigma^2)^{\frac{p}{2}}} e^{-\frac{\ x-\theta\ ^2}{2\sigma^2}}$	
Student, $t_p(\theta, \sigma^2, \nu)$	$\frac{\Gamma(\frac{p+\nu}{2})}{(\pi\sigma^2\nu)^{\frac{p}{2}}\Gamma(\frac{\nu}{2})} \left(1 + \frac{\ x-\theta\ ^2}{\nu\sigma^2}\right)^{-\frac{(\nu+p)}{2}}$	$\nu \geq 1$
Kotz, $\mathcal{K}_p(\theta, \sigma^2, N, r)$	$\frac{\pi^{-\frac{p}{2}}\Gamma(\frac{p}{2})r^{\frac{2N+p-2}{2}}}{\sigma^{2N+p-1}\Gamma(\frac{2N+p-2}{2})} \ x-\theta\ ^{2(N-1)} e^{-r\frac{\ x-\theta\ ^2}{\sigma^2}}$	$N > \frac{2-p}{2}, r > 0$
Puis.Exp, $\mathcal{PE}_p(\theta, \sigma^2, b)$	$\frac{p\Gamma(\frac{p}{2})}{(\pi\sigma^2)^{\frac{p}{2}}2^{1+\frac{1}{2b}}\Gamma(1+\frac{p}{2b})} e^{-\frac{1}{2}\left(\frac{\ x-\theta\ ^2}{\sigma^2}\right)^b}$	$b > 0$
Logistique, $\mathcal{LOG}_p(\theta, \sigma^2)$	$\frac{\left(\sum_{j=1}^{\infty} (-1)^{j-1} j^{1-\frac{p}{2}}\right)^{-1}}{(2\pi\sigma^2)^{\frac{p}{2}}} \frac{e^{-\frac{\ x-\theta\ ^2}{2\sigma^2}}}{\left(1 + e^{-\frac{\ x-\theta\ ^2}{2\sigma^2}}\right)^2}$	

TABLEAU 2.1 – Exemples de densités à symétrie sphérique.

2.3 Lois à symétrie elliptique

Nous avons défini dans la section précédente les distributions à symétrie sphérique grâce à la propriété d'invariance orthogonale associée au produit scalaire classique $\langle x, y \rangle = x^T y$. De manière équivalente, nous allons définir les distributions à symétrie elliptique grâce à la propriété d'invariance orthogonale associée au produit scalaire

$$\langle x, y \rangle_{\Sigma^{-1}} = x^T \Sigma^{-1} y,$$

où, Σ est une matrice symétrique est inversible (définie positive).

Définition 2.3.1 (Transformation Σ^{-1} -orthogonale)

Soit $x \in \mathbb{R}^p$, $y \in \mathbb{R}^p$ et soit Σ une matrice $p \times p$ définie positive. Une transformation H est Σ^{-1} -orthogonale si elle préserve le produit scalaire $\langle x, y \rangle_{\Sigma^{-1}}$, i.e.

$$\langle Hx, Hy \rangle_{\Sigma^{-1}} = x^T H^T \Sigma^{-1} H y = \langle x, y \rangle_{\Sigma^{-1}}$$

c'est à dire $H^T \Sigma^{-1} H = \Sigma^{-1}$.

Définition 2.3.2 (Invariance Σ^{-1} -orthogonale d'un vecteur aléatoire)

Soit Σ une matrice $p \times p$ définie positive. Un vecteur aléatoire $Y \in \mathbb{R}^p$ est invariant par transformation Σ^{-1} -orthogonale H , si et seulement si

$$HY \stackrel{d}{=} Y.$$

Il est intéressant de noter que cette propriété est utilisée dans la représentation canonique du modèle additif défini dans (1.4.1) tout en maintenant inchangée sa distribution (cf. section 2.5.2).

Nous pouvons maintenant définir les distributions à symétrie elliptique.

Définition 2.3.3 (Distribution à symétrie elliptique)

Soit Σ une matrice $p \times p$ définie positive. Un vecteur aléatoire $Y \in \mathbb{R}^p$ a une distribution à symétrie elliptique autour de $\theta \in \mathbb{R}^p$ si le vecteur $Y - \theta$ est invariant pour toute transformation Σ^{-1} -orthogonale H . On note alors

$$Y \sim ES(\theta, \Sigma).$$

Cette dernière définition, donnée par Fourdrinier, Strawderman et Wells [FSW18], est à mettre en parallèle avec la [définition 2.2.3](#) des distributions à symétrie sphérique de sorte que, la notion d'invariance orthogonale associée au produit scalaire standard $\langle x, y \rangle = x^T y$ pour $x \in \mathbb{R}^p$ et $y \in \mathbb{R}^p$ est généralisé à la notion d'invariance Σ^{-1} -orthogonale associée au produit scalaire $\langle x, y \rangle_{\Sigma^{-1}} = x^T \Sigma^{-1} y$. Par conséquent, cette approche nécessite que Σ soit inversible, ce qui n'est pas une condition nécessaire pour définir les distributions à symétrie elliptique. En effet, on peut les définir, sans supposer l'inversibilité de Σ , au travers des fonctions caractéristiques (cf. Fang et Zang [FKN90]) ou en utilisant la transformation linéaire des lois à symétrie sphérique suivante :

$$Y = \theta + A^T X, \quad X \sim SS(0) \quad (2.3.1)$$

où, A est une matrice $k \times p$ avec $k \leq p$, $X \in \mathbb{R}^k$ et $\Sigma = A^T A$ est de rang inférieur ($\text{rang}(\Sigma) = k$) qui est inversible que lorsqu'elle est de plein rang ($\text{rang}(\Sigma) = p$). Dans ce dernier cas, il est facile de voir à partir de (2.3.1) que, pour $A = \Sigma^{1/2}$ alors $\Sigma^{-1/2} Y \sim SS(\Sigma^{-1/2} \theta)$ et pour $A = I_p$ alors $Y \sim SS(\theta)$. Rappelons que dans ce travail la matrice d'échelle Σ est inversible.

Définition 2.3.4 (Représentation stochastique)

Un vecteur aléatoire $Y \in \mathbb{R}^p$ a une distribution symétrie elliptique autour de θ si et seulement si

$$Y \stackrel{d}{=} \theta + R U, \quad R = \|Y - \theta\|_{\Sigma^{-1}} \geq 0, \quad U = \frac{Y - \theta}{\|Y - \theta\|_{\Sigma^{-1}}}$$

où, R est le rayon de Y et sa direction U suit une distribution uniforme sur

$$\mathcal{S}_1^{\Sigma^{-1}} = \{x \in \mathbb{R}^p; \|x - \theta\|_{\Sigma^{-1}}^2 = 1\}$$

le Σ^{-1} -ellipsoïde de rayon unitaire dans \mathbb{R}^p . Les variables aléatoires U et R sont indépendantes.

Comme pour les distributions à symétrie sphérique, la densité de R est liée à celle de Y comme on peut le voir dans le théorème suivant (cf. Fourdrinier, Strawderman et Wells [FSW18], chapitre 1).

Théorème 2.3.1

Soit $Y \in \mathbb{R}^p$ un vecteur aléatoire de distribution à symétrie elliptique autour de $\theta \in \mathbb{R}^p$. Alors les deux affirmations suivantes sont équivalentes :

- (1) $R = \|Y - \theta\|_{\Sigma^{-1}}$ a une densité $h(\cdot)$ par rapport à la mesure de Lebesgue dans \mathbb{R}_+ .
- (2) Y admet une densité $f(\cdot)$ par rapport à la mesure de Lebesgue dans \mathbb{R}^p .

De plus, si (1) ou (2) est vérifiée, alors

$$f(y) = |\Sigma|^{-1/2} g(\|y - \theta\|_{\Sigma^{-1}}^2) \quad \text{presque partout.}$$

et

$$h(r) = \frac{2 \pi^{p/2}}{\Gamma(p/2)} |\Sigma|^{-1/2} r^{p-1} g(r^2) \quad \text{presque partout.}$$

où $g(\cdot)$ est appelée fonction générative et $h(\cdot)$ densité radiale.

La forme quadratique $\|y - \theta\|_{\Sigma^{-1}}^2$ dans la densité $f(y)$ est une mesure statistique de distance, appelée distance de Mahalanobis [Mah36] qui correspond à la distance euclidienne lorsque $\Sigma = I_p$. En effet, elle donne la distance des variables du vecteur Y par rapport à θ en tenant compte du fait que les variances peuvent être différentes et que les variables peuvent être corrélées. À la différence de la distance euclidienne où toutes les composantes des vecteurs sont traitées de la même façon, la distance de Mahalanobis accorde un poids moins important aux composantes les plus dispersées. Ceci peut être pratique, par exemple, dans le traitement de signal où chaque composante est une variable aléatoire de type gaussien. Cela revient à minimiser l'influence des composantes les plus bruitées (celles ayant des grandes variances). En outre, le nom de distribution elliptique est dû au fait que

$$\|y - \theta\|_{\Sigma^{-1}}^2 = (y - \theta) \Sigma^{-1} (y - \theta)^T = r^2$$

est l'équation de l'ellipsoïde (qui correspond à la sphère lorsque Σ est proportionnel à I_p) centré en θ (cf. section 5.2.1), dont les axes sont dans la direction des vecteurs propres de Σ .

Il convient de noter que, lorsque Σ n'est pas inversible, le vecteur Y n'admet pas de densité par rapport à la mesure de Lebesgue. Dans ce cas, on parle de distributions à symétrie elliptique singulière qui admettent des densités par rapport à la mesure de Hausdorff (cf. Billingsley [Bil95]). Nous renvoyons les lecteurs intéressés à Díaz-García, Leiva-Sánchez et Galea [DGLSG02] pour une définition formelle des lois elliptiques singulières et à Uhlig [Uhl94], Srivastava [Sri03] pour des exemples.

Maintenant, nous allons voir les conditions d'existence de la moyenne et de la matrice de covariance des distribution à symétrie elliptique (cf. Fourdrinier, Strawderman et Wells [FSW18], chapitre 1).

Théorème 2.3.2

Soit $Y \in \mathbb{R}^p$ un vecteur aléatoire de distribution à symétrie elliptique autour de $\theta \in \mathbb{R}^p$. Alors, la moyenne de Y existe si et seulement si la moyenne de $R = \|Y - \theta\|_{\Sigma^{-1}}$ existe. Dans ce cas, on note

$$E(Y) = \theta.$$

La matrice de covariance de Y existe si et seulement si $E(R^2)$ est fini. Dans ce cas, on note

$$\text{Cov}(Y) = \frac{E(R^2)}{p} \Sigma.$$

Notons que, dans le cas gaussien, R^2 suit une distribution Khi-deux dont l'espérance est égale à p et donc la matrice de covariance de Y coïncide avec la matrice d'échelle Σ .

Le théorème suivant (cf. Fang, Kotz et Ng [FKN90], chapitre 2) indique que toute combinaison linéaire d'une distribution à symétrie elliptique est aussi une distribution à symétrie elliptique.

Théorème 2.3.3

Soit $Y \in \mathbb{R}^p$ un vecteur aléatoire de distribution à symétrie elliptique autour de $\theta \in \mathbb{R}^p$, et soit, $B \in \mathbb{R}^{p \times k}$, $V \in \mathbb{R}^{k \times 1}$. Alors

$$V + B^T Y \sim ES(V + B^T \theta, B^T \Sigma B).$$

Le corollaire de ce théorème donne les distributions marginales de Y qui sont également elliptiquement distribués.

Corollaire 2.3.1

Soit $Y \sim ES(\theta, \Sigma)$. On partitionne Y, θ et Σ comme suit :

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}, \quad \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

où, $Y_1 \in \mathbb{R}^k$, $\theta_1 \in \mathbb{R}^k$, $\Sigma_{11} \in \mathbb{R}^{k \times k}$, $Y_2 \in \mathbb{R}^{p-k}$, $\theta_2 \in \mathbb{R}^{p-k}$ et $\Sigma_{22} \in \mathbb{R}^{(p-k) \times (p-k)}$. Alors

$$Y_1 \sim ES(\theta_1, \Sigma_{11}), \quad Y_2 \sim ES(\theta_2, \Sigma_{22}).$$

Cette propriété joue un rôle majeur dans l'estimation du vecteur moyenne θ , la matrice d'échelle Σ ou son inverse (la matrice de précision) Σ^{-1} du modèle additif (1.4.1). En effet, en passant à la représentation canonique de ce modèle $Y = (Z, U)$ il est clair d'après le corollaire précédent que U suit une loi elliptique et la connaissance de la distribution de U facilite la construction d'estimateurs améliorés car elle intervient dans l'identité de type Stein-Haff qui est un outil essentiel dans ce contexte.

2.4 Lois à symétrie elliptique vectorisée

Dans la section précédente nous avons défini les distributions à symétrie elliptique dans le cas vectoriel. Rappelons que dans cette thèse, nous cherchons à estimer la matrice d'échelle en se basant sur une matrice d'observation $Y \in \mathbb{R}^{m \times p}$ de m -réalisations d'un vecteur aléatoire elliptique $Y \in \mathbb{R}^p$. Ainsi nous avons besoin d'étendre le concept de distribution à symétrie elliptique vectoriel au cas matriciel. Comme dans le cas vectoriel, il existe plusieurs manières différentes pour définir les distributions à symétrie elliptique matricielle (représentation stochastique, fonction génératrice, invariance orthogonale) mais contrairement au cas vectoriel, chaque approche mène à une classe de distribution différente (cf. Fang et Zhang [FZ90]). Nous nous intéresserons ici à celle appelée par Fang et Zhang classe des distributions elliptiques vectorisées. Elle est basée sur la vectorisation $A_v = \text{vec}(A^T)$ où $A \in \mathbb{R}^{m \times p}$ et $A_v \in \mathbb{R}^{pm}$.

Définition 2.4.1 (Distribution à symétrie elliptique vectorisée)

Soit Σ une matrice $p \times p$ définie positive. Une matrice $Y \in \mathbb{R}^{m \times p}$ a une distribution à symétrie elliptique vectorisée autour de la matrice $M \in \mathbb{R}^{m \times p}$ si et seulement si, pour toute transformation $(I_m \otimes \Sigma^{-1})$ -orthogonale $Q \in \mathbb{R}^{pm \times pm}$, on a

$$Q(Y_v - M_v) \stackrel{d}{=} (Y_v - M_v).$$

On note dans ce cas

$$Y \sim ES(M, I_m \otimes \Sigma).$$

Un cas particulier important, qui sera crucial dans la représentation canonique du modèle additif, survient lorsque $M = (\theta^T, 0^T)$, avec θ une matrice $n \times q$ et 0 la matrice nulle $(m - q) \times p$. Notons que

$$(\theta^T, 0^T) = \begin{pmatrix} \theta \\ 0 \end{pmatrix}.$$

Le théorème suivant, tiré de Canu et Fourdrinier [CF17], donne la densité de Y .

Théorème 2.4.1

Soit $Y \in \mathbb{R}^{m \times p}$ ayant une distribution elliptique vectorisée autour d'une matrice $M \in \mathbb{R}^{m \times p}$. Alors les deux affirmations suivantes sont équivalentes :

- (1) $R = \|Y - M\|_{\Sigma^{-1}}$ a une densité $h(\cdot)$ par rapport à la mesure de Lebesgue dans \mathbb{R}_+ .
- (2) Y admet une densité $f(\cdot)$ par rapport à la mesure de Lebesgue dans \mathbb{R}^{pm} .

De plus, si (1) ou (2) est vérifiée, alors

$$f(y) = |\Sigma|^{-m/2} g(\|y - M\|_{\Sigma^{-1}}^2) \quad p.p.$$

et

$$h(r) = \frac{2 \pi^{pm/2}}{\Gamma(pm/2)} r^{pm-1} |\Sigma|^{-m/2} g(r^2) \quad p.p.$$

où $g(\cdot)$ est appelée fonction générative et $h(\cdot)$ densité radiale.

Comme pour les distributions à symétrie elliptique dans \mathbb{R}^p , la moyenne de Y existe si et seulement si la moyenne de $R = \|Y - M\|_{\Sigma^{-1}}$ existe. Dans ce cas en note

$$E(Y) = M.$$

Aussi, la covariance de Y existe si et seulement si $E(R^2)$ est finie. Dans ce cas elle est proportionnelle à $I_m \otimes \Sigma$, i.e.

$$Cov(Y) = \frac{E(R^2)}{pm} I_m \otimes \Sigma. \quad (2.4.1)$$

Il convient de noter que, le lien entre la norme matricielle et sa vectorisation est donné par

$$\begin{aligned} \|Y - M\|_{\Sigma^{-1}}^2 &= \text{tr} [(Y - M) \Sigma^{-1} (Y - M)^T] \\ &= \text{tr} [(Y_v - M_v)^T I_m \otimes \Sigma^{-1} (Y_v - M_v)] \\ &= \|(Y_v - M_v)\|_{I_m \otimes \Sigma^{-1}}^2. \end{aligned}$$

2.5 Le modèle additif

Rappelons que l'objectif de cette thèse est d'estimer la matrice d'échelle Σ du modèle additif

$$Y = M + \mathcal{E}, \quad \mathcal{E} \sim ES(0, I_m \otimes \Sigma) \quad (2.5.1)$$

où, la matrice d'observation $Y \in \mathbb{R}^{m \times p}$. On suppose que $p < m$ et que

$$\text{rang}(M) = q < p. \quad (2.5.2)$$

Cette hypothèse sur le rang de M est souvent satisfaite dans la pratique car les colonnes de M peuvent être très corrélées. Par exemple, ces colonnes peuvent être des images individuelles dans une séquence vidéo qui sont généralement très corrélées (cf. Ji et al. [JLSX10b]). Le modèle additif (2.5.1) a été considéré par plusieurs auteurs comme, Canu et Fourdrinier [CF17], Candès et al. [CST13] dans le cadre de l'estimation de la matrice des paramètres M .

Il est plus intéressant de considérer la représentation canonique du modèle additif (2.5.1). En effet, elle nous permet de réduire la dimension des observations dans le sens ou au lieu d'estimer Σ en exploitant toute l'information contenue dans la matrice d'observation Y de taille $m \times p$, nous pouvons l'estimer en se basant sur une matrice de plus petite dimension U de taille $(m - q) \times p$, qui contient la même quantité d'information dont on a besoin pour estimer Σ que celle contenue dans Y . C'est la raison pour laquelle, dans la suite, nous nous concentrons sur la forme canonique du modèle (2.5.1).

2.5.1 Représentation canonique

Grâce à la condition (2.5.2), nous pouvons construire une matrice semi-orthogonale $G_2 \in \mathcal{L}_{m,(m-q)}$, tel que

$$G_2^T M = 0_{(m-q) \times p}. \quad (2.5.3)$$

On complète G_2 par une matrice $G_1 \in \mathcal{L}_{m,q}$ pour former une matrice orthogonale $G \in \mathcal{O}_m$, *i.e.*

$$G = \begin{pmatrix} G_1 \\ G_2 \end{pmatrix}.$$

En pratique, la matrice G peut facilement être obtenue avec la décomposition QR de M en posant $G = Q^T$. Rappelons que, dans cette factorisation, la matrice R est décomposée en deux sous matrices, une matrice triangulaire supérieure R_1 et une matrice nulle R_2 . Ainsi, cette factorisation suppose que

$$M = Q_1 R_1$$

où la matrice semi-orthogonale Q_1 est par analogie la matrice G_1 . Notons que, cette factorisation n'est unique que dans la cas où la matrice M est de plein rang. Ainsi, dans le cadre du modèle (2.5.1), cette factorisation n'est pas unique et n'offre donc qu'un moyen pour construire la matrice G (*cf.* Golub et Van Loan [GVL12]).

En utilisant la matrice orthogonale G , à partir du modèle additif dans (2.5.1), nous avons

$$G^T Y = \begin{pmatrix} G_1^T \\ G_2^T \end{pmatrix} Y = \begin{pmatrix} Z \\ U \end{pmatrix} = \begin{pmatrix} G_1^T \\ G_2^T \end{pmatrix} M + G^T \mathcal{E} = \begin{pmatrix} \theta \\ 0 \end{pmatrix} + G^T \mathcal{E},$$

grâce à (2.5.3) et en utilisant le fait que $G_1^T M = \theta_{q \times p}$. Ainsi, la forme canonique du modèle additif (2.5.1) est donnée par

$$\begin{pmatrix} Z \\ U \end{pmatrix} = \begin{pmatrix} \theta \\ 0 \end{pmatrix} + G^T \mathcal{E}, \quad (2.5.4)$$

où Z et U sont respectivement des matrices de dimension $q \times p$ et $(m - q) \times p$.

Par construction de la forme canonique, Z incorpore l'information contenue dans M , tandis que U n'incorpore que l'information nécessaire pour estimer Σ . De ce fait, la représentation canonique dans (2.5.4) suggère de n'exploiter que l'information contenue dans la matrice résiduelle U qui est de plus petite dimension que Y et est suffisante pour l'estimation de Σ , de telle manière que les estimateurs usuels (*cf.* section 1.2) soient de la forme

$$\hat{\Sigma}_a = a S = a U^T U = a Y^T G_2 G_2^T Y.$$

où a est une constante positive.

Notons que, ce modèle a été considéré par Kubokawa et Srivastava [KS01] et aussi par Canu et Fourdrinier [CF17] afin d'estimer la matrice de paramètres M . En outre, la forme vectorielle de ce modèle (Z et U sont des vecteurs) à été considérée par plusieurs auteurs comme, Cellier et al. [CFR89], Maruyama [Mar03], Maruyama et Strawderman [MS05] et Fourdrinier et Strawderman [FS10].

2.5.2 Densité

Supposons que $\mathcal{E} = Y - M$ a une densité par rapport à la mesure de Lebesgue dans \mathbb{R}^{pm} (cf. ¹ théorème 2.4.1). Dans ce cas, celle-ci est de la forme

$$\begin{aligned} \mathcal{E} &\mapsto |\Sigma|^{-m/2} f(\|\mathcal{E}_v\|_{I_m \otimes \Sigma^{-1}}^2) \\ &= |\Sigma|^{-m/2} f\{\text{tr}(\mathcal{E} \Sigma^{-1} \mathcal{E}^T)\} \end{aligned}$$

et donc la densité de $G^T \mathcal{E}$ est donnée par

$$\begin{aligned} \mathcal{E} &\mapsto |\Sigma|^{-m/2} f\{\text{tr}(G^T \mathcal{E} \Sigma^{-1} \mathcal{E}^T G)\} \\ &= |\Sigma|^{-m/2} f\{\text{tr}(\mathcal{E} \Sigma^{-1} \mathcal{E}^T)\}. \end{aligned} \tag{2.5.5}$$

Puisque G est orthogonale, alors la densité de $G^T \mathcal{E} = ((Z^T - \theta^T), U^T)^T$ est la même que celle de $\mathcal{E} = (Y - M)$. Il s'ensuit que $(Z^T, U^T)^T$ a une distribution à symétrie elliptique vectorisée autour de $(\theta^T, 0)^T$. De plus, d'après le corollaire 2.3.1, Z et U ont aussi des distributions elliptiques vectorisées. Ainsi, de (2.5.4) et (2.5.5), la densité jointe de Z et U est

$$\begin{aligned} (z, u) &\mapsto |\Sigma|^{-m/2} f\left[\text{tr}\left\{\begin{pmatrix} z - \theta \\ u \end{pmatrix} \Sigma^{-1} \begin{pmatrix} z - \theta \\ u \end{pmatrix}^T\right\}\right] \\ &= |\Sigma|^{-m/2} f\left[\text{tr}\{(z - \theta)\Sigma^{-1}(z - \theta)^T\} + \text{tr}\{\Sigma^{-1} u^T u\}\right]. \end{aligned} \tag{2.5.6}$$

Nous faisons remarquer que le paramètre inconnue Σ^{-1} , intervient dans l'expression du risque quadratique (et sa modification dans (4.1.1)) d'un estimateur quelconque $\hat{\Sigma}$ de Σ ; à savoir

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

où $E_{\theta, \Sigma}(\cdot)$ désigne l'espérance par rapport à la densité (2.5.6). Par conséquent, dans notre approche, l'évaluation de la performance d'un quelconque estimateur $\hat{\Sigma}$, ou la comparaison entre différents estimateurs, est impossible. Pour pallier à cette contrainte, un moyen d'y parvenir est d'utiliser l'identité de type Stein-Haff (cf. section 2.7) où l'application de celle-ci nécessite l'introduction d'une nouvelle sous-classe de densité dite de type Berger.

1. Ici la fonction générative est notée par $f(\cdot)$ au lieu de $g(\cdot)$.

2.6 Une sous-classe de lois à symétrie elliptique de type Berger

Dans cette section, nous allons construire une nouvelle classe de densités qui repose sur la densité (2.5.6). À cette fin, dans le contexte du modèle (2.5.1), nous définissons $E_{\theta, \Sigma}^*$ comme l'espérance par rapport à la densité

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

où, pour tout $t \in \mathbb{R}^+$

$$F^*(t) = \frac{1}{2} \int_t^\infty f(v) dv \quad (2.6.1)$$

et la constante de normalisation

$$K^* = \int_{\mathbb{R}^{pm}} |\Sigma|^{-m/2} F^* \left[\text{tr}\{(z - \theta)\Sigma^{-1}(z - \theta)^T\} + \text{tr}\{\Sigma^{-1} u^T u\} \right] dz du \quad (2.6.2)$$

est supposée être finie.

Il convient de noter que, dans le cadre de l'estimation de la moyenne d'une distribution à symétrie sphérique, une classe de densités typique, considérée par Fourdrinier et Strawderman [FS08] et par Fourdrinier, Mezoued et Strawderman [FMS12], est celle introduite par Berger [Ber75] qui suppose que pour tout $t \in \mathbb{R}^+$

$$\frac{F^*(t)}{f(t)} \geq c, \quad (2.6.3)$$

où c est une constante positive. Dans ce travail, nous nous intéressons à la sous-classe de densité qui vérifie

$$c \leq \frac{F^*(t)}{f(t)} \leq b \quad (2.6.4)$$

où c et b sont deux constantes strictement positives.

Plusieurs exemples de densités dans la classe Berger (2.6.3) sont fournis dans Fourdrinier, Mezoued et Strawderman [FMS12]. Notons que, Fourdrinier, Mezoued et Wells [FMW16] ont considéré la sous-classe de densités (2.6.4) afin d'estimer la matrice de précision Σ^{-1} d'une distribution à symétrie elliptique. Dans notre contexte, lorsque l'on cherche à estimer la matrice d'échelle Σ du modèle additif (2.5.1) sous le coût quadratique (3.1.3), la complexité des calculs nécessite la définition d'une nouvelle espérance $E_{\theta, \Sigma}^{**}$ par rapport à la densité

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

tel que, pour tout $t \in \mathbb{R}^+$

$$F^{**}(t) = \frac{1}{2} \int_t^\infty F^*(v) dv \quad (2.6.5)$$

et la constante de normalisation

$$K^{**} = \int_{\mathbb{R}^{pm}} |\Sigma|^{-m/2} F^{**} \left[\text{tr}\{(z - \theta)\Sigma^{-1}(z - \theta)^T\} + \text{tr}\{\Sigma^{-1} u^T u\} \right] dz du \quad (2.6.6)$$

est supposée être finie.

Il convient de noter que sous la modification du coût quadratique dans (4.1.1), la définition de cette espérance n'est pas nécessaire. Par ailleurs, dans le cas particulier gaussien, où la fonction générative $f(\nu) \propto \exp(-\nu/2)$, nous avons $F^{**}(\cdot) = F^*(\cdot) = f(\cdot)$ et donc $c = b = 1$. Par conséquent les espérances $E_{\theta,\Sigma}(\cdot)$, $E_{\theta,\Sigma}^*(\cdot)$ et $E_{\theta,\Sigma}^{**}(\cdot)$ coïncident. Dans le cadre elliptique général, ces espérances sont liées puisque pour toute fonction intégrable $H(Z, U)$, tel que pour tout $z \in \mathbb{R}^{q \times p}$ et $u \in \mathbb{R}^{(m-q) \times p}$

$$\varphi_{\theta,\Sigma}^*(z, u) = \frac{F^*(v)}{f(v)} \quad \text{et} \quad \varphi_{\theta,\Sigma}^{**}(z, u) = \frac{F^{**}(v)}{F^*(v)}, \quad (2.6.7)$$

avec

$$v = \text{tr}\{(z - \theta) \Sigma^{-1} (z - \theta)^T\} + \text{tr}\{\Sigma^{-1} u^T u\},$$

nous avons

$$\begin{aligned} K^* E_{\theta,\Sigma}^*[H(Z, U)] &= \int_{\mathbb{R}^{pm}} \frac{F^*(v)}{f(v)} H(z, u) |\Sigma|^{-m/2} f(v) dz du \\ &= E_{\theta,\Sigma}[\varphi_{\theta,\Sigma}^*(Z, U) H(Z, U)] \end{aligned} \quad (2.6.8)$$

et

$$\begin{aligned} K^{**} E_{\theta,\Sigma}^{**}[H(Z, U)] &= K^* \int_{\mathbb{R}^{pm}} \frac{F^{**}(v)}{F^*(v)} H(z, u) \frac{1}{K^*} |\Sigma|^{-m/2} F^*(v) dz du \\ &= K^* E_{\theta,\Sigma}[\varphi_{\theta,\Sigma}^{**}(Z, U) H(Z, U)] \\ &= E_{\theta,\Sigma}[\varphi_{\theta,\Sigma}^{**}(Z, U) \varphi_{\theta,\Sigma}^*(Z, U) H(Z, U)]. \end{aligned} \quad (2.6.9)$$

Exemples

Nous allons donner quelques exemples de distributions, tirés de Fourdrinier Mezoued et Wells [FMW16], qui vérifient la condition (2.6.4). Il convient de noter que pour ces exemples, $F^*(t)/f(t)$ est monotone en t , croissante ou décroissante. Ce ratio est généralement croissant lorsque la densité (2.5.6) est un mélange de variance de distributions normale ($c = F^*(0)/f(0)$ et $b = \lim_{t \rightarrow \infty} F^*(t)/f(t)$).

Exemple 2.6.1

Soit la densité en (2.5.6) un mélange de variance de distributions normales où la variable de mélange V a une densité $g(\cdot)$ par rapport à la mesure de Lebesgue, c'est-à-dire pour $t \in \mathbb{R}^+$,

$$f(t) \propto \int_0^\infty v^{-p/2} \exp\left(\frac{-t}{2v}\right) g(v) dv.$$

On a

$$\frac{F^*(t)}{f(t)} = \frac{\int_0^\infty v^{-p/2+1} \exp\left(\frac{-t}{2v}\right) g(v) dv}{\int_0^\infty v^{-p/2} \exp\left(\frac{-t}{2v}\right) g(v) dv} = E_t[V] \quad (2.6.10)$$

où $E_t(\cdot)$ désigne l'espérance par rapport à la densité proportionnelle à $v \mapsto v^{-p/2} \exp(-t/2v) g(v)$. Comme cette densité a un rapport de vraisemblance monotone en t , strictement croissant, $E_t[V]$ est croissante en t . Alors, en désignant $E(\cdot)$ l'espérance à par rapport à $g(\cdot)$,

$$c = \frac{F^*(0)}{f(0)} = \frac{\int_0^\infty v^{-p/2+1} g(v) dv}{\int_0^\infty v^{-p/2} g(v) dv} = \frac{E[V^{-p/2+1}]}{E[V^{-p/2}]} > 0,$$

à condition que $E[V^{-p/2}] < \infty$.

Un exemple typique de mélange de variances de distributions normales est la loi de Student multivariée avec k degrés de liberté pour laquelle

$$g(v) \propto v^{-k/2-1} \exp(-k/2v).$$

Par le changement de variable $v = t/u$ dans les intégrales dans (2.6.10), on vérifie facilement que

$$\frac{F^*(t)}{f(t)} = t \frac{\int_0^\infty u^{(p+k)/2-2} \exp\left(-\frac{u}{2}\left(1 + \frac{k}{t}\right)\right) du}{\int_0^\infty u^{(p+k)/2-1} \exp\left(-\frac{u}{2}\left(1 + \frac{k}{t}\right)\right) du},$$

et donc $b = \lim_{t \rightarrow \infty} F^*(t)/f(t) = \infty$. Par conséquent la distribution de Student est dans la classe de Berger (2.6.3) mais ne vérifie pas la condition (2.6.4). Cependant, pour une densité de mélange tronquée $g\mathbb{1}_{[0,A]}$ avec $A > 0$ (normalisée par $\int_0^A g(v) dv$ pour que ce soit une vraie densité). D'après (2.6.10), il est clair que

$$b = \lim_{t \rightarrow \infty} \frac{F^*(t)}{f(t)} = A < \infty.$$

Par conséquent, de telles versions tronquées, en particulier les versions tronquées de la loi de Student sont dans la classe de distributions (2.6.4).

Nous donnons ci-dessous quelques exemples où $F^*(t)/f(t)$ est décroissant, et donc $b = F^*(0)/f(0)$ et $c = \lim_{t \rightarrow \infty} F^*(t)/f(t)$.

Exemple 2.6.2

Soit

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

tels que, $\beta > 0$ et $\gamma > 0$. Pour cette fonction, nous avons

$$c = \frac{1}{2\beta} \quad \text{et} \quad b = \frac{1 + e^{-\gamma}}{2\beta}.$$

Exemple 2.6.3

Soit

$$f(t) \propto \frac{1}{\cosh(\beta t + \gamma)}$$

tels que, $\beta > 0$ et $\gamma > 0$. Pour cette fonction, nous avons

$$c = \frac{1}{2\beta} \quad \text{et} \quad b = \frac{1}{\beta} \arctan(e^{-\gamma}) \cosh(\gamma).$$

Exemple 2.6.4

Soit

$$f(t) \propto (t + A) \exp\left(-\frac{t}{2}\right)$$

tel que, $A > 2$. Pour cette fonction, nous avons

$$c = 1 \quad \text{et} \quad b = 1 + \frac{2}{A}.$$

Maintenant, nous fournissons ici un dernier exemple pour lequel le ratio $F^*(t)/f(t)$ n'est pas monotone.

Exemple 2.6.5

Soit

$$f(t) \propto \frac{1}{t^2 - 2t + 2} \exp\left(-\frac{1}{2} \arctan(t-1)\right) \mathbb{1}_{[0,A]}$$

tel que, $A > 1$. On peut facilement vérifier que

$$\frac{F^*(t)}{f(t)} = t^2 - 2t + 2$$

qui est décroissante sur l'intervalle $[0, 1]$ et croissante sur $[1, A]$. Ensuite, on peut voir que $c \leq F^*(t)/f(t) \leq b$ avec $c = F^*(1)/f(1) = 1$ et $b = \{2 \vee F^*(A)/f(A)\}$.

2.7 Une identité de type Stein-Haff

L'identité de type Stein-Haff que nous allons donner est un outil essentiel pour la construction d'estimateurs améliorés de la matrice d'échelle Σ du modèle additif (2.5.1). Ce type d'identité est d'abord donné par Stein [Ste75] et Haff [Haf79] dans le cadre gaussien. Une extension au cas elliptique est donnée dans Kubokawa et Srivastava [KS99] au travers de l'opérateur différentiel de Haff \mathcal{D}_s , dont le terme générique est donné par

$$d_{ij}^s = \frac{1}{2} (1 + \delta_{ij}) \frac{\partial}{\partial S_{ij}}, \quad (2.7.1)$$

où, $S = U^T U$ où $U \in \mathbb{R}^{n \times p}$ avec $n = m - q$. Il convient de noter que, dans la représentation canonique (2.5.4), on distingue deux cas : S inversible ($p \leq n$) et S non inversible ($p > n$). Ainsi, l'identité de type Stein-Haff que nous allons établir traite les deux cas : $p \leq n$ et $p > n$, contrairement à celle donnée par Kubokawa et Srivastava qui est seulement adaptée au cas où $p \leq n$.

Le lemme suivant, qui servira pour la démonstration du [lemma 2.7.2](#), donne le lien entre les expressions différentielles $\text{tr}(U^T \nabla_U \{G^T S^+\})$ et $\text{tr}(SS^+ \mathcal{D}_s \{SS^+ G\}^T)$, où $(\nabla_U)_{ij} = \partial/\partial U_{ij}$ pour $i = 1, \dots, n$ et $j = 1, \dots, p$.

Lemme 2.7.1

Pour toute fonction matricielle $G(z, s)$ de dimension $p \times p$ faiblement différentiable par rapport à $s \in \mathbb{R}^{p \times p}$, pour tout $z \in \mathbb{R}^{q \times p}$ et pour tout $u \in \mathbb{R}^{n \times p}$, nous avons

$$\text{tr}(U^T \nabla_U \{G^T S^+\}) = \text{tr}[2SS^+ \mathcal{D}_s \{SS^+ G\}^T - ((p \wedge n) + 1) S^+ G]. \quad (2.7.2)$$

Preuve

Puisque $S^+ = SS^+ S^+$ (cf. [définition A.1](#)), nous avons

$$\text{tr}(U^T \nabla_U \{G^T S^+\}) = \sum_{i,j,l}^p \sum_k^n U_{ik}^T \frac{\partial}{\partial U_{kj}} \{(G^T SS^+)_{jl} S_{li}^+\} = R_1 + R_2$$

où

$$R_1 = \sum_{i,j,l}^p \sum_k^n S_{li}^+ U_{ik}^T \frac{\partial(G^T S S^+)_{jl}}{\partial U_{kj}} = \sum_{j,l}^p \sum_k^n (S^+ U^T)_{lk} \frac{\partial(G^T S S^+)_{jl}}{\partial U_{kj}} \quad (2.7.3)$$

et

$$R_2 = \sum_{i,j,l}^p \sum_k^n U_{ik}^T (G^T S S^+)_{jl} \frac{\partial S_{li}^+}{\partial U_{kj}} = \sum_j^p \sum_k^n \left(G^T S S^+ \frac{\partial S^+}{\partial U_{kj}} U^T \right)_{jk}. \quad (2.7.4)$$

Traitant avec R_1 , la règle de dérivation en chaîne nous donne

$$\frac{\partial(G^T S S^+)_{jl}}{\partial U_{kj}} = \sum_{r \geq q}^p \frac{\partial S_{rq}}{\partial U_{kj}} \frac{\partial(G^T S S^+)_{jl}}{\partial S_{rq}}.$$

Notons que, la règle de dérivation habituelle donne

$$\begin{aligned} \frac{\partial S_{qr}}{\partial U_{kj}} &= \frac{\partial}{\partial U_{kj}} \sum_{o=1}^n U_{qo}^T U_{or} \\ &= \sum_{o=1}^p \left(\frac{\partial U_{oq}}{\partial U_{kj}} U_{or} + U_{oq} \frac{\partial U_{or}}{\partial U_{kj}} \right) \\ &= U_{kr} \delta_{qj} + U_{kq} \delta_{rj}. \end{aligned} \quad (2.7.5)$$

Alors

$$\begin{aligned} \frac{\partial(G^T S S^+)_{jl}}{\partial U_{kj}} &= \left(\sum_{r \leq j}^p U_{kr} \frac{\partial}{\partial S_{rj}} + \sum_{q \geq j}^p U_{kq} \frac{\partial}{\partial S_{qj}} \right) (G^T S S^+)_{jl} \\ &= \left(\sum_{r \leq j}^p U_{kr} \frac{\partial}{\partial S_{rj}} + \sum_{r \geq j}^p U_{kr} \frac{\partial}{\partial S_{rj}} \right) (G^T S S^+)_{jl} \\ &= \left(\sum_r^p U_{kr} \frac{\partial}{\partial S_{rj}} + U_{kj} \frac{\partial}{\partial S_{jj}} \right) (G^T S S^+)_{jl} \\ &= 2 \sum_r^p U_{kr} \frac{1}{2} (1 + \delta_{rj}) \frac{\partial}{\partial S_{rj}} (G^T S S^+)_{jl} \\ &= 2 \sum_r^p U_{kr} d_{rj}^s (G^T S S^+)_{jl}, \end{aligned}$$

par définition du terme générique $d_{rj}^s(\cdot)$ de $\mathcal{D}_s(\cdot)$ dans (2.7.1). Par conséquent (2.7.3) devient

$$R_1 = 2 \sum_{l,j}^p \sum_k^n (S^+ U^T)_{lk} \sum_r^p U_{kr} d_{rj}^s (G^T S S^+)_{jl} = 2 \operatorname{tr} (S S^+ \mathcal{D}_s \{S S^+ G\}^T). \quad (2.7.6)$$

Pour R_2 dans (2.7.4), en utilisant la proposition B.1 avec $A = G^T S S^+$ et $B = U^T$, nous avons

$$\begin{aligned} R_2 &= \sum_j^p \sum_k^n ((G^T S S^+)(I - S S^+))_{jj} (U S S^+ U^T)_{kk} \\ &\quad + \sum_j^p \sum_k^n ((I - S S^+) U^T)_{jk} (U (G^T S^+ S^+)^T)_{kj} \\ &\quad - \sum_j^p \sum_k^n ((G^T S^+)_{jj} (U S^+ U^T)_{kk} + (S^+ U^T)_{jk} (U S^+ G)_{kj}). \end{aligned}$$

Puisque $S S^+ (I_p - S S^+) = S (I_p - S S^+) = 0_{p \times p}$, $(I_p - S S^+) U^T = 0_{p \times (p \wedge n)}$, $\text{tr}(S S^+) = p \wedge n$, $S = U^T U$ et $S^+ = S^+ S S^+$ (cf. définition A.1). Alors

$$R_2 = -(p \wedge n + 1) \text{tr}(S^+ G). \quad (2.7.7)$$

Enfin, la combinaison de (2.7.6) et (2.7.7) donne le résultat souhaité. \square

Nous donnons maintenant l'identité de type Stein-Haff dans le cadre des distributions à symétrie elliptique qui traite à la fois le cas inversible et non inversible.

Lemme 2.7.2

Soit $G(z, s)$ une fonction matricielle $p \times p$ tel que, pour tout $z \in \mathbb{R}^{q \times p}$ fixé, la fonction $G(z, s)$ est faiblement différentiable par rapport à $s \in \mathbb{R}^{p \times p}$. Supposons que $E_{\theta, \Sigma} [\text{tr}(\Sigma^{-1} S S^+ G)] < \infty$. Alors

$$E_{\theta, \Sigma} [\text{tr}(\Sigma^{-1} S S^+ G)] = K^* E_{\theta, \Sigma}^* [\text{tr}(2 S S^+ \mathcal{D}_s \{S S^+ G\}^T + (n - (p \wedge n) - 1) S^+ G)]. \quad (2.7.8)$$

Preuve

À l'origine, Fourdrinier Mezoued et Wells [FMW16] ont donné une identité de type Stein-Haff dans le cadre elliptique et pour les cas inversible et non inversible, comme suit

$$E_{\theta, \Sigma} [\text{tr}(\Sigma^{-1} S H)] = K^* E_{\theta, \Sigma}^* [n \text{tr}(H) + \text{tr}(U^T \nabla_U H^T)].$$

Les auteurs remarquent qu'une expression équivalente de cette identité est

$$E_{\theta, \Sigma} [\text{tr}(\Sigma^{-1} S H)] = K^* E_{\theta, \Sigma}^* [n \text{tr}(S S^+ H) + \text{tr}(U^T \nabla_U H^T S S^+)],$$

puisque $S = S S^+ S$. Pour $H = S^+ G$, cette identité devient

$$E_{\theta, \Sigma} [\text{tr}(\Sigma^{-1} S S^+ G)] = K^* E_{\theta, \Sigma}^* [n \text{tr}(S^+ G) + \text{tr}(U^T \nabla_U G^T S^+)]. \quad (2.7.9)$$

Par conséquent, nous déduisons l'identité de type Stein-Haff donnée dans (2.7.8) via l'opérateur Haff $\mathcal{D}_s(\cdot)$ en remplaçant (2.7.2) dans (2.7.9). \square

Il convient de noter que, dans le cas où S est inversible ($p \leq n$), puisque $S^+ = S^{-1}$, l'identité (2.7.8) correspond à

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

qui est l'identité donnée par Kubokawa et Srivastava [KS99]. Dans le cas non inversible, l'identité (2.7.8) devient

$$E_{\theta, \Sigma} [\text{tr}(\Sigma^{-1} S S^+ G)] = K^* E_{\theta, \Sigma}^* [\text{tr}(2 S S^+ \mathcal{D}_s \{S S^+ G\}^T - S^+ G)],$$

qui est, à notre connaissance, une nouvelle identité de type Stein-Haff.

2.8 Synthèse

Dans cette partie, nous avons d'abord défini les distributions à symétrie sphérique au travers de la propriété de l'invariance orthogonale, associée au produit scalaire standard $\langle x, y \rangle = x^T y$ pour $x \in \mathbb{R}^p$ et $y \in \mathbb{R}^p$. Ensuite, une extension de cette notion d'invariance orthogonale à la notion de Σ^{-1} -orthogonale, associée au produit scalaire $\langle x, y \rangle_{\Sigma^{-1}} = x^T \Sigma^{-1} y$, nous a permis la définition des distributions à symétrie elliptique, qui, par la suite, nous ont conduits à définir les distributions elliptiques vectorisées. Nous avons donné ensuite la représentation canonique du modèle additif dans (2.5.1) dont le bruit suit une distribution elliptique vectorisée. Dans les deux prochains chapitres, nous considérons cette représentation dans notre problème d'estimation de la matrice d'échelle Σ . Cette dernière permet de réduire la dimension des observations Y en exploitant la notion de faible rang dans (2.5.2). Par la suite, la complexité du coût quadratique (et sa modification) nous a conduit à réduire notre contexte distributionnel à la sous-classe de type Berger dans (2.6.4) et nous avons aussi donné des exemples de telles distributions. Enfin, le résultat principal de ce chapitre est l'identité de type Stein-Haff développée dans la [section 2.7](#). Celle-ci est l'outil majeur pour la construction d'estimateurs améliorés de Σ . Non seulement cette identité traite à la fois le cas où S est inversible et le cas où elle est non inversible mais aussi, elle est donnée dans le cadre général des distributions à symétrie elliptique.

ESTIMATION DE LA MATRICE D'ÉCHELLE SOUS UN COÛT QUADRATIQUE

Sommaire

3.1 Introduction	28
3.2 Estimateurs de forme générale	29
3.3 Estimateurs orthogonalement invariants	36
3.3.1 Extension de l'exemple de Haff	38
3.3.2 Estimateur de type Konno	43
3.4 Synthèse	43

3.1 Introduction

Dans ce chapitre, nous nous intéressons à l'amélioration des estimateurs usuels de la matrice d'échelle Σ du modèle additif (2.5.1) sous le coût quadratique $\text{tr}(\hat{\Sigma}\Sigma^{-1} - I_p)^2$. À cette fin, nous introduisons dans la section 3.2, une nouvelle classe d'estimateurs de forme générale et, au travers de la différence des risques entre ces estimateurs et les estimateurs usuels, nous donnons des conditions d'amélioration dans le théorème 3.2.1. L'outil essentiel de cette approche est l'identité de type Stein-Haff dans le lemme 2.7.2. Nous verrons dans la section 3.3 que notre théorème 3.2.1 traite de la classe des estimateurs orthogonalement invariants, considérée par plusieurs auteurs dans de tels problèmes d'estimation (cf. section 1.3). Aussi, nous donnons un exemple d'estimateur qui généralise les résultats de Haff [Haf80] au cas elliptique, ou plus précisément à la sous-classe des distributions elliptiques dans (3.2.18).

Reprenons la représentation canonique du modèle additif dans (2.5.1)

$$\begin{pmatrix} Z \\ U \end{pmatrix} = \begin{pmatrix} \theta \\ 0 \end{pmatrix} + G^T \mathcal{E}, \quad \mathcal{E} \sim ES(0, I_m \otimes \Sigma) \tag{3.1.1}$$

où Z et U sont respectivement des matrices $q \times p$ et $n \times p$ avec $n + q = m$. Rappelons que les

estimateurs usuels (cf. section 1.2) de la matrice d'échelle Σ sont de la forme

$$\hat{\Sigma}_a = a U^T U = a S, \quad (3.1.2)$$

où a est une constante positive. Il convient de noter que S est inversible quand $p \leq n$ (contexte en petite dimension) et non inversible lorsque $p > n$ (contexte en grande dimension). Pour ces deux cas, nous cherchons à estimer la matrice d'échelle Σ du modèle (3.1.1) sous le coût quadratique

$$L(\Sigma, \hat{\Sigma}) = \text{tr}(\hat{\Sigma} \Sigma^{-1} - I_p)^2 = \text{tr}(\Sigma^{-1} \hat{\Sigma} \Sigma^{-1} \hat{\Sigma} - 2 \Sigma^{-1} \hat{\Sigma} + I_p) \quad (3.1.3)$$

et son risque associé

$$R(\Sigma, \hat{\Sigma}) = E_{\theta, \Sigma}[L(\Sigma, \hat{\Sigma})], \quad (3.1.4)$$

où $\hat{\Sigma}$ est un estimateur de Σ .

3.2 Estimateurs de forme générale

Les estimateurs de Σ que nous considérons sont de la forme

$$\hat{\Sigma}_{a,G} = a(S + S S^+ G(Z, S)), \quad (3.2.1)$$

où a est une constante strictement positive, S^+ est l'inverse de Moore-Penrose (cf. définition A.1) de S , $G(Z, S)$ est une fonction matricielle $p \times p$ telle que le facteur de correction $S S^+ G(Z, S)$ est supposé symétrique. Notons que, lorsque S est inversible ($p \leq n$) alors S^+ coïncide avec l'inverse S^{-1} et, dans ce cas, $S S^+ = S S^{-1} = I_p$.

Nous développons des conditions sur $G := G(Z, S)$ afin que $\hat{\Sigma}_{a,G}$ améliore $\hat{\Sigma}_a$ sous le risque quadratique (3.1.4), c'est-à-dire, afin que la différence des risques

$$\Delta(G) = R(\Sigma, a(S + S S^+ G(Z, S))) - R(\Sigma, a S) \quad (3.2.2)$$

soit négative ou nulle. À cette fin, nous donnons dans la proposition suivante des conditions pour assurer la finitude des risques de $\hat{\Sigma}_a$ et $\hat{\Sigma}_{a,G}$ et, par conséquent, la différence des risques dans (3.2.2).

Proposition 3.2.1

Supposons que $E_{\theta, \Sigma} [\|\Sigma^{-1} S\|_F^2] < \infty$ et $E_{\theta, \Sigma} [\|\Sigma^{-1} S S^+ G\|_F^2] < \infty$. Alors les risques de $\hat{\Sigma}_a$ et de $\hat{\Sigma}_{a,G}$ sont finis. Dans ce cas, la différence des risques (3.2.2) est également finie et peut être écrite sous la forme

$$\Delta(G) = a^2 E_{\theta, \Sigma} [\text{tr}(\Sigma^{-1} S S^+ (2S + G) \Sigma^{-1} S S^+ G)] - 2a E_{\theta, \Sigma} [\text{tr}(\Sigma^{-1} S S^+ G)]. \quad (3.2.3)$$

Preuve

La preuve consiste à donner des conditions pour assurer la finitude des risques de $\hat{\Sigma}_a$ et de $\hat{\Sigma}_{a,G}$. Nous rappelons d'abord que, pour deux matrices A et B de dimension $p \times p$, l'inégalité de Cauchy-Schwarz exprime que

$$\text{tr}^2(A^T B) \leq \|A\|_F^2 \|B\|_F^2. \quad (3.2.4)$$

D'abord nous traitons la finitude du risque de $\hat{\Sigma}_a = a S$. Supposons que

$$E_{\theta, \Sigma} [\|\Sigma^{-1} S\|_F^2] < \infty. \quad (3.2.5)$$

Nous allons montrer que sous l'hypothèse (3.2.5), le risque de $\hat{\Sigma}_a = a S$ est nécessairement fini. En effet, le coût de $\hat{\Sigma}_a$ s'écrit comme

$$\begin{aligned} L(\Sigma, a S) &= \text{tr} (a S \Sigma^{-1} - I_p)^2 \\ &= a^2 \text{tr}(\Sigma^{-1} S)^2 - 2 a \text{tr}(\Sigma^{-1} S) + p. \end{aligned} \quad (3.2.6)$$

En remarquant que

$$\text{tr}^2(\Sigma^{-1} S) = \text{tr}^2(\Sigma^{-1/2} S \Sigma^{-1/2}),$$

l'application de (3.2.4) pour $A = I_p$ et $B = \Sigma^{-1/2} S \Sigma^{-1/2}$ donne

$$\begin{aligned} \text{tr}^2(\Sigma^{-1} S) &\leq p \|\Sigma^{-1/2} S \Sigma^{-1/2}\|_F^2 \\ &= p \text{tr} (\Sigma^{-1} S)^2. \end{aligned} \quad (3.2.7)$$

Ensuite, une deuxième application de (3.2.4) pour $A^T = B = \Sigma^{-1} S$ donne

$$\begin{aligned} \text{tr}(\Sigma^{-1} S)^2 &\leq \text{tr} ((\Sigma^{-1} S)^T \Sigma^{-1} S) \\ &= \|\Sigma^{-1} S\|_F^2. \end{aligned} \quad (3.2.8)$$

En prenant l'espérance de (3.2.6), il résulte de (3.2.8) que, selon l'hypothèse (3.2.5), l'espérance $E_{\theta, \Sigma} [\text{tr}(\Sigma^{-1} S)^2]$ est finie et donc, en vertu de (3.2.7), que l'espérance $E_{\theta, \Sigma} [\text{tr}^2(\Sigma^{-1} S)]$ est aussi finie, ce qui implique la finitude de $E_{\theta, \Sigma} [\text{tr}(\Sigma^{-1} S)]$. Ainsi le risque de $\hat{\Sigma}_a$ est fini sous l'hypothèse (3.2.8).

Maintenant, nous traitons de la finitude du risque de $\hat{\Sigma}_{a, G} = a S + a S S^+ G$. Supposons que

$$E_{\theta, \Sigma} [\|\Sigma^{-1} S S^+ G\|_F^2] < \infty. \quad (3.2.9)$$

Nous allons voir que, les conditions (3.2.5) et (3.2.9) assurent la finitude du risque de $\hat{\Sigma}_{a, G}$. Notons que, le coût en Σ de $\hat{\Sigma}_{a, G}$ est

$$L(\Sigma, \hat{\Sigma}_{a, G}) = L(\Sigma, a S) + a^2 \text{tr}(\Sigma^{-1} S S^+ G)^2 + 2 a \text{tr} [\Sigma^{-1} S S^+ G (a S \Sigma^{-1} - I_p)].$$

L'application de l'inégalité de Cauchy-Schwarz pour $A^T = B = \Sigma^{-1} S S^+ G$ donne

$$\text{tr}(\Sigma^{-1} S S^+ G)^2 \leq \|\Sigma^{-1} S S^+ G\|_F^2$$

et donc

$$E_{\theta, \Sigma} [\text{tr}(\Sigma^{-1} S S^+ G)^2] \leq E_{\theta, \Sigma} [\|\Sigma^{-1} S S^+ G\|_F^2] < \infty, \quad (3.2.10)$$

par hypothèse (3.2.9). Aussi, par l'inégalité de Cauchy-Schwarz associée au produit scalaire $\langle A, B \rangle = E_{\theta, \Sigma} [\text{tr}(A^T B)]$, on a

$$\begin{aligned} |E_{\theta, \Sigma} \{ \text{tr} [\Sigma^{-1} S S^+ G (a S \Sigma^{-1} - I_p)] \}| &\leq (E_{\theta, \Sigma} [\|\Sigma^{-1} S S^+ G\|_F^2])^{1/2} \times \\ &\quad (E_{\theta, \Sigma} [\|a S \Sigma^{-1} - I_p\|_F^2])^{1/2}. \end{aligned} \quad (3.2.11)$$

Par hypothèse (3.2.10), le premier terme à droite dans l'inégalité (3.2.11) est fini. Aussi, en développant le carré de la norme de Frobenius et en utilisant des arguments similaires à ceux utilisés pour déterminer la finitude du risque de $\hat{\Sigma}_a$, alors il est clair que le second terme est aussi fini.

Enfin, sous les conditions (3.2.5) et (3.2.9), la différence des risques $\Delta(G)$ dans (3.2.2) entre $\hat{\Sigma}_{a,G}$ et $\hat{\Sigma}_a$ est finie et peut être réécrite comme

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

□

La dépendance du paramètre inconnu Σ^{-1} dans (3.2.3) est problématique dans la mesure où ce paramètre intervient dans le terme intégrant de cette différence des risques. Afin d'y remédier, nous appliquons l'identité de Stein-Haff dans le lemme 2.7.2 au deuxième terme de droite dans (3.2.3). Pour le premier terme, où Σ^{-1} apparaît deux fois, nous appliquons le lemme 2.7.2 deux fois, ce qui donne lieu au corollaire suivant.

Corollaire 3.2.1

Soit $G(z, s)$ et $V(z, s)$ des fonctions matricielles $p \times p$ telles que, pour tout $z \in \mathbb{R}^{q \times p}$ fixé, $G(z, s)$ et $V(z, s)$ sont faiblement différentiables par rapport à $s \in \mathbb{R}^{p \times p}$. Supposons que $SS^+V(Z, S)$ est symétrique telles que les espérances $E_{\theta,\Sigma} \left[\left[\text{tr} \left(\Sigma^{-1} SS^+ V \Sigma^{-1} SS^+ G \right) \right] \right]$ et $E_{\theta,\Sigma}^* \left[\left[\text{tr} \left(\Sigma^{-1} SS^+ T^* \right) \right] \right]$ sont finies. Nous avons alors

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

avec

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

Il convient de noter que l'introduction de l'espérance $E_{\theta,\Sigma}^{**}$ est nécessaire à la double application de l'identité de type Stein-Haff, vu que le paramètre Σ^{-1} est présent deux fois dans le terme intégrant de la différence des risques dans (3.2.3). Nous verrons dans le chapitre 4 que l'on peut se dispenser de $E_{\theta,\Sigma}^{**}$ en modifiant le coût quadratique (3.1.3).

Preuve

Premièrement, nous appliquons le lemme 2.7.2 en prenant $V \Sigma^{-1} SS^+ G$ au lieu de G , ce qui donne

$$E_{\theta,\Sigma} \left[\text{tr} \left(\Sigma^{-1} SS^+ V \Sigma^{-1} SS^+ G \right) \right] = K^* E_{\theta,\Sigma}^* \left[\text{tr} \left[2SS^+ \mathcal{D}_s \{ G^T SS^+ \Sigma^{-1} V^T SS^+ \} + (n - (p \wedge n) - 1) \Sigma^{-1} SS^+ G S^+ V \right] \right]. \quad (3.2.12)$$

Deuxièmement, en vertu de (A.1) dans la proposition A.1, le premier terme intégrant dans le membre de droite de (3.2.12) peut être réécrit comme

$$\begin{aligned}\text{tr} \left[SS^+ \mathcal{D}_s \{ G^T SS^+ \Sigma^{-1} V^T SS^+ \} \right] &= \text{tr} \left[\Sigma^{-1} V^T SS^+ \mathcal{D}_s \{ G^T SS^+ \} \right] \\ &\quad + \text{tr} \left[SS^+ \left(\Sigma^{-1} SS^+ G \mathcal{D}_s \right)^T \{ V^T SS^+ \} \right].\end{aligned} \quad (3.2.13)$$

Maintenant, l'application de (A.2) dans la proposition A.1 au deuxième terme du membre de droite de (3.2.13) donne

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

De plus, en utilisant le fait que $SS^+ = SS^+ SS^+$, l'identité (A.1) dans la proposition A.1 nous donne

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

Par conséquent, en vertu de (B.13) et en utilisant le fait que $(I_p - SS^+) S = 0_{p \times p}$ (cf. définition A.1), on obtient

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

Donc

$$\operatorname{tr} \left[SS^+ (\Sigma^{-1} SS^+ G \mathcal{D}_s)^T \{V^T S^+ S\} \right] = \operatorname{tr} \left[\Sigma^{-1} SS^+ G (\mathcal{D}_s \{SS^+ V\} - \frac{1}{2}(p - p \wedge n) S^+ V) \right]. \quad (3.2.14)$$

Maintenant, en combinant (3.2.12), (3.2.13) et (3.2.14) et en utilisant la symétrie de $SS^+ V$ on a

$$E_{\theta, \Sigma} [\operatorname{tr} (\Sigma^{-1} SS^+ V \Sigma^{-1} SS^+ G)] = K^* E_{\theta, \Sigma}^* [\operatorname{tr} (\Sigma^{-1} SS^+ T^*)], \quad (3.2.15)$$

où

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

Enfin, sous la condition $E_{\theta, \Sigma}^* [\operatorname{tr} (\Sigma^{-1} SS^+ T^*)] < \infty$, l'application du lemme 2.7.2 à (3.2.15) avec T^* au lieu de G nous donne

$$E_{\theta, \Sigma}^* [\operatorname{tr} (\Sigma^{-1} SS^+ T^*)] = K^{**} E_{\theta, \Sigma}^{**} [\operatorname{tr} (2S^+ S \mathcal{D}_s \{SS^+ T^*\}^T - (n - (p \wedge n) - 1) S^+ T^*)],$$

qui est le résultat annoncé. □

Nous pouvons maintenant, grâce au lemme 2.7.2 et au corollaire 3.2.1, donner une expression de la différence des risques (3.2.3) dont le terme intégrant ne dépend pas du paramètre inconnu Σ^{-1} . À cette fin, il est intéressant de noter que, les conditions que nous allons utiliser dans la prochaine proposition impliquent celles qui sont exprimées dans lemme 2.7.2 et le corollaire 3.2.1.

Proposition 3.2.2

Soit $G(z, s)$ une fonction matricielle $p \times p$ tel que, pour tout $z \in \mathbb{R}^{q \times p}$ fixé, $G(z, s)$ est faiblement différentiable par rapport à $s \in \mathbb{R}^{p \times p}$. Supposons que, $SS^+ G(Z, S)$ est une matrice symétrique et que $E_{\theta, \Sigma} [\|\Sigma^{-1} S\|_F^2]$, $E_{\theta, \Sigma} [\|\Sigma^{-1} SS^+ G\|_F^2]$ et $E_{\theta, \Sigma}^* [\operatorname{tr} (\Sigma^{-1} SS^+ T^*)]$ soient des espérances finis. Alors la différence des risques dans (3.2.3) s'écrit

$$\begin{aligned} \Delta(G) = & a^2 K^* K^{**} E_{\theta, \Sigma}^{**} [\operatorname{tr} (2 SS^+ \mathcal{D}_s \{SS^+ T^*\}^T + (n - (p \wedge n) - 1) S^+ T^*)] \\ & - 2a K^* E_{\theta, \Sigma}^* [\operatorname{tr} (2 S^+ S \mathcal{D}_s \{SS^+ G\} + (n - (p \wedge n) - 1) S^+ G)] \end{aligned} \quad (3.2.16)$$

où

$$T^* = 4(S + SS^+ G) \mathcal{D}_s \{SS^+ G\} + G(2n SS^+ - (p - n + 1) S^+ G). \quad (3.2.17)$$

Preuve

Rappelons que les conditions $E_{\theta,\Sigma} [\|\Sigma^{-1} S\|_F^2] < \infty$ et $E_{\theta,\Sigma} [\|\Sigma^{-1} S S^+ G\|_F^2] < \infty$ garantissent la finitude de la différence des risques entre $\hat{\Sigma}_a$ et $\hat{\Sigma}_{a,G}$. Notons que la seconde condition implique celle utilisée dans le [lemme 2.7.2](#), à savoir

$$E_{\theta,\Sigma} [\|\text{tr}(\Sigma^{-1} S S^+ G)\|] < \infty,$$

puisque l'application de l'inégalité de Cauchy-Schwarz dans (3.2.4) pour $A^T = I_p$ et $B = \Sigma^{-1} S S^+ G$ donne

$$\begin{aligned} E_{\theta,\Sigma} [\|\text{tr}(\Sigma^{-1} S S^+ G)\|] &= E_{\theta,\Sigma} [\|\text{tr}(I_p \Sigma^{-1} S S^+ G)\|] \\ &\leq p^{1/2} E_{\theta,\Sigma} [\|\text{tr}(\Sigma^{-1} S S^+ G)\|_F^2]. \end{aligned}$$

Notons aussi que cette condition et la condition $E_{\theta,\Sigma}^* [\|\text{tr}(\Sigma^{-1} S S^+ T^*)\|] < \infty$ impliquent les conditions dans le [corollaire 3.2.1](#) avec T^* dans (3.2.17). En particulier, pour $V = 2S + G$ nous avons

$$E_{\theta,\Sigma} [\|\text{tr}(\Sigma^{-1} S S^+ V \Sigma^{-1} S S^+ G)\|] < \infty.$$

Ensuite, en appliquant le [lemme 2.7.2](#) à la deuxième espérance dans (3.2.3) et le [corollaire 3.2.1](#) au premier terme avec $V = 2S + G$ donne (3.2.16), puisque

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

où nous avons utilisé (B.12), le fait que $SS^+(I_p - SS^+) = 0_{p \times p}$ (cf. [Definition A.1](#)) et la symétrie de $SS^+ G$.

□

Comme la différence des risques implique des expressions au travers des espérances $E_{\theta,\Sigma}^*$ et $E_{\theta,\Sigma}^{**}$, pour comparer ces quantités et avoir une expression homogène, qui ne dépend que de $E_{\theta,\Sigma}$, nous réduisons le contexte de distribution à la sous-classe de densités de type Berger dans (2.6.4) tel que, pour tout $t \in \mathbb{R}^+$,

$$c \leq \frac{F^*(t)}{f(t)} \leq b. \quad (3.2.18)$$

Dans la classe des estimateurs usuels de la forme $\hat{\Sigma}_a = aS$, il existe une constante $a_o > 0$ tel que $\hat{\Sigma}_{a_o} = a_o S$ est optimal (en d'autres termes, le risque de $\hat{\Sigma}_{a_o}$ est inférieur ou égal au risque de $\hat{\Sigma}_a$ pour tout $a > 0$). L'amélioration de $\hat{\Sigma}_a = aS$ dans (3.1.2) sera illustrée par l'amélioration de $\hat{\Sigma}_{a_o,G} = a_o S + a_o SS^+ G(Z, S)$, puisqu'il est clair que, si on améliore l'optimal, alors on améliore tout les autres. La proposition suivante donne la valeur optimale de a_o .

Proposition 3.2.3

Pour la classe d'estimateurs usuels $\hat{\Sigma}_a$ dans (3.1.2), où $a > 0$, la constante optimale a_o qui minimise le risque en (3.1.4) est

$$a_o = \frac{1}{K^{**} (p + n + 1)}, \quad (3.2.19)$$

où la constante de normalisation K^{**} est donnée dans (2.6.6).

Preuve

Soit $\hat{\Sigma}_a = aS$, où a est une constante positive. Alors, en vertu de (3.1.4), le risque de $\hat{\Sigma}_a$ en Σ est

$$\begin{aligned} R(\Sigma, \hat{\Sigma}_a) &= E_{\theta, \Sigma} \left[\text{tr} (aS\Sigma^{-1} - I_p)^2 \right] \\ &= a^2 E_{\theta, \Sigma} \left[\text{tr} (\Sigma^{-1}SS^+S)^2 \right] - 2a E_{\theta, \Sigma} \left[\text{tr} (\Sigma^{-1}SS^+S) \right] + p, \end{aligned} \quad (3.2.20)$$

puisque $S = SS^+S$. Pour traiter le premier terme de droite dans (3.2.20), on applique le corollaire 3.2.1 pour $G = V = S$. Pour traiter le second terme, on applique le lemme 2.7.2 et on obtient

$$\begin{aligned} R(\Sigma, \hat{\Sigma}_a) &= a^2 K^* K^{**} E_{\theta, \Sigma}^* \left[\text{tr} \left[2S^+S \mathcal{D}_s \{SS^+T^*\}^T + (n - (p \wedge n) - 1)S^+T^* \right] \right] \\ &\quad - 2a K^* E_{\theta, \Sigma}^* \left[\text{tr} \left[2S^+S \mathcal{D}_s \{S\} + (n - (p \wedge n) - 1)S^+S \right] \right] + p, \end{aligned} \quad (3.2.21)$$

où

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

Grâce à (B.12) et à (B.13) et en utilisant le fait que $S(I_p - SS^+) = 0_{p \times p}$, les termes intégrands dans la partie droite de (3.2.21) sont évalués par

$$\text{tr} \left[2S^+S \mathcal{D}_s \{SS^+T^*\}^T + (n - (p \wedge n) - 1)S^+T^* \right] = (p + n + 1)(p + n - p \wedge n) \text{tr}(SS^+) \quad (3.2.22)$$

et

$$\text{tr} \left(2S^+S \mathcal{D}_s \{S\} + (n - (p \wedge n) - 1)SS^+ \right) = (p + n - p \wedge n) \text{tr}(SS^+). \quad (3.2.23)$$

En remplaçant (3.2.22) et (3.2.23) dans (3.2.21) on obtient

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

puisque $\text{tr}(SS^+) = p \wedge n$. Clairement, le choix de $a = 1/K^{**}(p + n + 1)$ est optimal sous le risque (3.1.4). □

Il est intéressant de noter que, la constante optimal a_o dans (3.2.19) est la même dans les deux cas d'inversibilité ou de non inversibilité de S ce qui peut ne pas être le cas pour un coût différent. Aussi, celle-ci dépend de la distribution du bruit \mathcal{E} dans (2.5.1) au travers de K^{**} . Dans le cas gaussien par exemple, où $K^{**} = 1$, notre résultat dans (3.2.19) coïncide avec celui de Haff [Haf80] dans le contexte en petite dimension (S inversible ou $p \leq n$) et avec celui de Konno [Kon09] dans le contexte en grande dimension (S non inversible ou $p > n$).

Dans le théorème suivant, qui est le résultat principal de cette section, nous donnons les conditions d'amélioration de $a_o(S + SS^+G(Z, S))$ sur $a_o S$.

Théorème 3.2.1

Considérons une densité de la forme (2.5.6) qui satisfait la condition (3.2.18). Si

$$\text{tr} [2 SS^+ \mathcal{D}_s\{SS^+ G\} + (n - (p \wedge n) - 1) S^+ G] \geq 0, \quad (3.2.24)$$

alors l'estimateur $\hat{\Sigma}_{a_o, G} = a_o(S + SS^+G(Z, S))$ améliore $\hat{\Sigma}_{a_o} = a_o S$ dès que

$$\text{tr} [2 S^+ S \mathcal{D}_s\{SS^+ T^*\}^T + (n - (p \wedge n) - 1) S^+ T^* - 2(p + n + 1) \frac{c^2}{b^2} (2 SS^+ \mathcal{D}_s\{SS^+ G\} + (n - (p \wedge n) - 1) S^+ G)] \leq 0, \quad (3.2.25)$$

où

$$T^* = 4(S + SS^+ G) \mathcal{D}_s\{SS^+ G\} + G(2n SS^+ - (p - n + 1) S^+ G). \quad (3.2.26)$$

Preuve

En utilisant (2.6.8) et (2.6.9), on exprime la différence des risques (3.2.16) uniquement au travers de l'espérance $E_{\theta, \Sigma}$ comme suit

$$\begin{aligned} \Delta(G) = & a_o^2 K^* E_{\theta, \Sigma} [\varphi_{\theta, \Sigma}^*(Z, U) \varphi_{\theta, \Sigma}^{**}(Z, U) (\text{tr} [2 S^+ S \mathcal{D}_s\{SS^+ T^*\}^T - \\ & (n - (p \wedge n) - 1) S^+ T^*] \\ & - \frac{2}{a_o K^* \varphi_{\theta, \Sigma}^{**}(Z, U)} \text{tr} [2 S^+ S \mathcal{D}_s\{SS^+ G\} - (n - (p \wedge n) - 1) S^+ G])] , \end{aligned}$$

où a_o est donné dans (3.2.19). En utilisant l'inégalité (B.22) dans le lemme B.3 et en notant que $K^{**} \geq c^2$, on a

$$\varphi_{\theta, \Sigma}^*(Z, U) \frac{K^{**}}{K^*} \geq \frac{c^2}{b^2}.$$

Alors, sous la condition (3.2.24),

$$\begin{aligned} \Delta(G) \leq & a^2 K^* E_{\theta, \Sigma} [\varphi_{\theta, \Sigma}^*(Z, U) \varphi_{\theta, \Sigma}^{**}(Z, U) (\text{tr} [2 S^+ S \mathcal{D}_s\{SS^+ T^*\}^T - \\ & (n - (p \wedge n) - 1) S^+ T^*] \\ & - 2(p + n + 1) \frac{c^2}{b^2} \text{tr} [2 S^+ S \mathcal{D}_s\{SS^+ G\} - (n - (p \wedge n) - 1) S^+ G])] . \end{aligned}$$

Par conséquent, l'estimateur $\hat{\Sigma}_{a_o, G}$ domine $\hat{\Sigma}_{a_o}$ dès que l'inégalité (3.2.25) est satisfaite puisque $\varphi_{\theta, \Sigma}^*(Z, U)$ et $\varphi_{\theta, \Sigma}^{**}(Z, U)$ sont positives. □

Comme il est bien connu, les estimateurs usuels de Σ sont inadmissibles et peuvent être améliorés. Ceci ce traduit par le fait que les plus grandes (les plus petites) valeurs propres de $\hat{\Sigma}_a$ ont tendance à être plus grandes (plus petites) que les valeurs propres correspondantes de Σ (cf. section 1.2). Ce fait suggère de ramener (*to shrink*) les valeurs propres de $\hat{\Sigma}_a$ vers une valeur centrale, ce qui donne lieu à des estimateurs orthogonalement invariants de la forme (1.3.2). Le théorème 3.2.1 est bien adapté pour traiter de cette classe d'estimateurs.

3.3 Estimateurs orthogonalement invariants

Dans cette section, nous proposons des estimateurs compétitifs de Σ basés sur la décomposition en valeurs propres de S (cf. lemme B.1), soit

$$S = H_1 L H_1^T,$$

où $H_1 \in \mathcal{L}_{p,p \wedge n}$ et $L \in \mathcal{D}_{(p \wedge n)}$. Les estimateurs alternatifs $\hat{\Sigma}_{a,G} = a (S + S S^+ G(Z, S))$ dans (3.2.1) que nous considérons sont orthogonalement invariants, c'est-à-dire tel que

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

pour une fonction $\Psi(L) \in \mathcal{D}_{p \wedge n}$ différentiable par rapport à L . Sous cette décomposition en valeurs singulières, comme $S S^+ G = G$, ces estimateurs peuvent être exprimés sous la forme

$$\hat{\Sigma}_{a,\Psi} = a H_1 (L + \Psi(L)) H_1^T. \quad (3.3.1)$$

Il convient de noter que, dans (3.3.1), la fonction $\Psi(L)$ modifie les valeurs propres de S , soit celles dans L . Ainsi ce type d'estimateur améliore l'estimation des valeurs propres sans affecter les vecteurs propres. Dans le contexte gaussien, Stein [Ste86], Dey et Srinivasan [DS85], Haff [Haf86], Kubokawa et al. [KRS92], Ledoit et Wolf [LW04], Fisher et Sun [FS11] utilisent ce type de modifications. Nous donnons dans le corollaire suivant des conditions générales sur le facteur correcteur $\Psi := \Psi(L)$, dans le cadre des distributions à symétrie elliptique, pour que de tels estimateurs améliorent ceux de la forme $\hat{\Sigma}_a$.

Corollaire 3.3.1

Considérons une densité de la forme (2.5.6) qui satisfait (3.2.18). Sous la condition

$$\sum_{i=1}^{p \wedge n} \left\{ (p + n - 2(p \wedge n) - 1) \frac{\psi_i}{l_i} + 2 \frac{\partial \psi_i}{\partial l_i} + \sum_{j \neq i}^{p \wedge n} \frac{\psi_i - \psi_j}{l_i - l_j} \right\} \geq 0, \quad (3.3.2)$$

l'estimateur $\hat{\Sigma}_{a_o,\Psi} = a_o H_1 \text{diag}(l_1 + \psi_1, l_2 + \psi_2, \dots, l_{p \wedge n} + \psi_{p \wedge n}) H_1^T$ améliore l'estimateur usuel $\hat{\Sigma}_{a_o} = a_o H_1 \text{diag}(l_1, l_2, \dots, l_{p \wedge n}) H_1^T$ dès que

$$\sum_{i=1}^{p \wedge n} \left\{ (p + n - 2(p \wedge n) - 1) \frac{\phi_i}{l_i} + 2 \frac{\partial \phi_i}{\partial l_i} + \sum_{j \neq i}^{p \wedge n} \frac{\phi_i - \phi_j}{l_i - l_j} - 2(p + n + 1) \frac{c^2}{b^2} \left((p + n - 2(p \wedge n) - 1) \frac{\psi_i}{l_i} + 2 \frac{\partial \psi_i}{\partial l_i} + \sum_{j \neq i}^{p \wedge n} \frac{\psi_i - \psi_j}{l_i - l_j} \right) \right\} \leq 0, \quad (3.3.3)$$

où

$$\begin{aligned} \phi_i &= 2(p + n - p \wedge n) \psi_i + (p + n - 2(p \wedge n) - 1) \frac{\psi_i^2}{l_i} \\ &\quad + 4(\psi_i + l_i) \frac{\partial \psi_i}{\partial l_i} + 2(\psi_i + l_i) \sum_{j \neq i}^{p \wedge n} \frac{\psi_i - \psi_j}{l_i - l_j}. \end{aligned} \quad (3.3.4)$$

Preuve

En utilisant la décomposition en valeurs propres de S (cf. lemme B.1), il est clair que $S^+ = H_1 L^{-1} H_1^T$, $SS^+ = H_1 H_1^T$, $SS^+G = G = H_1 \Psi H_1^T$ et $S^+G = H_1 L^{-1} \Psi H_1^T$. Dans ce contexte, la condition (3.2.24) dans le théorème 3.2.1 peut être réécrite sous la forme

$$\operatorname{tr} [2 S^+ S \mathcal{D}_s \{SS^+G\} + (n - (p \wedge n) - 1) S^+ G] = \operatorname{tr} [2 H_1 H_1^T \mathcal{D}_s \{H_1 \Psi H_1^T\} + (n - (p \wedge n) - 1) H_1 L^{-1} \Psi H_1^T] .$$

L'application de l'identité (B.11) dans le lemme B.2 et l'utilisation des faits que $H_1^T H_1 = I_{p \wedge n}$ et $H_1^T (I_p - H_1 H_1^T) = 0_{(p \wedge n) \times p}$, donnent

$$\operatorname{tr} [2 S^+ S \mathcal{D}_s \{SS^+G\} + (n - (p \wedge n) - 1) S^+ G] = \operatorname{tr} [2 \Psi^{(1)} + (n - (p \wedge n) - 1) L^{-1} \Psi] \quad (3.3.5)$$

où $\Psi^{(1)} \in \mathcal{D}_{p \wedge n}$ avec

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

ce qui correspond à (3.3.2).

Quant à la condition (3.2.25) dans le théorème 3.2.1, notons que T^* dans (3.2.26) devient

$$T^* = 4 H_1 (L + \Psi) H_1^T \mathcal{D}_s \{H_1 \Psi H_1^T\} + H_1 (2n \Psi - (p - n + 1) \Psi^2 L^{-1}) H_1^T .$$

L'application de l'identité (B.11) et utilisation des faits que $H_1 H_1^T = I_{p \wedge n}$ et $H_1^T (I_p - H_1 H_1^T) = 0_{(p \wedge n) \times (p \wedge n)}$ donnent

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

où le terme générique de $\Phi \in \mathcal{D}_{p \wedge n}$ est donné par

$$\begin{aligned} \phi_i &= 2(p + n - p \wedge n) \psi_i + (p + n - 2(p \wedge n) - 1) \frac{\psi_i^2}{l_i} \\ &\quad + 4(\psi_i + l_i) \frac{\partial \psi_i}{\partial l_i} + 2(\psi_i + l_i) \sum_{j \neq i}^{p \wedge n} \frac{\psi_i - \psi_j}{l_i - l_j} . \end{aligned}$$

De manière similaire à la condition (3.2.24), on applique (B.11) avec $G = T^*$ pour avoir

$$\operatorname{tr} [2 S^+ S \mathcal{D}_s \{SS^+T^*\} + (n - (p \wedge n) - 1) S^+ T^*] = \operatorname{tr} [2 \Phi^{(1)} + (n - (p \wedge n) - 1) L^{-1} \Phi] , \quad (3.3.6)$$

où $\Phi^{(1)} \in \mathcal{D}_{p \wedge n}$ avec

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

Par conséquent, le résultat souhaité découle de (3.3.5) et (3.3.6). □

3.3.1 Extension de l'exemple de Haff

L'exemple suivant est une extension de celui donné par Haff [Haf80] dans (1.3.4), qui a considéré le contexte gaussien et seulement le cas où S est inversible. Avec cette extension, nous traitons le cas elliptique et, aussi, les deux cas où S est inversible et non inversible dans un cadre unifiée. Notons que, dans ce contexte, cet estimateur qui est de la forme $\hat{\Sigma} = a_o (S + \nu t(\nu) C)$, nous prenons $C = SS^+ = H_1 H_1^T$.

Proposition 3.3.1

Considérons une densité de la forme (2.5.6) qui satisfait (3.2.18) et tel que

$$\frac{p+n-(p \wedge n)+2}{p+n+1} \leq \frac{c^2}{b^2} \leq \frac{2p+2n-5(p \wedge n)-3}{p+n+1}. \quad (3.3.7)$$

Soit $\hat{\Sigma}_{a_o, \Psi}$ un estimateur de la forme

$$\hat{\Sigma}_{a_o, \Psi} = a_o H_1 \text{diag}(l_1 + \nu t(\nu), l_2 + \nu t(\nu), \dots, l_{p \wedge n} + \nu t(\nu)) H_1^T,$$

où $\nu = 1/\text{tr}(S^+)$ et $t(\cdot)$ est une fonction convexe, strictement positive, décroissante et deux fois différentiable. Alors $\hat{\Sigma}_{a_o, \Psi}$ améliore $\hat{\Sigma}_{a_o}$ dès que

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

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

et

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

Preuve

La preuve consiste à prouver les conditions (3.3.2) et (3.3.3) pour $\Psi = \nu t(\nu) I_{p \wedge n}$, où $\nu = 1/\text{tr}(S^+)$ et $t(\cdot)$ est une fonction décroissante, convexe et deux fois différentiable. Notons que, la condition (3.3.7) est valide si et seulement si $p+n-(p \wedge n+2) \leq 2p+2n-5(p \wedge n)-3$, qui est équivalent à

$$p+n \geq 4(p \wedge n) + 5.$$

Dans ce contexte, la condition (3.3.2) est équivalente à

$$\sum_i^{p \wedge n} \left\{ (p+n-2(p \wedge n)-1) \frac{\nu t(\nu)}{l_i} + 2 \frac{\partial \{ \nu t(\nu) \}}{\partial l_i} \right\} \geq 0. \quad (3.3.8)$$

En utilisant le fait que, pour $r, k = 0, 1, 2, \dots$,

$$\frac{\partial}{\partial l_i} \left\{ \frac{\nu^k}{l_i^r} \right\} = k \frac{\nu^{k+1}}{l_i^{r+2}} - r \frac{\nu^k}{l_i^{r+1}} \quad \text{et} \quad \frac{\partial}{\partial l_i} \{ t(\nu) \} = \frac{\nu^2}{l_i^2} t'(\nu), \quad (3.3.9)$$

la condition (3.3.8) devient

$$(p+n-2(p \wedge n)-1) t(\nu) + 2\rho_2 t(\nu) + 2\rho_2 \nu t'(\nu) \geq 0, \quad (3.3.10)$$

où, pour tout $k = 1, \dots, \rho_k = \nu^k \text{tr}(S^{+k})$. Grâce à la deuxième inégalité dans (A.3), cette condition est satisfaite dès que la condition (i) est vérifiée.

Quant à la condition (3.3.3), nous utilisons également (3.3.9) pour exprimer (3.3.4) comme

$$\begin{aligned} \phi_i = & \left\{ 4 \frac{\nu^3}{l_i^2} + (p+n-2(p \wedge n) - 1) \frac{\nu^2}{l_i} \right\} t^2(\nu) + 2 \left\{ 2 \frac{\nu^2}{l_i} + \nu(p+n-(p \wedge n)) \right\} t(\nu) \\ & + 4 \left\{ \frac{\nu^3}{l_i^2} t(\nu) + \frac{\nu^2}{l_i} \right\} \nu t'(\nu). \end{aligned}$$

Il s'ensuit que

$$\begin{aligned} \sum_i^{p \wedge n} \frac{\phi_i}{l_i} = & \{4\rho_3 + (p+n-2(p \wedge n) - 1)\rho_2\} t^2(\nu) + 2 \{2\rho_2 + (p+n-p \wedge n)\} t(\nu) \\ & + 4 \{\rho_3 t(\nu) + \rho_2\} \nu t'(\nu) \end{aligned} \quad (3.3.11)$$

et

$$\begin{aligned} \sum_i^{p \wedge n} \sum_{j \neq i}^{p \wedge n} \frac{\phi_i - \phi_j}{l_i - l_j} = & \left\{ 4\nu^3 \sum_i^{p \wedge n} \sum_{j \neq i}^{p \wedge n} \frac{l_i^{-2} - l_j^{-2}}{l_i - l_j} + (p+n-2p \wedge n - 1) \nu^2 \sum_i^{p \wedge n} \sum_{j \neq i}^{p \wedge n} \frac{l_i^{-1} - l_j^{-1}}{l_i - l_j} \right\} t^2(\nu) \\ & + 4\nu^2 \sum_i^{p \wedge n} \sum_{j \neq i}^{p \wedge n} \frac{l_i^{-1} - l_j^{-1}}{l_i - l_j} t(\nu) \\ & + 4 \left\{ \nu^3 t(\nu) \sum_i^{p \wedge n} \sum_{j \neq i}^{p \wedge n} \frac{l_i^{-2} - l_j^{-2}}{l_i - l_j} + \nu^2 \sum_i^{p \wedge n} \sum_{j \neq i}^{p \wedge n} \frac{l_i^{-1} - l_j^{-1}}{l_i - l_j} \right\} \nu t'(\nu). \end{aligned} \quad (3.3.12)$$

De plus, grâce à (3.3.9), il vient

$$\begin{aligned} \sum_i^{p \wedge n} \frac{\partial \phi_i}{\partial l_i} = & \{12\rho_4 + 2(p+n-2(p \wedge n) - 5)\rho_3 - (p+n-2(p \wedge n) - 1)\rho_2\} t^2(\nu) \\ & + 2 \{4\rho_3 + (p+n-(p \wedge n) - 2)\rho_2\} t(\nu) \\ & + 2 \{2\rho_4 \nu t'(\nu) + 12\rho_4 t(\nu) + (p+n-2(p \wedge n) - 5)\rho_3 t(\nu) \\ & \quad + 16\rho_3 + (p+n-(p \wedge n) - 2)\rho_2\} \nu t'(\nu) \\ & + 4 \{\rho_4 t(\nu) + \rho_3\} \nu^2 t''(\nu). \end{aligned} \quad (3.3.13)$$

En remplaçant (3.3.11), (3.3.12), (3.3.13) et le terme du membre de gauche de (3.3.10) dans (3.3.3), cette dernière condition peut être exprimée par

$$H_1(z, u) + H_2(z, u) \leq 0 \quad (3.3.14)$$

où

$$\begin{aligned}
H_1(z, u) &= 2 \{12 \rho_4 + 2(p+n-2(p \wedge n) - 5) \rho_3 - (p+n-2(p \wedge n) - 1) \rho_2\} t^2(\nu) \\
&\quad + (p+m-2(p \wedge n) - 1) \{(p+n-2(p \wedge n) - 1) \rho_2 + 4 \rho_3\} t^2(\nu) \\
&\quad + 2 \left\{ (p+n-2(p \wedge n) - 1) \left(p+n-p \wedge n + 2 \rho_2 - \frac{c^2}{b^2} (p+n+1) \right) \right\} t(\nu) \\
&\quad + 4 \left\{ 4 \rho_3 + (p+n-(p \wedge n) - 2) \rho_2 - \frac{c^2}{b^2} (p+n+1) \rho_2 \right\} t(\nu) \\
&\quad + \nu^2 \left\{ 4\nu \sum_i^{p \wedge n} \sum_{j \neq i}^{p \wedge n} \frac{l_i^{-2} - l_j^{-2}}{l_i - l_j} + (p+n+2(p \wedge n) - 1) \sum_i^{p \wedge n} \sum_{j \neq i}^{p \wedge n} \frac{l_i^{-1} - l_j^{-1}}{l_i - l_j} \right\} t^2(\nu)
\end{aligned} \tag{3.3.15}$$

et

$$\begin{aligned}
H_2(z, u) &= \{48 \rho_4 t(\nu) + 8 \rho_4 \nu t'(\nu) + 8(p+n-2(p \wedge n) - 3) \rho_3 t(\nu) + 32 \rho_3 \\
&\quad + 4(2p+2n-3(p \wedge n) - 3) \rho_2 - 4 \frac{c^2}{b^2} (p+n+1) \rho_2 \\
&\quad + 4\nu^3 t(\nu) \sum_i^{p \wedge n} \sum_{j \neq i}^{p \wedge n} \frac{l_i^{-2} - l_j^{-2}}{l_i - l_j} + 4\nu^2 \sum_i^{p \wedge n} \sum_{j \neq i}^{p \wedge n} \frac{l_i^{-1} - l_j^{-1}}{l_i - l_j} \} \nu t'(\nu) \\
&\quad + 8 \{ \rho_4 t(\nu) + \rho_3 \} \nu^2 t''(\nu).
\end{aligned} \tag{3.3.16}$$

Nous allons maintenant construire des majorants successifs pour $H_1(z, u)$ dans (3.3.15) et $H_2(z, u)$ dans (3.3.16). Premièrement, il s'ensuit des deuxièmes inégalités dans (A.5) et (A.6) que le dernier terme de droite dans (3.3.15) est négatif. Puis, l'utilisation du fait que $\rho_4 \leq \rho_3 \leq \rho_2$ (cf. lemme A.1) donne

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

puisque $t(\cdot) \geq 0$. Par conséquent, il découle de la première inégalité dans (3.3.7) que le troisième terme de droite dans la dernière inégalité est négatif. Ensuite, puisque $\rho_2 \leq 1$, un majorant pour (3.3.15) est donné par

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

qui est négatif dès que la condition (ii) est vérifiée.

Quant au terme $H_2(z, u)$ dans (3.3.16), les premières inégalités (A.5) et (A.6) dans le lemme A.2 donnent

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

ce qui implique que

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

puisque $t'(\cdot) \leq 0$.

Maintenant, en vertu de la deuxième inégalité dans (3.3.7) et grâce au lemme A.1, il est clair que

$$\begin{aligned} 32 \rho_3 + 4 \left(2p + 2n - 5(p \wedge n) - 3 - (p+n+1) \frac{c^2}{b^2} \right) \rho_2 \geq \\ 4 \left(2p + 2n - 4(p \wedge n) + 5 - (p+n+1) \frac{c^2}{b^2} \right) \rho_4. \end{aligned}$$

Donc, comme $t'(\cdot)$ est négative, alors

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

Maintenant, en utilisant le fait que $\rho_3 \geq \rho_4$ et en factorisant $4 \rho_4$, on obtient

$$\begin{aligned} H_2(z, u) \leq & 4 \rho_4 \left\{ 2(p+n-4(p \wedge n) + 3) t(\nu) + 2 \nu t'(\nu) \right. \\ & \left. + \left(2p + 2n - 5(p \wedge n) + 5 - (p+n+1) \frac{c^2}{b^2} \right) \right\} \nu t'(\nu) \\ & + 8 \rho_4 \left\{ t(\nu) + \frac{\rho_3}{\rho_4} \right\} \nu^2 t''(\nu). \end{aligned} \tag{3.3.17}$$

Ensuite, il découle de l'inégalité de Cauchy-Schwarz que

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

où, pour la deuxième inégalité, nous appliquons (A.3) dans le lemme A.1 pour $k = 2$. Par conséquent, nous avons

$$\frac{\rho_3}{\rho_4} = \frac{\text{tr}(S^+) \text{tr}^2(S^{+3})}{\text{tr}(S^{+4}) \text{tr}(S^+)^2} \leq \frac{1}{\rho_3} \leq (p \wedge n)^2, \tag{3.3.18}$$

selon encore (A.3) dans le lemme A.1. Ainsi, grâce à (3.3.18) et en utilisant le fait que $t''(\cdot) \geq 0$, le majorant de $H_2(z, u)$ dans (3.3.17) est négatif si la condition (iii) dans la proposition 3.3.1 est satisfaite. Puisque nous avons $H_1(z, u) \leq 0$, la condition (3.3.14) est satisfaite et donc la condition (3.3.3) est également satisfaite, qui est la condition de dominance dans le corollaire 3.3.1. \square

Exemple

Maintenant, nous allons donner un exemple qui illustre la [proposition 3.3.1](#). Soit

$$t(\nu) = \alpha/(\beta + \nu)$$

où $\alpha > 0$ et $\beta > 0$. Il est facile de voir que

$$t'(\nu) = \frac{-\alpha}{(\beta + \nu)^2} = \frac{-1}{(\beta + \nu)} t(\nu)$$

et

$$t''(\nu) = \frac{2\alpha}{(\beta + \nu)^3} = \frac{2}{(\beta + \nu)^2} t(\nu).$$

Dans ce cas, la condition (i) dans la [proposition 3.3.1](#) devient

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

qui est satisfaite dès que $p + n - 2(p \wedge n) \geq 1$ puisque

$$\frac{\nu}{\beta + \nu} \leq 1. \quad (3.3.19)$$

Pour la condition (ii), puisque $0 \leq t(\nu) \leq t(0) = \alpha/\beta$ alors la condition (ii) est satisfaite si

$$\begin{aligned} \frac{\alpha}{\beta} &\leq \frac{2(p + n - 2(p \wedge n) - 1) \left((p + n + 1)(c^2/b^2) - p - n + (p \wedge) - 2 \right)}{(p + n - 2(p \wedge n) + 1)(p + n - 2(p \wedge n) + 3)} \\ &= t_1(p, n). \end{aligned}$$

Maintenant, la condition (iii) est équivalente à

$$\begin{aligned} &\left\{ 6 \frac{\nu}{\beta + \nu} - 2(p + n - 4(p \wedge n) + 3) \right\} \frac{1}{\beta + \nu} t^2(\nu) \\ &+ \left\{ 4(p \wedge n)^2 \frac{\nu}{\beta + \nu} - \left(2p + 2n - 5(p \wedge n) + 5 - (p + n + 1) \frac{c^2}{b^2} \right) \right\} \frac{1}{\beta + \nu} t(\nu) \leq 0. \end{aligned}$$

À partir de (3.3.19), l'inégalité ci-dessus est satisfaite si

$$\begin{aligned} &\left\{ -2(p + n - 4(p \wedge n)) t(\nu) + 4(p \wedge n)^2 \right. \\ &\quad \left. - \left(2p + 2n - 5(p \wedge n) + 5 - (p + n + 1) \frac{c^2}{b^2} \right) \right\} \frac{1}{\beta + \nu} t(\nu) \leq 0. \end{aligned}$$

Maintenant, en utilisant le fait que $t(\nu) \geq 0$, cette dernière inégalité est vérifiée dès que

$$\begin{aligned} t(\nu) &\leq \frac{(p + n + 1)(c^2/b^2) + 4(p \wedge n)^2 - 2p - 2n + 5(p \wedge n) - 5}{2(p + n - 4(p \wedge n))} \\ &= t_2(p, n), \end{aligned}$$

ce qui est réalisé si

$$\frac{\alpha}{\beta} \leq t_2(p, n),$$

puisque $t(\nu) \leq t(0) = \alpha/\beta$.

Finalement, une condition suffisante pour que les conditions (ii) et (iii) dans la [proposition 3.3.1](#) soient satisfaites, et que

$$\frac{\alpha}{\beta} \leq t_1(p, n) \wedge t_2(p, n).$$

□

3.3.2 Estimateur de type Konno

L'estimateur de Konno [Kon09] dans (1.3.6), donné dans le cadre gaussien et seulement pour le cas où S est inversible, correspond au cas particulier de l'estimateur de type Haff dans la proposition 3.3.1 où la fonction $t(\nu)$ est une constante positive. Par conséquent, l'estimateur que nous donnons dans la proposition suivante, qui découle immédiatement de la proposition 3.3.1, est une extension des résultats de Konno au cadre elliptique dans une approche unifiée (S inversible ou S non inversible).

Proposition 3.3.2

Considérons une densité de la forme (2.5.6) qui satisfait (3.2.18) et tel que

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

Soit $\hat{\Sigma}_{a_o, \Psi}$ un estimateur de la forme

$$\hat{\Sigma}_{a_o, \Psi} = a_o H_1 \text{diag}(l_1 + \nu t, l_2 + \nu t, \dots, l_{p \wedge n} + \nu t) H_1^T,$$

où $\nu = 1/\text{tr}(S^+)$ et t est une constante positive. Alors $\hat{\Sigma}_{a_o, \Psi}$ améliore $\hat{\Sigma}_{a_o}$ aussitôt que

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

Il convient de noter que, dans le contexte de grande dimension ($p > n$), l'estimateur dans la proposition ci-dessus correspond à celui donné par Konno [Kon09] dans le cas gaussien où $c = b = 1$. Dans le cas unifié, nous avons amélioration de l'estimateur de Konno aussitôt que

$$0 \leq t \leq \frac{2(p+n-2(p \wedge n)-1)(p \wedge n - 1)}{(p+n-2(p \wedge n)+1)(p+n-2(p \wedge n)+3)}. \quad (3.3.20)$$

3.4 Synthèse

Dans ce chapitre, nous avons présenté des résultats de dominance d'estimateurs de la forme $a(S + SS^+ G(Z, S))$ par rapport aux estimateurs habituels aS sous un coût quadratique. Grâce à une nouvelle identité de type Stein-Haff dans le lemme 2.7.2, notre approche unifie les cas où S est inversible et S est non inversible. Les résultats sur la structure des valeurs propres de S , dans le lemme B.2, permettent de fournir des exemples généraux d'estimateurs améliorés orthogonalement invariants. Comme mentionné par Haff [Haf80], même dans le cadre gaussien et dans le cas non singulier, les calculs sous le coût quadratique en (3.1.3) sont difficiles. Nous avons pu exécuter de tels calculs dans le contexte étendu des distributions à symétrie elliptique. De plus, nous traitons ici des fonctions plus générales $t(\cdot)$ que celles considérées par Haff [Haf80] et Konno [Kon09]. Nous verrons dans le prochain chapitre qu'une modification de ce coût quadratique facilite les calculs et donne des résultats plus conséquents.

**ESTIMATION D'UNE MATRICE D'ÉCHELLE SOUS UN COÛT
BASÉ SUR LES DONNÉES**

Sommaire

4.1 Introduction	44
4.2 Estimateurs de forme générale	45
4.3 Estimateurs orthogonalement invariants	50
4.3.1 Estimateur de type Haff	52
4.3.2 Estimateur de type Konno	54
4.4 Approche spécifique pour l'estimateur de Konno	54
4.5 Synthèse	55

4.1 Introduction

Comme mentionné par Haff [Haf80], dans le cas gaussien et dans le contexte de petite dimension, nous avons vu dans le chapitre précédent, dans le cas elliptique et dans un cadre unifié, que le coût quadratique (3.1.3) n'est pas facile à manipuler pour comparer différents estimateurs de Σ . Afin de voir où réside cette difficulté, on rappelle que celui-ci est de la forme

$$L(\Sigma, \hat{\Sigma}) = \text{tr}(\Sigma^{-1} \hat{\Sigma} \Sigma^{-1} \hat{\Sigma}) - 2 \text{tr}(\Sigma^{-1} \hat{\Sigma}) + p.$$

C'est le premier terme de droite, où le paramètre inconnu Σ^{-1} est présent deux fois, qui est difficile à gérer. En effet, bien que l'application du corollaire 3.2.1 semble être un moyen pour pallier cette difficulté, ce dernier implique des calculs longs et inabordables. De plus, l'hypothèse de symétrie du facteur de correction $SS^+G(Z, S)$ dans (3.2.1) est nécessaire pour son application. Dans ce chapitre, nous verrons qu'une modification du coût quadratique (3.1.3) nous permettra de construire des estimateurs améliorés, dans la sous-classe de Berger en (3.2.18), pour lesquels le facteur de correction n'est pas nécessairement symétrique. Nous verrons aussi que, sous ce coût, les calculs sont plus simples et les conditions d'amélioration sur le facteur de correction et sur la distribution sont moins fortes.

Considérons une modification du coût (3.1.3) de la forme

$$\begin{aligned} L_S(\Sigma, \hat{\Sigma}) &= \text{tr}(S^+ \Sigma (\Sigma^{-1} \hat{\Sigma} - I_p)^2) \\ &= \text{tr}(\Sigma^{-1} \hat{\Sigma} S^+ \hat{\Sigma}) - 2 \text{tr}(S^+ \hat{\Sigma}) + \text{tr}(S^+ \Sigma). \end{aligned} \quad (4.1.1)$$

Ce type de coût, dit basé sur les données puisqu'il contient S^+ à été introduit par Efron et Morris [EM76] dans le cadre de l'estimation du paramètre de position d'une distribution gaussienne. Il a ensuite été considéré par plusieurs auteurs comme, Fourdrinier et Strawderman [FS15] pour l'estimation du vecteur moyenne dans le cas sphérique, et dans le cas gaussien, par Kubokawa et Srivastava [KS08] pour l'estimation de la matrice de précision Σ^{-1} , et par Tsukuma et Kubokawa [TK15] pour l'estimation de la matrice de covariance Σ . Notons que ce dernier a toutes les propriétés attrayantes des fonctions de coûts (cf. section 1.2).

Il convient de noter que, dans le premier terme à droite dans (4.1.1), le paramètre inconnu Σ^{-1} n'est présent qu'une seule fois, contrairement au coût quadratique (3.1.3) où il est présent deux fois. Ceci représente un avantage considérable pour ce coût basé sur les données par rapport au coût quadratique. En effet l'application du corollaire 3.2.1 n'est plus nécessaire, ce qui implique, comme nous le verrons dans le théorème 4.2.1, des calculs plus faciles et des conditions d'améliorations simples et moins fortes que celles du théorème 3.2.1. Ceci est reflété dans les simulations de la section 5.3.

Dans ce chapitre, sous le coût (4.1.1) et son risque associé

$$R_S(\Sigma, \hat{\Sigma}) = E_{\theta, \Sigma}[L_S(\Sigma, \hat{\Sigma})], \quad (4.1.2)$$

on cherche à améliorer les estimateurs usuels de la forme $\hat{\Sigma}_a = a S$ de la matrice d'échelle du modèle additif (2.5.1). Nous donnons tout d'abord, dans la section 4.2, des résultats d'amélioration d'estimateurs de forme générale, pour la sous-classe de type Berger (3.2.18). Ensuite, dans la section 4.3, nous verrons que ces résultats s'appliquent à la classe des estimateurs orthogonalement invariant, en particulier, aux estimateurs de type Haff dans (1.3.4). Finalement, dans la section 4.4, pour tout estimateur de type Konno dans (1.3.6), nous donnons des conditions d'améliorations qui valent pour toutes les distributions elliptiques (sans être restreint à la sous-classe de Berger).

4.2 Estimateurs de forme générale

Nous considérons des estimateurs de Σ de la forme

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

où a est une constante strictement positive, S^+ est l'inverse de Moore-Penrose de S (cf. définition A.1) et $G(Z, S)$ est une fonction matricielle de dimension $p \times p$. Notons que ici, on élargi la classe d'estimateurs considérés puisque l'on impose pas de symétrie sur $G(Z, S)$ comme dans la section 3.2. Avant d'établir le théorème 4.2.1 sur les conditions d'amélioration des estimateurs de forme $\hat{\Sigma}_{a,G}$, nous donnons au préalable, dans la prochaine proposition, les conditions de finitude de la différence des risques,

$$\Delta(G) = R_S(\Sigma, a(S + S S^+ G(Z, S))) - R_S(\Sigma, a S), \quad (4.2.1)$$

sous le coût (4.1.1) et son risque associé (4.1.2).

Proposition 4.2.1

Supposons que $E_{\theta, \Sigma} [\|S^+ G\|_F^2]$, $E_{\theta, \Sigma} [\|\Sigma^{-1} S S^+ G\|_F^2]$, $E_{\theta, \Sigma} [\text{tr}(\Sigma S^+)]$ et $E_{\theta, \Sigma} [\text{tr}(\Sigma^{-1} S)]$ soient des espérances finies. Alors les risques de $\hat{\Sigma}_a$ et de $\hat{\Sigma}_{a,G}$ sont finis. Dans ce cas, la différence des risques est également finie et peut être écrite sous la forme

$$\Delta(G) = a^2 K^* E_{\theta, \Sigma} [\text{tr}(\Sigma^{-1} S S^+ G \{I_p + S^+ G + S S^+\})] - 2 a E_{\theta, \Sigma} [\text{tr}(S^+ G)]. \quad (4.2.2)$$

Preuve

Commençons d'abord par donner le coût de $\hat{\Sigma}_a$ en Σ , soit

$$\begin{aligned} L_S(\Sigma, \hat{\Sigma}_a) &= a^2 \text{tr}(\Sigma^{-1} S S^+ S) - 2 a \text{tr}(S S^+) + \text{tr}(\Sigma S^+) \\ &= a^2 \text{tr}(\Sigma^{-1} S) - 2 a (p \wedge n) + \text{tr}(\Sigma S^+), \end{aligned} \quad (4.2.3)$$

puisque $S S^+ S = S$ et $\text{tr}(S S^+) = p \wedge n$. En prenant l'espérance dans (4.2.3), il est clair que le risque de $\hat{\Sigma}_a$, soit $R_S(\Sigma, \hat{\Sigma}_a)$, est fini tant que

$$E_{\theta, \Sigma} [\text{tr}(\Sigma^{-1} S)] < \infty \quad (4.2.4)$$

et

$$E_{\theta, \Sigma} [\text{tr}(\Sigma S^+)] < \infty. \quad (4.2.5)$$

Maintenant, nous allons voir sous quelles conditions le risque de $\hat{\Sigma}_{a,G}$ est fini. Le coût de $\hat{\Sigma}_{a,G}$ en Σ est donné par

$$\begin{aligned} L_S(\Sigma, \hat{\Sigma}_{a,G}) &= \text{tr}(S^+ \Sigma (\Sigma^{-1} \hat{\Sigma}_{a,G} - I_p)^2) \\ &= L_S(\Sigma, \hat{\Sigma}_a) + a^2 \text{tr}(\Sigma^{-1} S S^+ G) + a^2 \text{tr}(\Sigma^{-1} S S^+ G S S^+) \\ &\quad + a^2 \text{tr}(\Sigma^{-1} S S^+ G S^+ G) - 2 \text{tr}(S^+ G). \end{aligned} \quad (4.2.6)$$

Sous l'hypothèse

$$E_{\theta, \Sigma} [\|\Sigma^{-1} S S^+ G\|_F^2] < \infty \quad (4.2.7)$$

et

$$E_{\theta, \Sigma} [\|S^+ G\|_F^2] < \infty, \quad (4.2.8)$$

nous allons voir que, en plus des conditions (4.2.4) et (4.2.5) qui assurent la finitude du risque de $\hat{\Sigma}_a$, les conditions (4.2.7) et (4.2.8) suffisent pour assurer la finitude du risque de $\hat{\Sigma}_{a,G}$. La suite de la démonstration repose sur plusieurs applications de l'inégalité de Cauchy-Schwarz dans (3.2.4). On l'applique une première fois pour $A = I_p$ et $B = S^+ G$ et on obtient

$$\text{tr}^2(\Sigma^{-1} S S^+ G) \leq p \|\Sigma^{-1} S S^+ G\|_F^2. \quad (4.2.9)$$

On l'applique une deuxième fois pour $A = S S^+$ et $B = \Sigma^{-1} S S^+$ pour obtenir

$$\text{tr}^2(\Sigma^{-1} S S^+ G S S^+) \leq \|S S^+\|_F^2 \|\Sigma^{-1} S S^+ G\|_F^2 = p \wedge n \|\Sigma^{-1} S S^+ G\|_F^2, \quad (4.2.10)$$

puisque $\|S S^+\|_F^2 = \text{tr}(S^+ S S S^+) = \text{tr}(S^+ S) = p \wedge n$. Une autre application de (3.2.4) pour $A = \Sigma^{-1} S S^+ G$ et $B = S^+ G$ donne

$$|\text{tr}(\Sigma^{-1} S S^+ G S^+ G)| \leq \|\Sigma^{-1} S S^+ G\|_F \|S^+ G\|_F. \quad (4.2.11)$$

Une dernière application de (3.2.4) pour $A = I_p$, $B = S^+G$ donne

$$\mathrm{tr}^2(I_p S^+G) \leq p \|S^+G\|_F^2. \quad (4.2.12)$$

En prenant l'espérance dans (4.2.9), (4.2.10), (4.2.11) et (4.2.12), il est facile de voir que, sous les hypothèses (4.2.7) et (4.2.8), les espérances $E_{\theta,\Sigma}[\mathrm{tr}^2(\Sigma^{-1}SS^+GSS^+)]$, $E_{\theta,\Sigma}[\mathrm{tr}^2(S^+G)]$ et $E_{\theta,\Sigma}[\mathrm{tr}^2(\Sigma^{-1}SS^+GS^+G)]$ sont finies, puisque l'inégalité de Cauchy-Schwarz donne

$$\begin{aligned} E_{\theta,\Sigma}|\mathrm{tr}(\Sigma^{-1}SS^+GS^+G)| &\leq E_{\theta,\Sigma}(\|\Sigma^{-1}SS^+G\|_F \|S^+G\|_F) \\ &\leq E_{\theta,\Sigma}(\|\Sigma^{-1}SS^+G\|_F^2 E_{\theta,\Sigma}\|S^+G\|_F^2)^{1/2}. \end{aligned}$$

Ainsi, en prenant l'espérance dans (4.2.6), les termes qui interviennent dans le risque de $\hat{\Sigma}_{a,G}$, à savoir, $E_{\theta,\Sigma}[\mathrm{tr}(\Sigma^{-1}SS^+GSS^+)]$, $E_{\theta,\Sigma}[\mathrm{tr}(\Sigma^{-1}SS^+GS^+G)]$ et $E_{\theta,\Sigma}[\mathrm{tr}(S^+G)]$, sont aussi finis. Par conséquent, sous les hypothèses (4.2.4), (4.2.5), (4.2.7) et (4.2.8), les risques de $\hat{\Sigma}_a$ et de $\hat{\Sigma}_{a,G}$ sont finis. Dans ce cas, la différence des risques dans (4.2.1) est également finie et peut être écrite sous la forme

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

□

L'amélioration de la classe d'estimateurs usuels de la forme $\hat{\Sigma}_a = aS$ sera illustrée par l'amélioration de $\hat{\Sigma}_{a_o,G} = a_oS + a_oSS^+G(Z,S)$ où a_o est la constante optimale (celle qui minimise le coût (4.1.1)) dans la classe des estimateurs $\hat{\Sigma}_{a_o} = a_oS$. La proposition suivante donne la valeur de a_o .

Proposition 4.2.2

Pour la classe d'estimateurs $\hat{\Sigma}_a = aS$, où $a > 0$, la constante optimale sous le risque (4.1.2) est

$$a_o = \frac{1}{K^*(p \vee n)}, \quad (4.2.13)$$

où la constante de normalisation K^* est donnée dans (2.6.2).

Preuve

Soit $\hat{\Sigma}_a = aS$, où a est une constante strictement positive. Alors, en vertu de (3.1.4), le risque de Σ en $\hat{\Sigma}_a$ est donné par

$$\begin{aligned} R_S(\Sigma, \hat{\Sigma}_a) &= E_{\theta,\Sigma}[L_S(\Sigma, \hat{\Sigma}_a)] \\ &= a^2 E_{\theta,\Sigma} [\mathrm{tr}(\Sigma^{-1}SS^+S)] - 2a E_{\theta,\Sigma} [\mathrm{tr}(SS^+)] + E_{\theta,\Sigma} [\mathrm{tr}(S^+\Sigma)], \end{aligned}$$

puisque $S = SS^+S$. Pour traiter le terme de droite qui dépend du paramètre inconnu Σ^{-1} , on applique l'identité de type Stein-Haff dans le lemme 2.7.2 pour $G = S$ et on obtient

$$\begin{aligned} R_S(\Sigma, \hat{\Sigma}_a) &= a^2 K^* E_{\theta,\Sigma}^* [\mathrm{tr} [2S^+S \mathcal{D}_s\{S\} + (n - (p \wedge n) - 1) SS^+]] \\ &\quad - 2a E_{\theta,\Sigma} [\mathrm{tr}(SS^+)] + E_{\theta,\Sigma} [\mathrm{tr}(S^+\Sigma)]. \end{aligned} \quad (4.2.14)$$

Maintenant, grâce à (B.12) et en utilisant les faits que $S(I_p - SS^+) = 0_{p \times p}$ et $\mathrm{tr}(SS^+) = p \wedge n$, le risque dans (4.2.14) devient

$$R_S(\Sigma, \hat{\Sigma}_a) = a^2 K^* (p \wedge n) (p + n - p \wedge n) - 2a (p \wedge n) + E_{\theta,\Sigma} [\mathrm{tr}(S^+\Sigma)].$$

Clairement, le choix de $a = 1/K^*(p + n - p \wedge n) = 1/K^*(p \vee n)$ est optimal sous le risque (4.1.2). \square

Il convient de noter que, sous le coût quadratique basé sur les données (4.1.1), la constante optimale dans (4.2.13) dépend de l'inversibilité de S . En effet, lorsque $p \leq n$ (S est inversible), nous avons $a_o = 1/n K^*$ et, lorsque $p > n$ (S n'est pas inversible), nous avons $a_o = 1/p K^*$. Ceci est en contraste avec le coût quadratique (3.1.3) puisque, dans les deux derniers cas, $a_o = 1/(p + n + 1)K^{**}$.

Dans la différence des risques (4.2.2) où le paramètre inconnu Σ^{-1} n'est présent qu'une fois, une seule application de l'identité de type Stein-Haff dans le lemme 2.7.2 suffit à le remplacer. Il n'est donc plus nécessaire de définir une nouvelle espérance $E_{\theta, \Sigma}^{**}$, comme dans la section 3.2. Nous donnons dans la proposition suivante une expression de la différence des risques dans (4.2.2) où le paramètre de nuisance Σ^{-1} n'intervient pas.

Proposition 4.2.3

Soit $G(z, s)$ une fonction matricielle de dimension $p \times p$ telle que, pour tout $z \in \mathbb{R}^{q \times p}$ fixé, la fonction $G(z, s)$ est faiblement différentiable par rapport à $s \in \mathbb{R}^{p \times p}$. Supposons que les espérances $E_{\theta, \Sigma} [\|S^+ G\|_F^2]$, $E_{\theta, \Sigma} [\|\Sigma^{-1} S S^+ G\|_F^2]$, $E_{\theta, \Sigma} [\text{tr}(\Sigma S^+)]$ et $E_{\theta, \Sigma} [\text{tr}(\Sigma^{-1} S)]$ soient finies. Alors

$$\begin{aligned} \Delta(G) = & a^2 K^* E_{\theta, \Sigma}^* [(n - (p \wedge n) - 1) \text{tr}(S^+ G + (S^+ G)^2 + S^+ G S S^+) + \\ & 2 \text{tr}(S S^+ \mathcal{D}_s \{S S^+ G + S S^+ G S^+ G + S S^+ G S S^+\}^T)] - 2 a E_{\theta, \Sigma} [\text{tr}(S^+ G)]. \end{aligned} \quad (4.2.15)$$

Preuve

Rappelons que les conditions annoncées dans cette proposition sont celles qui assurent la finitude de la différence des risques. De plus, comme nous allons le voir, les deux dernières, à savoir, $E_{\theta, \Sigma} [\|S^+ G\|_F^2] < \infty$, $E_{\theta, \Sigma} [\|\Sigma^{-1} S S^+ G\|_F^2] < \infty$ impliquent la condition d'application du lemme 2.7.2 pour $G(I_p + S^+ G + S S^+)$ au lieu de G , c'est-à-dire

$$E_{\theta, \Sigma} |\text{tr}(\Sigma^{-1} S S^+ \{G + G S^+ G + G S S^+\})| < \infty. \quad (4.2.16)$$

En effet, l'inégalité triangulaire donne

$$\begin{aligned} |\text{tr}(\Sigma^{-1} S S^+ \{G + G S^+ G + G S S^+\})| \leq & |\text{tr}(\Sigma^{-1} S S^+ G)| + |\text{tr}(\Sigma^{-1} S S^+ G S^+ G)| \\ & + |\text{tr}(\Sigma^{-1} S S^+ G S S^+)|. \end{aligned} \quad (4.2.17)$$

L'application de l'inégalité de Cauchy-Schwarz (3.2.4) pour tous les termes de droite dans l'inégalité (4.2.17) donne

$$\begin{aligned} |\text{tr}(\Sigma^{-1} S S^+ G)| & \leq p \|\Sigma^{-1} S S^+ G\|_F, \\ |\text{tr}(\Sigma^{-1} S S^+ G S^+ G)| & \leq \|\Sigma^{-1} S S^+ G\|_F \|S^+ G\|_F \end{aligned}$$

et

$$|\text{tr}(\Sigma^{-1} S S^+ G S S^+)| \leq (p \wedge n) \|\Sigma^{-1} S S^+ G\|_F.$$

En prenant l'espérance dans (4.2.17) et, par conséquent, dans ces trois dernières inégalités, il est clair que la condition (4.2.16) est satisfaite si

$$E_{\theta, \Sigma} [\|\Sigma^{-1} S S^+ G\|_F] < \infty \quad (4.2.18)$$

et

$$E_{\theta, \Sigma} [\|\Sigma^{-1} S S^+ G\|_F \|S^+ G\|_F] < \infty. \quad (4.2.19)$$

Maintenant, grâce, une fois de plus, à l'inégalité de Cauchy-Schwarz, nous avons

$$E_{\theta,\Sigma}[\|\Sigma^{-1}SS^+G\|_F \|S^+G\|_F] \leq (E_{\theta,\Sigma}\|\Sigma^{-1}SS^+G\|_F^2 E_{\theta,\Sigma}\|S^+G\|_F^2)^{1/2}.$$

Ainsi, il est facile de voir que, si les espérances $E_{\theta,\Sigma}[\|S^+G\|_F^2]$ et $E_{\theta,\Sigma}[\|\Sigma^{-1}SS^+G\|_F^2]$ sont finies, alors les conditions (4.2.18) et (4.2.19) sont satisfaites et, par conséquent, celle dans (4.2.16) est aussi satisfaite.

Enfin, sous la condition (4.2.16), l'application du lemme 2.7.2 pour $G(I_p + S^+G + SS^+)$ donne le résultat souhaité. \square

La différence des risques (4.2.15) dépend des espérances $E_{\theta,\Sigma}$ et $E_{\theta,\Sigma}^*$. Afin d'avoir une expression homogène qui ne dépend que de $E_{\theta,\Sigma}$, nous verrons (dans la preuve du prochain théorème) qu'il suffit de réduire le contexte distributionnel à la sous-classe de Berger dans (3.3.7) où, pour tout $t \in \mathbb{R}^+$,

$$c \leq \frac{F^*(t)}{f(t)} \leq b$$

où c et b sont des constantes strictement positives (cf. section 2.6) et où $F^*(t)$ et $f(t)$ sont respectivement donnés dans (2.6.1) et (2.5.6).

Nous donnons dans le théorème suivant, qui est le résultat principal de cette section, des conditions d'amélioration de $\hat{\Sigma}_{a_o,G}$ sur $\hat{\Sigma}_{a_o}$ sous le coût (4.1.1), où a_o est la constante optimale dans (4.2.13).

Théorème 4.2.1

Considérons une densité de la forme (2.5.6) qui satisfait (2.6.4). Sous la condition

$$\text{tr}(S^+G) \geq 0, \tag{4.2.20}$$

l'estimateur $\hat{\Sigma}_{a_o,G} = a_o(S + SS^+G(Z, S))$ améliore $\hat{\Sigma}_{a_o} = a_o S$ dès que

$$\begin{aligned} \text{tr} \left[(n - (p \wedge n) - 1)(S^+GSS^+ + (S^+G)^2) + (n - (p \wedge n) - 2\frac{c}{b})(p \vee n) - 1)S^+G \right. \\ \left. + 2SS^+ \mathcal{D}_s\{SS^+G + SS^+GS^+G + SS^+GSS^+\}^T \right] \leq 0. \end{aligned} \tag{4.2.21}$$

Preuve

En utilisant (2.6.8), on n'exprime la différence des risques (4.2.15) qu'au travers de l'espérance $E_{\theta,\Sigma}$ comme suit

$$\begin{aligned} \Delta(G) = a_o^2 E_{\theta,\Sigma}[\varphi_{\theta,\Sigma}^*(Z, U) \left(\text{tr} \left[(n - (p \wedge n) - 1)(SG^+ + (S^+G)^2 + S^+GSS^+) \right. \right. \\ \left. \left. + 2SS^+ \mathcal{D}_s\{SS^+G + SS^+GS^+G + SS^+GSS^+\}^T \right] \right. \\ \left. - \frac{2}{a_o \varphi_{\theta,\Sigma}^*(Z, U)} \text{tr}(S^+G) \right)], \end{aligned}$$

où a_o est donné dans (4.2.13). Maintenant, en posant $H(Z, U) = 1$ dans (2.6.8) nous avons

$$K^* = E_{\theta,\Sigma}[\varphi_{\theta,\Sigma}^*(Z, U)]. \tag{4.2.22}$$

Comme nous sommes dans la sous-classe de Berger dans (2.6.4), il est clair que $c \leq \varphi_{\theta,\Sigma}^*(Z, U) \leq b$ et donc, grâce à (4.2.22), nous avons $c \leq K^* \leq b$. Par conséquent

$$\frac{-1}{a_o \varphi_{\theta,\Sigma}^*(Z, U)} = \frac{-K^*(p \vee n)}{\varphi_{\theta,\Sigma}^*(Z, U)} \leq -\frac{c}{b}(p \vee n).$$

Alors, sous la condition (4.2.20), la différence des risques est majorée par

$$a_o^2 E_{\theta, \Sigma} [\varphi_{\theta, \Sigma}^*(Z, U) (\text{tr} [2 SS^+ \mathcal{D}_s \{SS^+ G + SS^+ GS^+ G + SS^+ GSS^+\}^T + (n - (p \wedge n) - 1)(S^+ GSS^+ + (S^+ G)^2) + (n - (p \wedge n) - 2 \frac{c}{b}(p \vee n) - 1)S^+ G])]$$

Ainsi, puisque $\varphi_{\theta, \Sigma}^*(Z, U)$ est positif, l'estimateur $\hat{\Sigma}_{a_o, G}$ domine $\hat{\Sigma}_{a_o}$ dès que l'inégalité (4.2.21) est satisfaite. \square

Comme mentionner en début de chapitre (cf. section 4.1), nous constatons que les conditions d'amélioration dans le théorème 4.2.1, impliquant le coût basé sur les données (4.1.1), sont plus simples à interpréter et moins fortes que celles du théorème 3.2.1, impliquant le coût quadratique (3.1.3). Ceci nous permettra (comme nous le verrons dans la prochaine section où nous appliquons notre théorie à la classe des estimateurs orthogonalement invariants) d'élargir la classe d'estimateurs améliorés de type Haff.

4.3 Estimateurs orthogonalement invariants

Dans cette partie, en se basant sur la décomposition en valeurs propres de S , nous cherchons, sous le coût (4.1.1), à améliorer l'estimateur optimal

$$\hat{\Sigma}_{a_o} = a_o S = a_o H_1 \text{diag}(l_1, l_2, \dots, l_{p \wedge n}) H_1^T,$$

où a_o est donné dans (4.2.13), par un estimateur alternatif de la forme

$$\hat{\Sigma}_{a_o, \Psi} = a_o H_1 (L + \Psi(L)) H_1^T,$$

où $H_1 \in \mathcal{L}_{p, p \wedge n}$ et $L, \Psi(L) \in \mathcal{D}_{p \wedge n}$ (cf. lemme B.1). L'application du théorème 4.2.1 conduit au corollaire suivant.

Corollaire 4.3.1

Considérons une densité de la forme (2.5.6) qui satisfait (2.6.3). Sous la condition

$$\sum_{i=1}^{p \wedge n} \frac{\psi_i}{l_i} \geq 0, \quad (4.3.1)$$

l'estimateur $\hat{\Sigma}_{a_o, \Psi} = a_o H_1 \text{diag}(l_1 + \psi_1, l_2 + \psi_2, \dots, l_{p \wedge n} + \psi_{p \wedge n}) H_1^T$ améliore l'estimateur optimal $\hat{\Sigma}_{a_o} = a_o H_1 \text{diag}(l_1, l_2, \dots, l_{p \wedge n}) H_1^T$ dès que

$$\sum_{i=1}^{p \wedge n} \left\{ (n - (p \wedge n) - 1) \frac{\psi_i^2}{l_i^2} + 2 \left(n - (p \wedge n) - \frac{c}{b}(p \vee n) - 1 \right) \frac{\psi_i}{l_i} + (p - p \wedge n) \frac{\phi_i}{l_i} + 2 \frac{\partial \phi_i}{\partial l_i} + \sum_{i \neq j} \frac{\phi_i - \phi_j}{l_i - l_j} \right\} \leq 0, \quad (4.3.2)$$

où

$$\phi_i = 2 \psi_i + \frac{\psi_i^2}{l_i}. \quad (4.3.3)$$

Preuve

La preuve consiste à prouver les conditions (4.2.20) et (4.2.21) du [théorème 4.2.1](#). Rappelons que, $S^+ = H_1 L^{-1} H_1^T$, $SS^+ = H_1 H_1^T$, $SS^+G = G = H_1 \Psi H_1^T$ et $S^+G = H_1 L^{-1} \Psi H_1^T$. Dans ce contexte, la condition (4.2.20) devient

$$\text{tr}(H_1 L^{-1} \Psi H_1^T) = \text{tr}(H_1^T H_1 L^{-1} \Psi) = \sum_{i=1}^{p \wedge n} \frac{\psi_i}{l_i} \geq 0,$$

puisque $H_1^T H_1 = I_{p \wedge n}$ qui est réalisée par hypothèse. Quant à la condition (4.2.21), notons tout d'abord que

$$(SS^+G + SS^+GS^+G + SS^+GSS^+)^T = H_1 H_1^T H_1 \Psi H_1^T + H_1 H_1^T H_1 \Psi H_1^T H_1 L^{-1} \Psi H_1^T + H_1 H_1^T H_1 \Psi H_1^T H_1 H_1^T.$$

Puisque $H_1^T H_1 = I_{p \wedge n}$ alors

$$(SS^+G + SS^+GS^+G + SS^+GSS^+)^T = H_1 [2\Psi + \Psi^2 L^{-1}] H_1^T = H_1 \Phi H_1^T,$$

où $\Phi = [2\Psi + \Psi^2 L^{-1}] \in \mathcal{D}_{p \wedge n}$ puisque $\Psi(L)$ et $L^{-1} \in \mathcal{D}_{p \wedge n}$. En passant à la trace, on a

$$\text{tr} [SS^+ \mathcal{D}_s \{SS^+G + SS^+GS^+G + SS^+GSS^+\}^T] = \text{tr} [H_1 H_1^T \mathcal{D}_s \{H_1 \Phi H_1^T\}].$$

Maintenant, puisque $H_1 H_1^T (I_p - H_1 H_1^T) = 0$, l'application du [lemme B.2](#) donne

$$\begin{aligned} \text{tr} [SS^+ \mathcal{D}_s \{SS^+G + SS^+GS^+G + SS^+GSS^+\}^T] &= \text{tr}(H_1 \Phi^{(1)} H_1^T) \\ &= \text{tr}(\Phi^{(1)}), \end{aligned} \tag{4.3.4}$$

où $\Phi^{(1)} \in \mathcal{D}_{p \wedge n}$ avec

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

Par ailleurs, nous avons

$$\begin{aligned} &\text{tr} \left[(n - (p \wedge n) - 1)(S^+GSS^+ + (S^+G)^2) + (n - (p \wedge n) - 2 \frac{c}{b} (p \vee n) - 1)S^+G \right] \\ &= \text{tr} \left[(n - (p \wedge n) - 1)(H_1 L^{-1} \Psi H_1^T + H_1 (L^{-1} \Psi)^2 H_1^T) \right] \\ &\quad + \text{tr} \left[(n - (p \wedge n) - 2 \frac{c}{b} (p \vee n) - 1)H_1 L^{-1} \Psi H_1^T \right] \\ &= \text{tr} \left[(n - (p \wedge n) - 1)(L^{-1} \Psi)^2 + 2 \left(n - (p \wedge n) - \frac{c}{b} (p \vee n) - 1 \right) L^{-1} \Psi \right]. \end{aligned} \tag{4.3.5}$$

Finalement, en combinant (4.3.4) et (4.3.5), la condition (4.2.21) coïncide avec (4.3.2). \square

Comme sous le coût quadratique, nous donnons maintenant, sous le coût basé sur les données, un exemple d'estimateur de type Haff de la forme (1.3.4).

4.3.1 Estimateur de type Haff

Proposition 4.3.1

Considérons une densité de la forme (2.5.6) qui satisfait (2.6.3) et telle que

$$\frac{c}{b} \geq \frac{(p \vee n) - (p \wedge n) + 1}{(p \vee n)}. \quad (4.3.6)$$

Soit $\hat{\Sigma}_{a_o, \Psi}$, où a_o est donné dans (4.2.13), et soit un estimateur de la forme

$$\begin{aligned} \hat{\Sigma}_{a_o, \Psi} &= a_o H_1 (L + \Psi(L)) H_1^T \\ &= a_o H_1 \text{diag}(l_1 + \nu t(\nu), l_2 + \nu t(\nu), \dots, l_{p \wedge n} + \nu t(\nu)) H_1^T, \end{aligned}$$

où $\nu = 1/\text{tr}(S^+)$ et $t(\cdot)$ est une fonction positive, décroissante et différentiable. Alors $\hat{\Sigma}_{a_o, \Psi}$ améliore $\hat{\Sigma}_{a_o}$ dès que

$$0 \leq t(\nu) \leq 2 \frac{((p \wedge n) - 1) + (p \vee n)(c/b - 1)}{(p \vee n) - (p \wedge n) + 1}. \quad (4.3.7)$$

Preuve

La preuve consiste à vérifier les conditions (4.3.1) et (4.3.2) du corollaire 4.3.1. Dans le contexte des estimateurs orthogonalement invariants, puisque $\sum_{i=1}^{p \wedge n} (1/l_i) = 1/\nu$, la première condition devient

$$\sum_{i=1}^{p \wedge n} \frac{\nu}{l_i} t(\nu) = t(\nu) \geq 0,$$

qui est vérifiée par hypothèse.

Quant à la condition (4.3.2), notons tout d'abord que, dans ce contexte, (4.3.3) devient

$$\phi_i = 2\nu t(\nu) + \frac{\nu^2}{l_i} t^2(\nu).$$

Il s'ensuit que

$$\sum_{i=1}^{p \wedge n} \frac{\phi_i}{l_i} = 2t(\nu) + \nu^2 t^2(\nu), \quad (4.3.8)$$

$$\sum_{i=1}^{p \wedge n} \frac{\partial \phi_i}{\partial l_i} = \sum_{i=1}^{p \wedge n} \frac{\partial}{\partial l_i} \left(2\nu t(\nu) + \frac{\nu^2}{l_i} t^2(\nu) \right) \quad (4.3.9)$$

et

$$\sum_{i=1}^{p \wedge n} \sum_{j \neq i}^{p \wedge n} \frac{\phi_i - \phi_j}{l_i - l_j} = \nu^2 t^2(\nu) \sum_{i=1}^{p \wedge n} \sum_{j \neq i}^{p \wedge n} \frac{l_i^{-1} - l_j^{-1}}{l_i - l_j}, \quad (4.3.10)$$

puisque $\rho_k = \nu^k \operatorname{tr}(S^{+k})$. Grâce à (3.3.9), nous avons

$$\begin{aligned} \frac{\partial \phi_i}{\partial l_i} &= 2 \frac{\partial}{\partial l_i} (\nu t(\nu)) + \frac{\partial}{\partial l_i} \left(\frac{\nu^2}{l_i} t^2(\nu) \right) \\ &= \left(2 \frac{\nu^3}{l_i^3} - \frac{\nu^2}{l_i^2} \right) t^2(\nu) + 2 \frac{\nu^2}{l_i^2} t(\nu) + 2 \frac{\nu^3}{l_i^2} t'(\nu) + 2 \frac{\nu^4}{l_i^3} t(\nu) t'(\nu). \end{aligned} \quad (4.3.11)$$

Ainsi, en remplaçant (4.3.11) dans (4.3.9), on obtient

$$\sum_{i=1}^{p \wedge n} \frac{\partial \phi_i}{\partial l_i} = (2\rho_3 - \rho_2) t^2(\nu) + 2\rho_2 t(\nu) + 2\nu(\rho_2 + \rho_3 t(\nu)) t'(\nu). \quad (4.3.12)$$

Maintenant, en combinant (4.3.8), (4.3.10) et (4.3.12), le terme de gauche dans l'inégalité (4.3.2), qu'on note par $H(z, u)$, devient

$$\begin{aligned} H(z, u) &= (n - (p \wedge n) - 1) \rho_2 t^2(\nu) + 2 \left(n - (p \wedge n) - \frac{c}{b}(p \vee n) - 1 \right) t(\nu) \\ &\quad + 2(2\rho_3 - \rho_2) t^2(\nu) + 4\rho_2 t(\nu) + 4\nu(\rho_2 + \rho_3 t(\nu)) t'(\nu) \\ &\quad + (p - p \wedge n) (2t(\nu) + \rho_2 t^2(\nu)) + \nu^2 t^2(\nu) \sum_{i=1}^{p \wedge n} \sum_{j \neq i}^{p \wedge n} \frac{l_i^{-1} - l_j^{-1}}{l_i - l_j}. \end{aligned} \quad (4.3.13)$$

Puisque nous avons

$$\sum_{i=1}^{p \wedge n} \sum_{j \neq i}^{p \wedge n} \frac{l_i^{-1} - l_j^{-1}}{l_i - l_j} \leq 0,$$

on majore (4.3.13) comme suit :

$$\begin{aligned} H(z, u) &\leq (4\rho_3 + (p + n - 2(p \wedge n) - 3)\rho_2) t^2(\nu) + 4\nu(\rho_3 t(\nu) + \rho_2) t'(\nu) \\ &\quad \left(4\rho_2 + 2(p - p \wedge n) + 2 \left(n - (p \wedge n) - \frac{c}{b}(p \vee n) - 1 \right) \right) t(\nu) \\ &\leq (4\rho_3 + (p + n - 2(p \wedge n) - 3)\rho_2) t^2(\nu) \\ &\quad \left(4\rho_2 + 2(p - p \wedge n) + 2 \left(n - (p \wedge n) - \frac{c}{b}(p \vee n) - 1 \right) \right) t(\nu) \end{aligned} \quad (4.3.14)$$

puisque $t'(\cdot) \leq 0$ et $t(\cdot) \geq 0$. Maintenant, en utilisant le fait que $\rho_3 \leq \rho_2 \leq 1$ et grâce à (4.3.6), on majore (4.3.14) par

$$H(z, u) \leq (p + n - 2(p \wedge n) + 1) t^2(\nu) + 2 \left(p + n + 1 - 2(p \wedge n) - \frac{c}{b}(p \vee n) \right) t(\nu). \quad (4.3.15)$$

Puisque

$$p + n - 2(p \wedge n) = (p \vee n) - (p \wedge n),$$

alors, grâce à (4.3.6), la condition (4.3.2) est satisfaite si (4.3.15) est négatif, c'est-à-dire si la condition (4.3.7) est vérifiée. \square

En comparant les conditions d'amélioration de type Haff énoncées dans la proposition [proposition 4.3.1](#) avec celles de la [proposition 3.3.1](#), on constate que l'utilisation du coût basé sur les données (4.1.1) nous a conduit à obtenir des conditions plus faibles sur la fonction $t(\nu)$, soit la condition (4.3.7). De plus, on a élargi la classe distributionnelle qui dans ce cas nécessite uniquement que le rapport c/b soit minoré comme dans la condition (4.3.6).

4.3.2 Estimateur de type Konno

L'estimateur de type Konno dans (1.3.6), que nous donnons dans la prochaine proposition et qui découle immédiatement de la proposition 4.3.1, est le cas particulier de l'estimateur de type Haff donné dans la section 4.3.1 quand la fonction $t(\cdot)$ est une constante positive.

Proposition 4.3.2

Considérons une densité de la forme (2.5.6) qui satisfait (2.6.3) et telle que

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

Soit $\hat{\Sigma}_{a_o, \Psi}$, où a_o est donné dans (4.2.13), et soit un estimateur

$$\begin{aligned} \hat{\Sigma}_{a_o, \Psi} &= a_o H_1 (L + \Psi(L)) H_1^T \\ &= a_o H_1 \text{diag}(l_1 + \nu t, l_2 + \nu t, \dots, l_{p \wedge n} + \nu t) H_1^T, \end{aligned}$$

où $\nu = 1/\text{tr}(S^+)$ et t est une constante positive. Alors $\hat{\Sigma}_{a_o, \Psi}$ améliore $\hat{\Sigma}_{a_o}$ dès que

$$0 \leq t \leq 2 \frac{((p \wedge n) - 1) + (p \vee n)(c/b - 1)}{(p \vee n) - (p \wedge n) + 1}. \quad (4.3.16)$$

Il convient de remarquer que, sous le coût basé sur les données (4.1.1), la borne supérieure de la condition d'amélioration dans (4.3.16) est plus grande que celle, associée au coût quadratique, dans (3.3.20). Ceci représente un avantage du coût basé sur les données dans la mesure où il offre la possibilité de construire un plus grand nombre d'estimateurs améliorés.

4.4 Approche spécifique pour l'estimateur de Konno

Dans cette section, nous montrons, par une démonstration spécifique, la dominance de l'estimateur de Konno $\hat{\Sigma}_{a_o, G}^K$ dans (1.3.6) sur l'estimateur optimal $\hat{\Sigma}_{a_o}$ où a_o est donné dans (4.2.13). Contrairement à la proposition 4.3.2, cette approche ne nécessite pas la restriction à la sous-classe de type Berger dans (3.2.18), autrement dit la proposition 4.4.1 ne suppose aucune condition sur la classe distributionnelle.

Proposition 4.4.1

Considérons une densité de la forme (2.5.6). Soit

$$\hat{\Sigma}_{a_o, G}^K = a_o (S + \nu t S S^+)$$

où $\nu = 1/\text{tr}(S^+)$ et t est une constante positive. Alors $\hat{\Sigma}_{a_o, G}^K$ améliore $\hat{\Sigma}_{a_o}$ dès que

$$0 \leq t \leq \frac{2((p \wedge n) - 1)}{(p \vee n) - (p \wedge n) + 1}. \quad (4.4.1)$$

Preuve

Il est clair que l'estimateur $\hat{\Sigma}_{a_o, G}^K$ est de la forme (3.2.1) où $G(Z, S) = t I_p / \text{tr}(S^+)$. Dans ce contexte, grâce à la proposition 4.2.3, la différence des risques entre $\hat{\Sigma}_{a_o, G}^K$ et $\hat{\Sigma}_{a_o}$ est donnée par

$$\begin{aligned} \Delta(G) &= a_o^2 K^* E_{\theta, \Sigma}^* \left[(n - (p \wedge n) - 1) \text{tr}(t \nu S^+ + t^2 \nu^2 S^{+2} + t \nu S^+) + \right. \\ &\quad \left. 2 \text{tr}(S S^+ \mathcal{D}_s \{t \nu S S^+ + t^2 \nu^2 S^+ + t \nu S S^+\}) \right] - 2 a_o E_{\theta, \Sigma} \left[\text{tr}(t \nu S^+) \right] \\ &= a_o^2 K^* E_{\theta, \Sigma}^* \left[(n - (p \wedge n) - 1) \text{tr}(2 t \nu S^+ + t^2 \nu^2 S^{+2}) + \right. \\ &\quad \left. 2 \text{tr}(S S^+ \mathcal{D}_s \{2 t \nu S S^+ + t^2 \nu^2 S^+\}) \right] - 2 a_o t, \end{aligned} \quad (4.4.2)$$

puisque $\text{tr}(\nu S^+) = \text{tr}(S^+)/\text{tr}(S^+) = 1$.

En outre, l'application du [lemme B.11](#) donne

$$2 \text{tr} (SS^+ \mathcal{D}_s \{2t \nu SS^+\}) = 2 [(p - p \wedge n) + 2\rho_2] t$$

et

$$2 \text{tr} (SS^+ \mathcal{D}_s \{t^2 \nu^2 S^+\}) = \left[(p - p \wedge n - 2) \rho_2 + 4 \rho_3 + 2\nu^2 \sum_i^{p \wedge n} \sum_{i \neq j} \frac{l_i^{-1} - l_j^{-1}}{l_i - l_j} \right] t^2$$

où $\rho_i = \nu^i/l_i$ (cf. [lemme A.1](#)). Ainsi, en combinant ces deux dernières égalités, nous avons

$$\begin{aligned} 2 \text{tr} (SS^+ \mathcal{D}_s \{2t \nu SS^+ + t^2 \nu^2 S^+\}) &= 2(p - p \wedge n)t + 4\rho_2 t + (p - p \wedge n)\rho_2 t^2 \\ &\quad - 2\rho_2 t^2 + 4\rho_3 t^2 + 2\nu^2 t^2 \sum_i^{p \wedge n} \sum_{i \neq j} \frac{l_i^{-1} - l_j^{-1}}{l_i - l_j}. \end{aligned} \quad (4.4.3)$$

En remplaçant (4.4.3) dans (4.4.2) et en utilisant le fait que $\sum_i^{p \wedge n} \sum_{i \neq j} (l_i^{-1} - l_j^{-1}/(l_i - l_j)) < 0$, la différence des risques dans (4.4.2) est majorée comme suit :

$$\begin{aligned} \Delta(G) &\leq a_o^2 K^* [(p + n - 2(p \wedge n) - 1)t^2 + 2(p + n - 2(p \wedge n) + 1)t] - 2a_o t \\ &= a_o^2 K^* \left[(p + n - 2(p \wedge n) - 1)t^2 + 2(p + n - 2(p \wedge n) + 1)t - \frac{2}{a_o K^*} t \right] \end{aligned}$$

puisque $\rho_3 \leq \rho_2 \leq 1$ (cf. [lemme A.1](#)). Maintenant comme $a_o K^* = 1/(p \wedge n)$, il vient que

$$\begin{aligned} \Delta(G) &\leq a_o^2 K^* [(p + n - 2(p \wedge n) + 1)t^2 + 2(p + n - 2(p \wedge n) - (p \vee n) + 1)t] \\ &= a_o^2 K^* [((p \vee n) - (p \wedge n) + 1)t^2 + 2(1 - p \wedge n)t], \end{aligned}$$

puisque $p + n - 2(p \vee n) = (p \vee n) - (p \wedge n)$.

Par conséquent, $\hat{\Sigma}_{a_o, G}^K$ améliore $\hat{\Sigma}_{a_o}$ tant que la condition (4.4.1) est satisfaite. \square

4.5 Synthèse

Dans ce chapitre, sous une modification du coût quadratique (coût basé sur les données), nous avons donné, grâce à une nouvelle identité de type Stein-Haff, des résultats de dominance d'estimateurs de la forme générale $a(S + SS^+ G(Z, S))$ par rapport aux estimateurs habituels aS , dans une approche qui unifie les cas où S est inversible et où S est non inversible. Ces résultats sont donnés au travers de l'estimateur optimal associé au coût basé sur les données, qui, contrairement à celui associé au coût quadratique, change en fonction des dimensions p et n (selon l'inversibilité de S). Notre théorie générale est appliquée aux estimateurs orthogonalement invariant dans la classe de type Berger (3.2.18). Nous avons donné aussi un exemple d'estimateur orthogonalement invariant de type Haff et un cas particulier de celui-ci, *i.e.*, l'estimateur de Konno. Notons qu'une démonstration spécifique est donnée pour ce dernier où nous avons fourni des résultats de dominance pour toute la classe des distributions à symétrie elliptique.

Il convient de noter que l'utilisation du coût basé sur les données (4.1.1) facilite les techniques de calcul. Quant au problème d'estimation amélioré de la matrice d'échelle Σ , les résultats obtenus sont plus intéressants que ceux du [chapitre 3](#). En effet, nous proposons une plus large classe d'estimateurs améliorés pour un contexte distributionnel plus grand. Afin d'évaluer la qualité de ses estimateurs, des simulations seront effectuées dans le prochain chapitre.

Sommaire

5.1	La théorie moderne du portefeuille	56
5.2	Simulation des lois à symétrie elliptique	59
5.2.1	Représentations graphique de la loi normale multidimensionnelle	60
5.3	Étude numérique	62
5.3.1	Comparaisons des PRIAL	62
5.3.2	Effet du facteur de correction sur le PRIAL	64

5.1 La théorie moderne du portefeuille

En finance, un portefeuille est un ensemble d’actifs détenus par un investisseur. Ces actifs peuvent provenir de différentes classes : actions, obligations, produits dérivés, matières premières, fonds, cash, *etc.* Un investisseur, rationnel, cherche à maximiser le rendement espéré du portefeuille pour un niveau de risque donné ou bien de minimiser le risque pour un niveau de rendement espéré. Dans la théorie moderne du portefeuille, développée par Markowitz [Mar52], le rendement d’un actif R_i est une variable aléatoire possédant, une espérance $E(R_i) = \theta_i$ et une variance $Var(R_i) = \sigma_i^2$. Un portefeuille est une combinaison linéaire des p –actifs qui le composent pondérés par leur poids w_i dans le portefeuille, *i.e.*,

$$R_{po} = \sum_{i=1}^p w_i R_i .$$

Par conséquent, le rendement d’un portefeuille est également une variable aléatoire dont l’espérance est donnée par

$$E(R_{po}) = \sum_{i=1}^p w_i E(R_i) = \mathbf{w}^T \boldsymbol{\theta} , \tag{5.1.1}$$

où $\mathbf{w} = (w_1, \dots, w_i, \dots, w_p)^T$ avec $\sum_{i=1}^p w_i = 1$ et $\boldsymbol{\theta} = (\theta_1, \dots, \theta_i, \dots, \theta_p)^T$. Sa variance qui dépend des corrélations ρ_{ij} entre les rendements, est de la forme

$$\text{Var}(R_{po}) = \sum_{i=1}^p \sum_{j=1}^p w_i w_j \sigma_i \sigma_j \rho_{ij}, = \mathbf{w}^T \Sigma \mathbf{w}, \quad (5.1.2)$$

où Σ est la matrice de covariance des p -actifs du portefeuille.

Le risque de rendement du portefeuille est quantifié par la volatilité de ce dernier, qui peut être définie par l'écart-type $\sqrt{\text{Var}[R_{po}]}$. Un portefeuille a une forte volatilité lorsque les variations de son rendement dans le temps sont très dispersées autour de sa moyenne, tandis qu'un portefeuille a une faible volatilité si ces variations sont regroupées autour de sa moyenne. Un investisseur peut vouloir chercher à réduire le risque (volatilité) de son portefeuille. Ceci peut être fait en choisissant des actifs qui ne soient pas ou peu positivement corrélés, donc en diversifiant ses investissements. Cela permet d'obtenir la même espérance de rendement en diminuant la volatilité du portefeuille. Comme dans la notion de corrélation intervient la covariance, puisque

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j},$$

où σ_{ij} est la covariance entre les rendements R_i et R_j . Dans ce cas, une bonne estimation de la matrice de covariance Σ permet à l'investisseur de sélectionner les actifs les moins corrélés afin de réduire la volatilité (le risque) de son portefeuille. Comme mentionné par Ledoit et Wolf [LW03], lorsque le nombre d'actifs considérés est important, en particulier par rapport au nombre d'observations de leurs rendements, les estimateurs usuels à l'instar de la matrice de covariance empirique ne sont pas de bons estimateurs et peuvent être remplacés par des estimateurs de la forme (3.2.1).

Une fois ces actifs sélectionnés, l'investisseur est ramené à tester différentes combinaisons d'actifs avec des pondérations w_i diverses pour calculer le risque et la rentabilité espérée des différents portefeuilles. Il est alors possible d'obtenir un graphique (*cf.* figure 5.1) représentant le risque et le rendement de chaque portefeuille. Ainsi on peut déterminer une frontière d'efficacité sur laquelle se situent les portefeuilles offrant, une rentabilité maximale pour un certain niveau de risque ou un risque minimal pour un certain niveau de rendement. Une façon de calculer la frontière d'efficacité est de fixer le rendement (5.1.1) à une valeur μ_{po} , puis la faire varier et minimiser la variance du portefeuille dans (5.1.2) pour chaque valeur. Ceci se ramène à un problème d'optimisation classique de la forme

$$\min_{\mathbf{w}} \left\{ \frac{1}{2} \mathbf{w}^T \Sigma \mathbf{w} \mid \mathbf{w}^T \mathbf{1} = 1, \mathbf{w}^T \boldsymbol{\theta} = \mu_{po} \right\}, \quad (5.1.3)$$

où $\mathbf{1}$ est le vecteur colonne composé de 1. D'après la méthode de multiplicateur de Lagrange, le lagrangien de ce problème d'optimisation est donné par

$$L(\mathbf{w}, \lambda_1, \lambda_2) = \frac{1}{2} \mathbf{w}^T \Sigma \mathbf{w} - \lambda_1 (\mathbf{w}^T \mathbf{1} - 1) - \lambda_2 (\mathbf{w}^T \boldsymbol{\theta} - \mu_{po})$$

où λ_1 et λ_2 sont les multiplicateurs de Lagrange. En fixant les trois dérivées partielles de $L(\mathbf{w}, \lambda_1, \lambda_2)$ à zéro, nous avons

$$\frac{\partial L(\mathbf{w}, \lambda_1, \lambda_2)}{\partial \mathbf{w}} = \Sigma \mathbf{w} - \lambda_1 \mathbf{1} - \lambda_2 \boldsymbol{\theta} = 0, \quad (5.1.4)$$

$$\frac{\partial L(\mathbf{w}, \lambda_1, \lambda_2)}{\partial \lambda_1} = 1 - \mathbf{w}^T \mathbf{1} = 0 \quad (5.1.5)$$

et

$$\frac{\partial L(\mathbf{w}, \lambda_1, \lambda_2)}{\partial \lambda_2} = \mu_{po} - \mathbf{w}^T \boldsymbol{\theta} = 0. \quad (5.1.6)$$

D'après (5.1.4), nous avons une solution optimale la forme

$$\mathbf{w}_o = \Sigma^{-1} (\lambda_1 \mathbf{1} + \lambda_2 \boldsymbol{\theta}). \quad (5.1.7)$$

Afin de déterminer λ_1 et λ_2 , nous utilisons (5.1.5) et (5.1.6), ce qui donne

$$\begin{aligned} \mathbf{w}_o^T \mathbf{1} &= \lambda_1^T \Sigma^{-1} \mathbf{1} - \lambda_2 \boldsymbol{\theta}^T \Sigma^{-1} \mathbf{1} = 1, \\ \mathbf{w}_o^T \boldsymbol{\theta} &= \lambda_1 \mathbf{1}^T \Sigma^{-1} \boldsymbol{\theta} + \lambda_2 \boldsymbol{\theta}^T \Sigma^{-1} \boldsymbol{\theta} = \mu_{po}. \end{aligned}$$

Ainsi, la solution de ce système est donnée par

$$\lambda_1 = \frac{c - b\mu_{po}}{ac - b^2} \text{ et } \lambda_2 = \frac{a\mu_{po} - b}{ac - b^2}, \quad (5.1.8)$$

où $a = \mathbf{1}^T \Sigma^{-1} \mathbf{1}$, $b = \boldsymbol{\theta}^T \Sigma^{-1} \mathbf{1}$ et $c = \boldsymbol{\theta}^T \Sigma^{-1} \boldsymbol{\theta}$. Alors, la solution optimale du problème d'optimisation dans (5.1.3) est obtenue en remplaçant (5.1.8) dans (5.1.7), *i.e.*

$$\mathbf{w}_o = \Sigma^{-1} \left(\frac{c - b\mu_{po}}{ac - b^2} \mathbf{1} + \frac{a\mu_{po} - b}{ac - b^2} \boldsymbol{\theta} \right),$$

et la variance minimale est donnée par

$$\frac{a^2 \mu_{po} - 2b\mu_{po} + c}{ac - b^2}.$$

Par conséquent, l'ensemble des portefeuilles de variance minimale, qui forment la frontière efficiente, est représenté par la courbe parabolique sur le plan $(Var(R_{po}), \mu_{po})$ dans la figure 5.1, qui est obtenue en faisant varier le rendement souhaité μ_{po} .

Il convient de noter que dans sa théorie, Markowitz ne fait aucune hypothèse de distribution sur les rendements des actifs. Toutefois, un problème fondamental provient du fait que les applications, comme le MEDAF¹, utilisant cette théorie reposent sur l'hypothèse de normalité des rendements (*cf.* Mandelbrot [Man63]). Or, il s'avère que la loi normale, qui possède des queues légères, sous-estime très fortement les événements improbables (rares) comme les crises ou les krachs boursiers puisque ses derniers se répartissent sur les deux extrémités de cette distribution. Dans ce cas, l'utilisation des distributions à symétrie elliptiques avec des queues plus épaisses que celles de la loi normale, peut être une alternative intéressante (*cf.* Hamada et Valdez [HV08]).

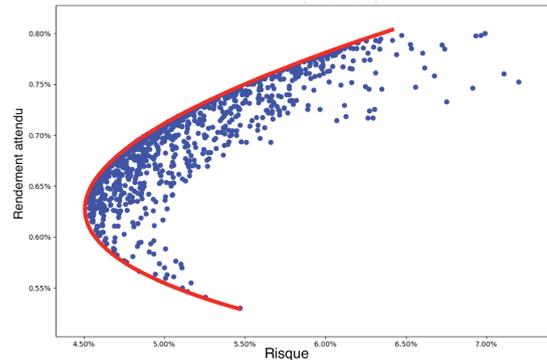


FIGURE 5.1 – Frontière efficiente.

1. Modèle d'Évaluation des Actifs Financiers. Ce modèle fournit une estimation du taux de rentabilité attendu pour un actif financier en fonction de son risque (*cf.* Sharpe [S.64])

5.2 Simulation des lois à symétrie elliptique

Afin de simuler une distribution à symétrie elliptique Y de moyenne θ et de matrice d'échelle Σ , nous pouvons, tout d'abord, simuler une distribution à symétrie sphérique X autour de 0, ensuite, par la transformation affine

$$Y = \theta + \Sigma^{1/2} X$$

on construit $Y \sim ES(\theta, \Sigma)$. Nous allons voir maintenant comment simuler $X \sim SS(0)$.

Bien qu'il existe plusieurs manières différentes pour simuler des vecteurs aléatoires de distribution à symétrie sphérique (cf. Devroye [Dev86]), nous avons choisi la méthode qui repose sur la représentation stochastique de X en raison de sa simplicité à être programmée. Rappelons (cf. définition 2.2.4) qu'un vecteur aléatoire $X \in \mathbb{R}^p$ a une distribution à symétrie sphérique si et seulement si il est de la forme

$$X \stackrel{d}{=} RU, \quad R = \|X\| \geq 0, \quad U = \frac{X}{\|X\|} \quad (5.2.1)$$

où R est le rayon de X et où sa direction U suit une distribution uniforme sur la sphère unité

$$\mathcal{S}_1 = \{x \in \mathbb{R}^p; \|x\|^2 = 1\},$$

les variables U et R étant indépendantes. Ainsi, si nous pouvons générer U , toute distribution dans \mathbb{R}^+ peut être utilisée pour générer le rayon R et ainsi générer un vecteur aléatoire sphérique X . De ce fait, notre intérêt se tourne maintenant vers la simulation d'un vecteur aléatoire uniforme sur la sphère \mathcal{S}_1 . Ceci peut être fait si un générateur efficace de variables aléatoires gaussiennes est disponible. En effet, la représentation stochastique (5.2.1) est valable pour toute distribution à symétrie sphérique, en particulier, pour la distribution gaussienne. Par conséquent si $Y \sim \mathcal{N}_p(0, 1)$, alors $U = Y/\|Y\|$ suit une distribution uniforme sur la sphère unité \mathcal{S}_1 (voir la figure 5.2). Nous expliquons dans l'algorithme 1 comment simuler des réalisations de U .

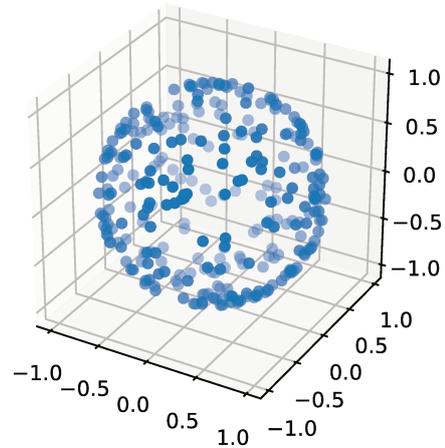


FIGURE 5.2 – Loi uniforme sur la sphère unité.

Algorithme 1:

Requérir : $p \in \mathbb{N}^*$.

1. Générer p variables aléatoires gaussienne $Y_i \sim \mathcal{N}_1(0, 1)$.
 2. Calculer $S = \sum_{i=1}^p Y_i^2$.
 3. Calculer $U = (Y_1, \dots, Y_i, \dots, Y_p)/S^{1/2}$.
-

Nous donnons dans le tableau 5.1 quelques exemples de vecteurs aléatoires à symétrie sphérique X pouvant être générés grâce à leur représentation stochastique, où nous spécifions à chaque fois la distribution du carré du rayon R^2 (cf. Fang, Kotz et Ng [FKN90]). Pour ces distributions, les lois de R^2 , correspondantes, sont préprogrammées dans la plupart des logiciels. Pour d'autres lois, la méthode d'acceptation-rejet peut être utilisée.

Loi	Densité de $R^2 = \ X\ ^2$
Normal $\mathcal{N}_p(\theta, \sigma^2)$	$R^2 \sim \mathcal{X}^2(p)$
Student $t_p(\nu, \sigma^2)$	$R^2/p \sim Fisher(p, \nu)$
Kotz $\mathcal{K}_p(N, r, \sigma^2)$	$R^2 \sim gamma(N + 1 + \frac{p}{2}, \frac{r}{\sigma^2})$
Puis. Expo $\mathcal{PE}_p(b, \sigma^2)$	$R^{2b} \sim gamma(\frac{p}{2b}, 2\sigma^{2b})$

TABLEAU 5.1 – Distribution du carre du rayon pour quelques lois sphériques

Maintenant, nous expliquons dans l’[algorithme 2](#) comment simuler une distribution à symétrie elliptique.

Algorithm 2:

Requérir : $p \in \mathbb{N}^*$, $\Sigma \in \mathbb{R}^{p \times p}$, $\theta \in \mathbb{R}^p$.

1. Générer le vecteur uniforme U .
 2. Générer le rayon R .
 3. Calculer $X = RU$.
 4. Calculer $Y = \theta + \Sigma^{1/2}X$.
-

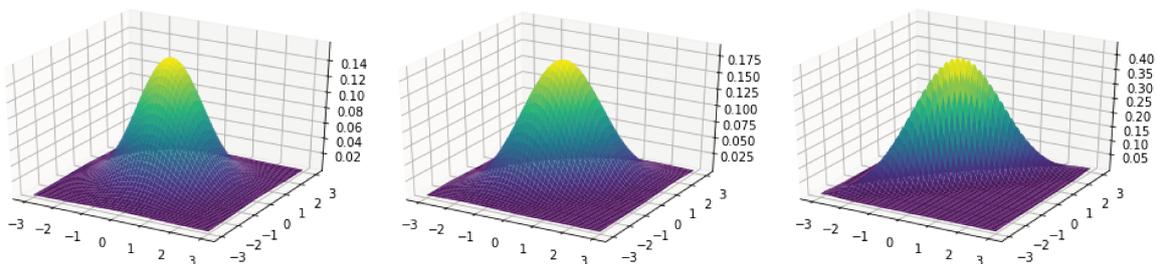
5.2.1 Représentations graphique de la loi normale multidimensionnelle

La loi normale multidimensionnelle est caractérisée par le vecteur moyenne $\theta \in \mathbb{R}^p$ et la matrice d’échelle (qui correspond à la matrice de covariance) $\Sigma \in \mathbb{R}^{p \times p}$. Si sa densité par rapport à la mesure de Lebesgue dans \mathbb{R}^p existe (cf. [théorème 2.3.1](#)), alors elle est nécessairement de la forme

$$f(y) = \frac{1}{(2\pi)^{p/2}} |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} (y - \theta) \Sigma^{-1} (y - \theta)^T \right\}$$

Nous donnons quelques aperçus de la densité joint de $Y \in \mathbb{R}^2$, pour différentes valeurs de Σ où nous supposons, sans perdre de généralité, que la moyenne $\theta = 0$.

Dans la [figure 5.3](#) nous donnons différentes représentations de la densité lorsque Σ est proportionnel à l’identité. La figure la plus à gauche montre une gaussienne avec une matrice de covariance $\Sigma = 0,5 I_2$. Celle du milieu est donnée pour $\Sigma = I_2$ et celle de droite pour $\Sigma = 2 I_2$. Lorsque la variance augmente, la densité est plus étalée et plus elle est petite, plus la densité devient comprimée.

FIGURE 5.3 – Densités de la loi normale pour Σ proportionnel à I_2 .

Dans la [figure 5.4](#), nous donnons la représentation des contours lorsque la covariance augmente. Les figures, de gauche à droite, correspondent respectivement aux valeurs de la matrice de covariance suivantes :

$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \Sigma = \begin{bmatrix} 1 & 0,5 \\ 0,5 & 1 \end{bmatrix} \text{ et } \Sigma = \begin{bmatrix} 1 & 0,8 \\ 0,8 & 1 \end{bmatrix}.$$

Nous remarquons que la densité devient plus comprimée (vers la ligne 45°) lorsque la covariance augmente.

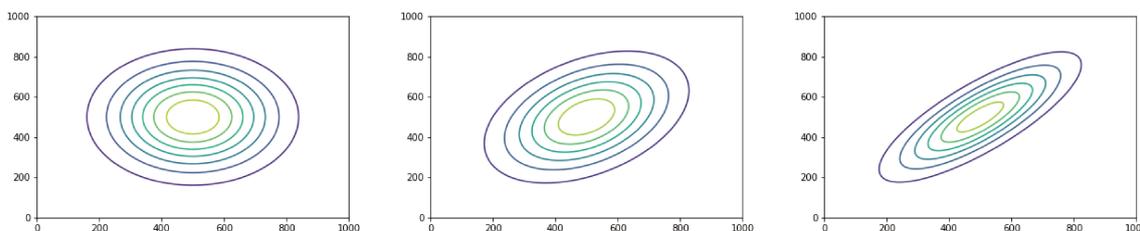


FIGURE 5.4 – Contours de la loi normale avec des covariances positives.

Enfin, nous donnons dans [figure 5.5](#) la représentation des contours lorsque la covariance (négative) diminue. Les figures, de gauche à droite, correspondent respectivement aux valeurs de la matrice de covariance suivantes :

$$\Sigma = \begin{bmatrix} 1 & -0,5 \\ -0,5 & 1 \end{bmatrix}, \Sigma = \begin{bmatrix} 1 & -0,8 \\ -0,8 & 1 \end{bmatrix} \text{ et } \Sigma = \begin{bmatrix} 2 & -0,8 \\ -0,8 & 1 \end{bmatrix}.$$

Nous remarquons dans la figure de gauche et dans celle du milieu que la densité se comprime comme dans la [figure 5.4](#) mais dans le sens opposé. On remarque dans la figure de droite que les contours forment des ellipses lorsque nous modifions les variances.

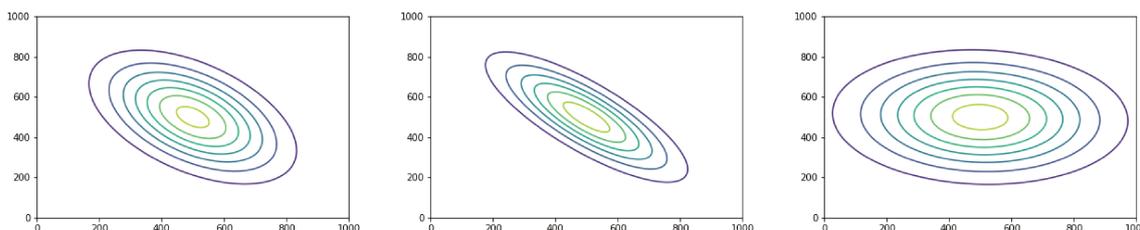


FIGURE 5.5 – Contours de la loi normale avec des covariances négatives.

5.3 Étude numérique

Dans cette section, nous effectuons des simulations de type Monte Carlo afin d'étudier numériquement la performance de l'estimateur alternatif de type Konno, dans (1.3.6), de la forme

$$\hat{\Sigma}^K = a_o (S + \nu t S S^+), \quad \text{avec } t \geq 0 \text{ et } \nu = \frac{1}{\text{tr}(S^+)}, \quad (5.3.1)$$

par rapport à l'estimateur optimal $\hat{\Sigma}_{a_o} = a_o S$, sous chacun des coûts (3.1.3) et (4.1.1). Nous nous intéressons au cadre dimensionnel large ($p > n$) dans le contexte distributionnel gaussien ($c = b = K^* = K^{**} = 1$).

Comme les constantes a_o et t changent en fonction du coût choisi, alors, l'estimateur de Konno change aussi. En effet, au vu de la proposition 3.3.2 et la proposition 4.3.2, au coût quadratique correspond l'estimateur

$$\hat{\Sigma}_Q^K = \frac{1}{p+n+1} \left(S + \nu \frac{2(n-1)}{(p-n+1)(p-n+3)} S S^+ \right)$$

au coût basé sur les données correspond l'estimateur

$$\hat{\Sigma}_S^K = \frac{1}{p} \left(S + \nu \frac{2(n-1)}{(p-n+1)} S S^+ \right).$$

5.3.1 Comparaisons des PRIAL

Pour deux structures différentes de la matrice d'échelle Σ , soit l'identité ($\Sigma = I_p$) et la structure autorégressive où $(\Sigma)_{ij} = 0.9^{|i-j|}$, nous générons 1000 répétitions d'un n -échantillon d'une distribution gaussienne p -dimensionnelle de moyenne $\theta = 0$. Ensuite, nous évaluons, dans les tableaux 5.2, 5.4, 5.3 et 5.5, les pourcentages d'amélioration de l'estimateur $\hat{\Sigma}_Q^K$ par rapport à l'optimal sous le coût quadratique (cf. proposition 3.2.3), soit $\hat{\Sigma}_{a_o} = S/(p+n+1)$ et de $\hat{\Sigma}_S^K$ par rapport à l'optimal sous le coût basé sur les données (cf. proposition 4.2.2), soit $\hat{\Sigma}_{a_o} = S/p$. Le pourcentage d'amélioration (PRIAL²) est défini par :

$$\text{PRIAL}(\hat{\Sigma}^K) = \frac{\text{coût moyen de } \hat{\Sigma}_{a_o} - \text{coût moyen de } \hat{\Sigma}^K}{\text{coût moyen de } \hat{\Sigma}_{a_o}} \times 100,$$

pour $(\hat{\Sigma}^K, \hat{\Sigma}_{a_o}) = (\hat{\Sigma}_Q^K, S/(p+n+1))$ ou $(\hat{\Sigma}_S^K, S/p)$.

Quand le rapport p/n est fixé à 2, nous disposons des tableaux 5.2 et 5.3. Nous constatons que, lorsque nous augmentons p de 10 à 100, les pourcentages d'amélioration diminuent.

Quand nous diminuons p/n pour p fixe, nous avons les tableaux 5.3 et 5.5. Nous voyons que, lorsque nous diminuons p/n de 5 à 5/4 pour $p = 20$ et $p = 100$, les deux pourcentages d'amélioration augmentent. Les meilleurs résultats sont atteints pour $p/n = 5/4$.

En outre, nous remarquons dans les tableaux 5.2, 5.3, 5.4 et 5.5 que, pour les deux structures de Σ , les pourcentages d'amélioration sous le coût basé sur les données (4.1.1) sont meilleurs que sous le coût quadratique (3.1.3). On remarque aussi qu'on gagne plus en amélioration lorsque nous considérons une structure autorégressive pour la matrice d'échelle.

2. Percentage Relative Improvement in Average Loss.

p	n	$S/(p+n+1)$	$\hat{\Sigma}_Q^K$	PRIAL($\hat{\Sigma}_Q^K$)%	S/p	$\hat{\Sigma}_S^K$	PRIAL($\hat{\Sigma}_S^K$)%
10	5	6.890	6.775	1.66	5.038	4.407	12.5
20	10	13.569	13.359	1.55	9.993	9.217	7.75
40	20	26.884	26.586	1.10	19.998	19.129	4.34
60	30	40.203	39.864	0.84	29.991	29.081	3.03
80	40	53.538	53.177	0.67	40.018	39.087	2.32
100	50	66.912	66.536	0.56	49.965	49.022	1.88

TABLEAU 5.2 – Résultat de 1000 simulations de type Monte Carlo pour $p/n = 2$,
 $\Sigma = I_p$.

p	n	$S/(p+n+1)$	$\hat{\Sigma}_Q^K$	PRIAL($\hat{\Sigma}_Q^K$)%	S/p	$\hat{\Sigma}_S^K$	PRIAL($\hat{\Sigma}_S^K$)%
10	5	6.881	6.644	3.442	5.023	3.199	36.29
20	10	13.553	12.960	4.37	10.002	7.868	21.33
40	20	26.902	26.104	2.96	20.015	17.730	11.41
60	30	40.218	39.328	2.21	29.987	27.620	7.89
80	40	53.538	52.600	1.75	40.018	37.620	5.99
100	50	66.896	65.930	1.44	49.950	47.533	4.83

TABLEAU 5.3 – Résultat de 1000 simulations de type Monte Carlo pour $p/n = 2$,
 $(\Sigma)_{ij} = 0.9^{|i-j|}$.

p	n	$S/(p+n+1)$	$\hat{\Sigma}_Q^K$	PRIAL($\hat{\Sigma}_Q^K$)%	S/p	$\hat{\Sigma}_S^K$	PRIAL($\hat{\Sigma}_S^K$)%
20	4	16.785	16.738	0.27	15.990	15.720	1.69
20	8	14.461	14.299	1.11	11.979	11.367	5.10
20	12	12.715	12.472	1.90	7.985	7.035	11.89
20	16	11.346	11.134	1.86	3.953	2.688	31.99
100	20	83.472	83.376	0.11	80.010	79.642	0.45
100	40	71.637	71.356	0.39	60.015	59.261	1.25
100	60	62.766	62.303	0.73	39.999	38.868	2.82
100	80	55.800	55.240	1.004	20.016	18.529	7.43

TABLEAU 5.4 – Résultats pour deux valeurs de p fixées, la valeur de n variant,
 $\Sigma = I_p$.

p	n	$S/(p+n+1)$	$\hat{\Sigma}_Q^K$	PRIAL($\hat{\Sigma}_Q^K$)%	S/p	$\hat{\Sigma}_S^K$	PRIAL($\hat{\Sigma}_S^K$)%
20	4	16.785	16.632	0.91	15.992	13.592	15.00
20	8	14.463	14.092	2.56	11.98	9.760	18.56
20	12	12.722	12.079	5.05	7.995	5.951	25.56
20	16	11.337	10.865	4.17	3.983	2.110	47.034
100	20	83.460	83.119	0.40	80.012	77.294	3.39
100	40	71.622	70.733	1.24	59.97	57.45	4.19
100	60	62.747	61.767	1.56	39.996	37.690	5.76
100	80	55.805	54.950	1.531	19.956	17.876	10.42

TABLEAU 5.5 – Résultats pour deux valeurs de p fixées, la valeur de n variant,
 $(\Sigma)_{ij} = 0.9^{|i-j|}$.

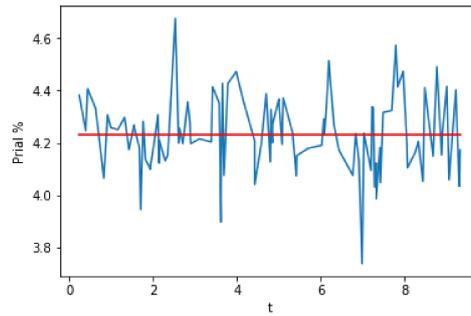


FIGURE 5.6 – Effet de t sur le PRIAL lorsque $\Sigma = I_p$ pour $(p, n) = (120, 100)$

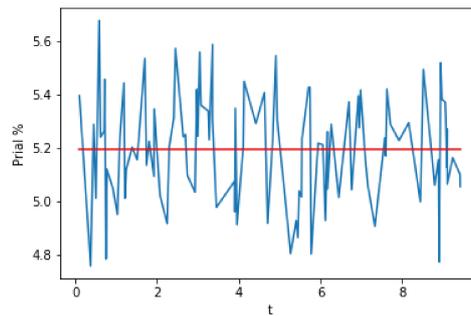


FIGURE 5.7 – Effet de t sur le PRIAL lorsque $(\Sigma)_{ij} = 0.9^{|i-j|}$ pour $(p, n) = (120, 100)$

5.3.2 Effet du facteur de correction sur le PRIAL

Intéressons nous maintenant à l'effet de la constante t , qui intervient dans l'estimateur de Konno dans (5.3.1), sur le pourcentage d'amélioration. Comme l'amélioration est meilleure sous le coût basé sur les données que sous le coût quadratique, et comme les résultats de simulation ont montré un comportement similaire du PRIAL sous ces deux coûts, nous ne nous intéressons alors qu'à l'effet de t sur le pourcentage d'amélioration de Σ_S^K par rapport à S/p .

Nous considérons deux structures différentes pour la matrice l'échelle, soit l'identité et la structure autorégressive. Pour $(p, n) = (120, 100)$, nous faisons varier t dans $]0, 2(n-1)/(p-n+1)]$ et, pour chaque valeur, nous calculons le pourcentage d'amélioration. Nous remarquons dans les figures 5.6 et 5.7 que ce dernier est constant dans l'intervalle $]0, 2(n-1)/(p-n+1)]$.

CONCLUSION

Dans cette thèse nous avons considéré, d'un point de vue de la théorie de la décision, le problème d'estimation de la matrice d'échelle Σ du modèle additif $Y_{p \times m} = M + \mathcal{E}$, où le bruit \mathcal{E} suit une distribution à symétrie elliptique de matrice de covariance proportionnelle à $I_m \otimes \Sigma$ et M une matrice de rang $q < p$. Nous avons vu qu'il est judicieux d'aborder ce problème d'estimation sous la représentation canonique de ce modèle dans (3.1.1), où la matrice d'observation Y est décomposée en deux matrices, à savoir, Z qui résume l'information sur M , et une matrice U qui résume l'information suffisante pour l'estimation de Σ . En effet, au lieu d'estimer Σ en exploitant toute l'information contenue dans la matrice d'observation Y , nous pouvons l'estimer en se basant sur une matrice de plus petite dimension $U_{n \times p}$ où $n = m - q$. Comme les estimateurs usuels de la forme

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

où $S = U^T U$ et a est une constante positive, sont inadmissibles lorsque le nombre de variables p et lorsque le rapport p/n sont grands, nous avons proposé des estimateurs améliorés de la forme générale

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

Dans un premier temps, la performance de ces estimateurs est évalué au travers de la fonction de coût quadratique

$$L(\Sigma, \hat{\Sigma}) = \text{tr}(\Sigma^{-1} \hat{\Sigma} \Sigma^{-1} \hat{\Sigma}) - 2 \text{tr}(\Sigma^{-1} \hat{\Sigma}) + p.$$

Pour montrer l'amélioration des ces estimateurs par rapport au estimateurs usuels, nous avons donné des conditions sur le facteur de correction $SS^+ G(Z, S)$, dans le [théorème 3.2.1](#), afin que la différence des risque entre $\hat{\Sigma}_{a,G}$ et $\hat{\Sigma}_a = a S$, *i.e.*

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

soit négative. Comme le terme intégrand de $\Delta(G)$ dépend du paramètre inconnu Σ^{-1} , nous avons développé une nouvelle identité de type Stein-Haff pour pallier à cette difficulté. À contrario des identités de ce type existantes dans la littérature, celle-ci traite à la fois le cas où S est inversible et le cas où S n'est pas inversible. De plus, elle est donnée dans le cadre général des distributions à symétrie elliptique. L'application de cette identité nous conduit à réduire le contexte de

distribution à la grande sous-classe de densité de type Berger dans (3.2.18) qui contient la loi gaussienne. Grâce au traitement de la structure en valeurs propre de S dans le lemme B.2, nous avons réussi à appliquer notre théorie à la classe des estimateurs orthogonalement invariants de la forme (1.3.2). Ceci nous a permis d'étendre les résultats de Haff [Haf80] et de Konno [Kon09], donnés dans le contexte gaussien, au contexte elliptique et dans une approche unifiée en terme d'inversibilité de S . Comme mentionné par Haff, ce coût quadratique est difficile à manipuler car la présence du paramètre inconnu Σ^{-1} deux fois dans le terme intégrant de la différence des risques, nécessite une double application de l'identité de type Stein-Haff. En effet, ceci implique des calculs long et inabordable et nous impose la symétrie du facteur de correction $SS^+G(Z, S)$. Afin de pallier à cette difficulté, nous avons modifier ce coût en introduisant S^+ l'inverse de Moore-Penrose de l'observation S , ce qui nous a conduit à considérer notre coût basé sur les données

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

Ainsi, comme dans ce coût le paramètre inconnu Σ^{-1} n'est présent qu'une fois dans la différence des risques

$$\Delta(G) = a^2 E_{\theta, \Sigma} [\text{tr}(\Sigma^{-1} SS^+ G \{I_p + S^+ G + SS^+\})] - 2a E_{\theta, \Sigma} [\text{tr}(S^+ G)],$$

l'identité de type Stein-Haff n'est appliquée qu'une seule fois. Par conséquent, sous ce coût, les calculs sont plus abordables et les conditions d'amélioration, données dans le théorème 4.2.1, sont moins fortes et le facteur de correction $SS^+G(Z, S)$ n'est pas nécessairement symétrique (d'où une plus grande classe d'estimateurs améliorés). Ce coût, nous a permis d'élaborer une démonstration particulière pour l'estimateur de type Konno dans (1.3.6) où des résultats de dominance sont fournis pour toute la classe des distributions à symétrie elliptique (sans se restreindre à la sous-classe de type Berger). De plus, pour cet estimateur, les simulations donnent de meilleurs résultats sous le coût basé sur les données que sous le coût quadratique.

Sous ces deux coûts, l'application de l'identité de Stein-Haff nous a amenés à comparer différentes espérances sous $E_{\theta, \Sigma}$, $E_{\theta, \Sigma}^*$ et $E_{\theta, \Sigma}^{**}$. À cette fin, pour disposer d'expressions homogènes des différences des risques, nous nous sommes placés dans la sous-classe de distributions type Berger dans (3.2.18). Bien que cette sous-classe contienne de nombreuses distributions, à l'instar de la loi gaussienne et la loi de Student tronquée, il serait intéressant dans l'avenir de trouver une approche différente pour comparer ces espérances et, par conséquent, de considérer une plus grande classe de loi. Notons que, sous le coût basé sur les données, nous avons donné une telle approche différente pour l'estimateur de Konno. En effet nous étions amenés à comparer entre $E_{\theta, \Sigma}$ et $E_{\theta, \Sigma}^*$ et le fait de choisir un terme correcteur $G(Z, S)$ de telle sorte que $E_{\theta, \Sigma}[\text{tr}(S^+ G)]$ soit constant, nous permis de considérer toute la classe des distributions à symétrie elliptique.

Comme il a été signalé par Tsukuma et Kubokawa [TK16] dans le cadre gaussien, des estimateurs $\hat{\Sigma}$, comme celui de Stein (1.3.3) ou celui de Konno (1.3.6), peuvent être améliorés par des estimateurs dont le facteur de correction $SS^+G(Z, S)$ contient aussi l'information apporté par Z sur la matrice des paramètres M . Il serait intéressant de considérer dans le cas elliptique de tels estimateurs.

Dans le cadre gaussien, une attention particulière a été récemment accordée au cas où la matrice d'échelle Σ n'est pas inversible. Chételat et Wells [CW16] donnent des résultats d'amélioration de l'estimateur de type Konno sous le coût quadratique modifié $\text{tr}(\hat{\Sigma}\Sigma^+ - I_p)^2$, où Σ^+ est l'inverse de Moore-Penrose de Σ qui est de rang $r < n \wedge p$. Une extension de ces travaux au cadre elliptique semble être possible si nous développons une nouvelle identité de type Stein-Haff.

Appendices

DÉFINITIONS ET CALCULS MATRICIELS

Dans cette section, nous donnons les définitions des notions utilisées dans ce manuscrit. Nous donnons aussi la preuve de certains résultats connus relatifs à l'opérateur Haff $\mathcal{D}_s(\cdot)$ défini dans (2.7.1) et à la trace de S et S^+ .

Définition A.1 (L'inverse de Moore-Penrose)

Soit $A \in \mathbb{R}^{n \times p}$. L'inverse de Moore-Penrose de A , noté A^+ , est défini comme une matrice $A^+ \in \mathbb{R}^{n \times p}$ qui satisfait les quatre critères suivants, connus sous le nom de conditions de Moore-Penrose :

$$\begin{aligned} AA^+A &= A. \\ A^+AA^+ &= A^+. \\ (AA^+)^T &= AA^+. \\ (A^+A)^T &= A^+A. \end{aligned}$$

Dans le cas particulier où $A = S = UU^T$ est symétrique, nous avons les propriétés suivantes :

$$\begin{aligned} SS^+S &= SSS^+ = S^+SS = S, \\ S^+SS^+ &= S^+S^+S = SS^+S^+ = S^+, \\ SS^+ &= S^+S. \end{aligned}$$

Ainsi

$$\begin{aligned} S(I_p - SS^+) &= 0. \\ S^+(I_p - SS^+) &= 0. \\ SS^+(I_p - SS^+) &= 0. \end{aligned}$$

Par conséquent puisque $S = UU^T$ nous avons aussi

$$\begin{aligned} USS^+ &= UU^+(U^T)^+U^T U = (UU^+)(UU^+)^T U = U, \\ SS^+U^T &= U^T UU^+(U^T)^+U^T = U^T UU^+(UU^+)^T = U^T \end{aligned}$$

et donc

$$\begin{aligned} U(I_p - SS^+) &= 0, \\ (I_p - SS^+)U^T &= 0. \end{aligned}$$

De plus, en utilisant la décomposition en valeurs propres de $S = H_1 L H_1^T$, puisque $SS^+ = H_1 H_1^T$, est un projecteur nous avons

$$\begin{aligned} H_1 H_1^T H_1 H_1^T &= H_1 H_1^T \\ H_1^T (I_p - H_1 H_1^T) &= 0 \\ (I_p - H_1 H_1^T) H_1 &= 0 \end{aligned}$$

Définition A.2 (Estimateur)

Soit (U_1, \dots, U_n) un n -échantillon d'un vecteur aléatoire $U \in \mathbb{R}^p$ dont la loi dépend d'un paramètre (matriciel) A . Un estimateur \hat{A}_n du paramètre A est une fonction qui fait correspondre à chaque réalisation possible (Y_1, \dots, Y_n) de l'échantillon à n -éléments la valeur \hat{A}_n que l'on nomme estimé ou estimation i.e.

$$\hat{A}_n = f(u_1, \dots, u_n).$$

Un estimateur ne doit évidemment jamais dépendre de A , il ne dépend que des observations empiriques (i.e. de la réalisation de l'échantillon).

Définition A.3 (Estimateur sans biais)

Le biais $B(\hat{A}_n)$ d'un estimateur \hat{A}_n de A est défini par

$$B(\hat{A}_n) = E_A(\hat{A}_n) - A$$

où $E_A(\cdot)$ désigne l'espérance par rapport à la loi de Y . Lorsque $B(\hat{A}_n) = 0$, l'estimateur \hat{A}_n est appelé estimateur sans biais ou non biaisé de A .

Définition A.4 (Estimateur consistant)

Un estimateur \hat{A}_n de A est dit consistant s'il converge en probabilité vers A lorsque n tend vers l'infinie i.e.

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} P\left(\|\hat{A}_n - A\| \geq \varepsilon\right) \rightarrow 0$$

où $P(\cdot)$ désigne la probabilité et $\|\cdot\|$ désigne une norme matricielle quelconque.

Proposition A.1

Soit A , B et C des fonctions $p \times p$ matricielles de S . En supposant que toutes les dérivées et produits partiels existent au besoin, nous avons

$$\mathcal{D}_s \{A B\} = \mathcal{D}_s \{A\} B + (A^T \mathcal{D}_s)^T \{B\} \tag{A.1}$$

$$\text{tr}[C (A^T \mathcal{D}_s)^T \{B\}] = \text{tr}[A^T (C \mathcal{D}_s)^T \{B\}^T] \tag{A.2}$$

Preuve

La première identité a été donnée par Haff [Haf80], nous ne donnons donc que la preuve de (A.2). On a

$$\begin{aligned} \text{tr}[C (A^T \mathcal{D}_s)^T \{B\}] &= \sum_{i,j,k}^p C_{ij} (A^T \mathcal{D}_s)_{jk}^T B_{ki} \\ &= \sum_{i,j,k}^p C_{ij} (A^T \mathcal{D}_s)_{kj} B_{ik}^T. \end{aligned}$$

Alors, par symétrie de l'opérateur de Haff $\mathcal{D}_s(\cdot)$,

$$\begin{aligned} \operatorname{tr}[C(A^T \mathcal{D}_s)^T \{B\}] &= \sum_{i,j,k,l}^p A_{kl}^T C_{ij} d_{jl} B_{ik}^T \\ &= \sum_{k,l,i}^p A_{kl}^T (C \mathcal{D}_s)_{li}^T B_{ik}^T \\ &= \operatorname{tr}[A^T (C \mathcal{D}_s)^T \{B\}^T]. \end{aligned}$$

□

Lemme A.1

Soit $\rho_k = \operatorname{tr}(S^{+k})/\operatorname{tr}^k(S^+) = \nu^k \operatorname{tr}(S^{+k})$. Alors, pour $k = 2, 3, 4, \dots$,

$$(p \wedge n)^{1-k} \leq \rho_k \leq \rho_{k-1} \leq 1. \quad (\text{A.3})$$

Preuve

Premièrement, nous allons voir que ρ_k décroissant en k . En utilisant la décomposition en valeur propre de $S^+ = H_1 L^{-1} H_1^T$, nous avons

$$\begin{aligned} \operatorname{tr}(S^{+k}) &= \operatorname{tr}(L^{-k}) \\ &= \sum_{i=1}^{p \wedge n} \frac{1}{l_i^k} \leq \sum_{i=1}^{p \wedge n} \frac{1}{l_i^{k-1}} \sum_{j=1}^{p \wedge n} \frac{1}{l_j} \\ &= \operatorname{tr}(S^{+(k-1)}) \operatorname{tr}(S^+). \end{aligned}$$

Par conséquent

$$\rho_k = \frac{\operatorname{tr}(S^{+k})}{\operatorname{tr}^k(S^+)} \leq \frac{\operatorname{tr}(S^{+(k-1)})}{\operatorname{tr}^{k-1}(S^+)} = \rho_{k-1}. \quad (\text{A.4})$$

Deuxièmement, il est facile de voir dans (A.4) que pour $k = 2$ on a $\rho_2 \leq 1$. Maintenant, l'inégalité de Hölder donne

$$\operatorname{tr}(L^{-1} I_{p \wedge n}) \leq \left(\operatorname{tr}(L^{-k}) \right)^{1/k} (p \wedge n)^{(k-1)/k}.$$

$k = 2, 3, 4, \dots$, Par conséquent

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

et donc

$$(p \wedge n)^{1-k} \leq \frac{\operatorname{tr}(L^{-k})}{\operatorname{tr}^k(L^{-1})} = \frac{\operatorname{tr}(S^{+k})}{\operatorname{tr}^k(S^+)} = \rho_k.$$

□

Lemme A.2

Soit $S^+ = H_1 L^{-1} H_1^T$ l'inverse de Moore-Penrose de S . Alors

$$-4(p \wedge n) \operatorname{tr}(S^{+3}) < \sum_{i=1}^{p \wedge n} \sum_{j \neq i}^{p \wedge n} \frac{l_i^{-2} - l_j^{-2}}{l_i - l_j} < 0, \quad (\text{A.5})$$

$$-2(p \wedge n) \operatorname{tr}(S^{+2}) < \sum_{i=1}^{p \wedge n} \sum_{j \neq i}^{p \wedge n} \frac{l_i^{-1} - l_j^{-1}}{l_i - l_j} < 0. \quad (\text{A.6})$$

Preuve

Comme $S^+ = H_1 L^{-1} H_1^T = H_1 \text{diag}(l_1^{-1}, l_2^{-1}, \dots, l_{p \wedge n}^{-1}) H_1^T$, alors

$$\begin{aligned} \sum_{i=1}^{p \wedge n} \sum_{j \neq i}^{p \wedge n} \frac{l_i^{-2} - l_j^{-2}}{l_i - l_j} &= 2 \sum_{i=1}^{p \wedge n} \sum_{j > i}^{p \wedge n} \frac{l_i^{-2} - l_j^{-2}}{l_i - l_j} \\ &= -2 \sum_{i=1}^{p \wedge n} \sum_{j > i}^{p \wedge n} \left[\frac{1}{l_i l_j^2} + \frac{1}{l_i^2 l_j} \right], \end{aligned}$$

qui est un terme négatif. Comme pour $j > i$, nous avons $l_i^{-1} < l_j^{-1}$ et $l_i^{-2} < l_j^{-2}$. Alors

$$\begin{aligned} \sum_{i=1}^n \sum_{j \neq i}^{p \wedge n} \frac{l_i^{-2} - l_j^{-2}}{l_i - l_j} &> -4 \sum_{i=1}^{p \wedge n} \sum_{j > i}^{p \wedge n} \frac{1}{l_j^3} \\ &= -4 \sum_{i=1}^{p \wedge n} \left[\sum_{j=1}^{p \wedge n} \frac{1}{l_j^3} - \sum_{j \leq i} \frac{1}{l_j^3} \right]. \end{aligned}$$

Maintenant, puisque $\sum_{i=1}^{p \wedge n} \sum_{j \leq i}^{p \wedge n} l_j^{-3} > 0$, alors

$$\begin{aligned} \sum_{i=1}^{p \wedge n} \sum_{j \neq i}^{p \wedge n} \frac{l_i^{-2} - l_j^{-2}}{l_i - l_j} &> -4 \sum_{i=1}^{p \wedge n} \sum_{j=1}^{p \wedge n} \frac{1}{l_j^3} \\ &= -4(p \wedge n) \text{tr}(S^{+3}). \end{aligned}$$

ce qui correspond à (A.5).

De manière similaire, pour (A.6) nous avons

$$\begin{aligned} \sum_{i=1}^{p \wedge n} \sum_{j \neq i}^{p \wedge n} \frac{l_i^{-1} - l_j^{-1}}{l_i - l_j} &= -2 \sum_{i=1}^{p \wedge n} \sum_{j > i}^{p \wedge n} \frac{1}{l_i l_j} \\ &> -2 \sum_{i=1}^{p \wedge n} \sum_{j > i}^{p \wedge n} \frac{1}{l_j^2} \\ &= -2 \sum_{i=1}^{p \wedge n} \left[\sum_{j=1}^{p \wedge n} \frac{1}{l_j^2} - \sum_{j \leq i} \frac{1}{l_j^2} \right]. \end{aligned}$$

En utilisant le fait que $\sum_{i=1}^{p \wedge n} \sum_{j \leq i}^{p \wedge n} l_j^{-2} > 0$, il s'en suit que

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

ce qui correspond à (A.6).

□

EXPRESSIONS DIFFÉRENTIELLES

Nous donnons dans cette section des résultats de dérivations de S et de S^+ .

Proposition B.1

Rappelons que $S = U^T U$ est une matrice $p \times p$ symétrique. Pour un entier non nul N , soit A et B deux matrices de dimensions $N \times p$ et $p \times N$, respectivement. Alors

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

Preuve

En appliquant un résultat bien connu sur les dérivées de l'inverse de Moore-Penrose S^+ (voir par exemple Harville [Har08], page 510), nous avons pour tout $1 \leq k \leq n$ et tout $1 \leq j \leq p$,

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

et donc

$$\begin{aligned} \left(A \frac{\partial S^+}{\partial U_{kj}} B \right)_{li} &= - \left(AS^+ \frac{\partial S}{\partial U_{kj}} S^+ B \right)_{li} + \left(A(I_p - SS^+) \frac{\partial S}{\partial U_{kj}} S^+ S^+ B \right)_{li} \\ &\quad + \left(AS^+ S^+ \frac{\partial S}{\partial U_{kj}} (I_p - SS^+) B \right)_{li}. \end{aligned} \tag{B.1}$$

Grâce à (2.7.5), le premier terme situé à droite de (B.1) devient

$$\begin{aligned} \left(AS^+ \frac{\partial S}{\partial U_{kj}} S^+ B \right)_{li} &= \sum_{q,r}^p (AS^+)_{lq} \left(\frac{\partial S}{\partial U_{kj}} \right)_{qr} (S^+ B)_{ri} \\ &= \sum_r^m (AS^+)_{lj} (S^+ B)_{ri} U_{kr} + \sum_{q=1}^n (AS^+)_{lq} (S^+ B)_{ji} U_{kq} \\ &= (AS^+)_{lj} (US^+B)_{ki} + (S^+B)_{ji} (US^+A^T)_{kl}. \end{aligned} \tag{B.2}$$

De manière similaire, les deux autres termes dans la partie droite de (B.1) deviennent respectivement

$$\begin{aligned} \left(A (I_p - SS^+) \frac{\partial S}{\partial U_{kj}} S^+ S^+ B \right)_{li} &= (A (I_p - SS^+))_{lj} (U S^+ S^+ B)_{ki} \\ &+ (S^+ S^+ B)_{ji} (U (I_p - SS^+) A^T)_{kl} \end{aligned} \quad (\text{B.3})$$

et

$$\begin{aligned} \left(A S^+ S^+ \frac{\partial S}{\partial U_{kj}} (I_p - SS^+) B \right)_{li} &= (A S^+ S^+)_{lj} (U (I_p - SS^+) B)_{ki} \\ &+ ((I_p - SS^+) B)_{ji} (U S^+ S^+ A^T)_{kl}. \end{aligned} \quad (\text{B.4})$$

Par conséquent, en remplaçant (B.2), (B.3) et (B.4) dans (B.1) et en utilisant le fait que $U(I_p - SS^+) = 0_{(p \wedge n) \times p}$, le résultat souhaité est obtenu. \square

Lemme B.1

Soit $1 \leq i, j, a, r \leq p$ et $1 \leq k \leq p \wedge n$. Alors

$$d_{ij}^s l_k = (H_1)_{ik} (H_1)_{jk} \quad (\text{B.5})$$

et

$$\begin{aligned} (d_{ij}^s H_1)_{ak} &= \frac{1}{2} \sum_{r \neq k}^{p \wedge n} \frac{(H_1)_{ar}}{l_k - l_r} [(H_1)_{ir} (H_1)_{jk} + (H_1)_{jr} (H_1)_{ik}] \\ &+ \frac{1}{2l_k} [(I_p - H_1 H_1^T)_{ai} (H_1)_{jk} + (I_p - H_1 H_1^T)_{aj} (H_1)_{ik}]. \end{aligned} \quad (\text{B.6})$$

Preuve

Soit $S = H_1 L H_1^T$ la décomposition en valeurs propres de S dans le contexte unifié (S inversible ou S non inversible), avec $H_1 \in \mathcal{L}_{p \times (p \wedge n)}$ et $L \in \mathcal{D}_{(p \wedge n) \times (p \wedge n)}$. on prend une matrice $H_2 \in \mathcal{L}_{p \times (p - p \wedge n)}$ tel que $H_2^T H_1 = 0_{(p - p \wedge n) \times (p \wedge n)}$ afin de construire une matrice orthogonale $H = [H_1, H_2] \in \mathcal{O}_p$. Notons que, dans le cas inversible ($p \leq n$), $H = H_1 \in \mathcal{O}_p$ et on a pas besoin de la matrice H_2

Pour tout opérateur différentiel de S , nous avons

$$dS = (dH_1) L H_1^T + H_1 (dL) H_1^T + H_1 L (dH_1^T),$$

ce qui donne

$$\begin{aligned} H^T (dS) H_1 &= \begin{bmatrix} H_1^T (dS) H_1 \\ H_2^T (dS) H_1 \end{bmatrix} \\ &= \begin{bmatrix} H_1^T (dH_1) L + H_1^T H_1 (dL) + H_1^T H_1 L (dH_1^T) H_1 \\ H_2^T (dH_1) L + H_2^T H_1 (dL) + H_2^T H_1 L (dH_1^T) H_1 \end{bmatrix}. \end{aligned}$$

L'expression différentielle de $H_1^T H_1 = I_{p \wedge n}$ donne

$$(dH_1^T) H_1 = -H_1^T (dH_1).$$

Alors

$$\begin{bmatrix} H_1^T (dS) H_1 \\ H_2^T (dS) H_1 \end{bmatrix} = \begin{bmatrix} H_1^T (dH_1) L - L H_1^T (dH_1) + (dL) \\ H_2^T (dH_1) L \end{bmatrix}$$

puisque $H_2^T H_1 = 0_{(p-p \wedge n) \times (p \wedge n)}$. Par conséquent, pour $1 \leq k \leq p \wedge n$

$$\begin{aligned} (dL)_k &= \{H_1^T (dS) H_1\}_{kk} - \{H_1^T (dH_1) L\}_{kk} + \{L H_1^T (dH_1)\}_{kk} \\ &= \{H_1^T (dS) H_1\}_{kk}, \end{aligned} \quad (\text{B.7})$$

puisque $\{L H_1^T (dH_1)\}_{kk} - \{H_1^T (dH_1) L\}_{kk} = 0$.

Maintenant, pour $r \neq k$, nous avons

$$\begin{aligned} \{H_1^T (dS) H_1\}_{rk} &= \{H_1^T (dH_1)\}_{rk} (l_k - l_r) \quad \text{pour } r, k = 1, \dots, p \wedge n \\ \{H_2^T (dS) H_1\}_{rk} &= \{H_2^T (dH_1)\}_{rk} l_k \quad \text{pour } r = p \wedge n + 1, \dots, p \text{ et } k = 1, \dots, p \wedge n. \end{aligned}$$

Donc,

$$\begin{aligned} (H_1^T (dH_1))_{rk} &= \frac{1}{l_k - l_r} \{H_1^T (dS) H_1\}_{rk} \quad \text{pour } r, k = 1, \dots, p \wedge n \\ (H_2^T (dH_1))_{rk} &= \frac{1}{l_k} \{H_2^T (dS) H_1\}_{rk} \quad \text{pour } r = p \wedge n + 1, \dots, p \text{ and } k = 1, \dots, p \wedge n. \end{aligned} \quad (\text{B.8})$$

Maintenant, si "d" est l'opérateur de Haff dans (2.7.1), nous avons

$$(d_{ij}^s S)_{cd} = \frac{1}{2} (\delta_{ic} \delta_{jd} + \delta_{id} \delta_{jc}).$$

alors

$$\begin{aligned} \{H_1^T (d_{ij}^s S) H_1\}_{rk} &= \frac{1}{2} [(H_1)_{ir} (H_1)_{jk} + (H_1)_{jr} (H_1)_{ik}], \\ \{H_2^T (d_{ij}^s S) H_1\}_{rk} &= \frac{1}{2} [(H_2)_{ir} (H_1)_{jk} + (H_2)_{jr} (H_1)_{ik}]. \end{aligned} \quad (\text{B.9})$$

Par conséquent, en utilisant (B.9), l'égalité (B.7) devient

$$d_{ij}^s l_k = (H_1)_{ik} (H_1)_{jk}$$

qui est le résultat annoncé dans (B.5).

Pour (B.6), notons que, puisque $H H^T = I_p$, alors

$$\begin{aligned} (d_{ij}^s H_1)_{ak} &= \{H H^T (d_{ij}^s H_1)\}_{ak} \\ &= \sum_{r \neq k}^{p \wedge n} (H_1)_{ar} \{H_1^T (d_{ij}^s H_1)\}_{rk} + \sum_{r=(p \wedge n)+1}^p (H_2)_{ar} \{H_2^T (d_{ij}^s H_1)\}_{rk}. \end{aligned} \quad (\text{B.10})$$

On combine (B.8) et (B.9), pour avoir que, lorsque $r \neq k$

$$\begin{aligned} (H_1^T (d_{ij}^s H_1))_{rk} &= \frac{(H_1)_{ir} (H_1)_{jk} + (H_1)_{jr} (H_1)_{ik}}{2(l_k - l_r)}, \\ (H_2^T (d_{ij}^s H_1))_{rk} &= \frac{(H_2)_{ir} (H_1)_{jk} + (H_2)_{jr} (H_1)_{ik}}{2l_k}. \end{aligned}$$

Par conséquent (B.10) devient

$$(d_{ij}^s H_1)_{ak} = \frac{1}{2} \sum_{r \neq k} \frac{(H_1)_{ar}}{l_k - l_r} [(H_1)_{ir}(H_1)_{jk} + (H_1)_{jr}(H_1)_{ik}] \\ + \frac{1}{2l_k} [(H_2 H_2^T)_{ai}(H_1)_{jk} + (H_2 H_2^T)_{aj}(H_1)_{ik}],$$

ce qui donne le résultat annoncé dans (B.6) puisque $H_2 H_2^T = (I_p - H_1 H_1^T)$.

□

Le lemme suivant est une généralisation des résultats obtenus par Haff [Haf91] aux cas inversible et non inversible.

Lemme B.2

Sous les notations de la preuve du lemme B.1, soit $G = H_1 \Psi H_1^T$ tel que $\Psi \in \mathcal{D}_{p \wedge n}$. Alors $SS^+ G = G$ et

$$\mathcal{D}_s\{SS^+ G\} = \mathcal{D}_s\{H_1 \Psi H_1^T\} = H_1 \Psi^{(1)} H_1^T + \frac{1}{2} \text{tr}(L^{-1} \Psi) (I_p - H_1 H_1^T) \quad (\text{B.11})$$

tel que $\Psi^{(1)} \in \mathcal{D}_{p \wedge n}$ où

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

En outre

$$\mathcal{D}_s\{S\} = \mathcal{D}_s\{H_1 L H_1^T\} \\ = \frac{1}{2}(p+1)H_1 H_1^T + \frac{1}{2}(p \wedge n)(I_p - H_1 H_1^T), \quad (\text{B.12})$$

$$\mathcal{D}_s\{SS^+\} = \mathcal{D}_s\{H_1 H_1^T\} \\ = \frac{1}{2}(p - p \wedge n)H_1 L H_1^T + \frac{1}{2} \text{tr}(L^{-1})(I_p - H_1 H_1^T) \quad (\text{B.13})$$

et

$$(SS^+ \mathcal{D}_s)^T\{SS^+ G\} = (H_1 H_1^T \mathcal{D}_s)^T\{H_1 \Psi H_1^T\} \\ = \mathcal{D}_s\{H_1 \Psi H_1\} - \frac{1}{2}(p-n)(H_1 L^{-1} \Psi H_1^T). \quad (\text{B.14})$$

Preuve

Les éléments (i, a) de la matrice $\mathcal{D}_s\{H_1 \Psi H_1^T\}$ sont exprimés comme suit

$$(\mathcal{D}_s\{H_1 \Psi H_1^T\})_{ia} = \sum_j^p \sum_b^{p \wedge n} d_{ij}^s \{(H_1)_{jb} \psi_b (H_1)_{ab}\} = A_{ia} + B_{ia} + C_{ia},$$

où

$$A_{ia} = \sum_j^p \sum_b^{p \wedge n} \psi_b (H_1)_{ab} (d_{ij}^s H_1)_{jb}, \quad (\text{B.15})$$

$$B_{ia} = \sum_j^p \sum_b^{p \wedge n} \psi_b (H_1)_{jb} (d_{ij}^s H_1)_{ab} \quad (\text{B.16})$$

et

$$C_{ia} = \sum_j^p \sum_b^{p \wedge n} (H_1)_{jb} (H_1)_{ab} (d_{ij}^s \psi)_b. \quad (\text{B.17})$$

Nous traitons d'abord (B.15) et (B.16). Grâce à (B.6), nous avons

$$\begin{aligned} A_{ia} &= \frac{1}{2} \sum_j^p \sum_b^{p \wedge n} \sum_{r \neq b}^{p \wedge n} \frac{\psi_b (H_1)_{ab} (H_1)_{jr}}{l_b - l_r} [(H_1)_{ir} (H_1)_{jb} + (H_1)_{jr} (H_1)_{ib}] \\ &\quad + \frac{1}{2} \sum_j^p \sum_b^{p \wedge n} \frac{\psi_b}{l_b} [(I_p - H_1 H_1^T)_{ji} (H_1)_{jb} + (I_p - H_1 H_1^T)_{jj} (H_1)_{ib}] \end{aligned}$$

et

$$\begin{aligned} B_{ia} &= \frac{1}{2} \sum_j^p \sum_b^{p \wedge n} \sum_{r \neq b}^{p \wedge n} \frac{\psi_b (H_1)_{jb} (H_1)_{ar}}{l_b - l_r} [(H_1)_{ir} (H_1)_{jb} + (H_1)_{jr} (H_1)_{ib}] \\ &\quad + \frac{1}{2} \sum_j^p \sum_b^{p \wedge n} \frac{\psi_b}{l_b} [(I_p - H_1 H_1^T)_{ai} (H_1)_{jb} + (I_p - H_1 H_1^T)_{aj} (H_1)_{ik}]. \end{aligned}$$

En sommant sur j , on a

$$\begin{aligned} A_{ia} &= \frac{1}{2} \sum_b^{p \wedge n} \sum_{r \neq b}^{p \wedge n} \frac{\psi_b (H_1)_{ab}}{l_b - l_r} [(H_1)_{ir} (H_1^T H_1)_{rb} + (H_1^T H_1)_{rr} (H_1)_{ib}] \\ &\quad + \frac{1}{2} \sum_b^{p \wedge n} \frac{\psi_b}{l_b} [(H_1^T (I_p - H_1 H_1^T))_{bi} + \text{tr}(I_p - H_1 H_1^T) (H_1)_{ib}] \end{aligned}$$

et

$$\begin{aligned} B_{ia} &= \frac{1}{2} \sum_b^{p \wedge n} \sum_{r \neq b}^{p \wedge n} \frac{\psi_b (H_1)_{ar}}{l_b - l_r} [(H_1)_{ir} (H_1^T H_1)_{bb} + (H_1^T H_1)_{br} (H_1)_{ib}] \\ &\quad + \frac{1}{2} \sum_b^{p \wedge n} \frac{\psi_b}{l_b} [(I_p - H_1 H_1^T)_{ai} (H_1^T H_1)_{bb} + ((I_p - H_1 H_1^T) H_1)_{ab} (H_1)_{ik}]. \end{aligned}$$

Puisque $H_1^T H_1 = I_{p \wedge n}$ nous avons, $H_1^T (I_p - H_1 H_1^T) = ((I_p - H_1 H_1^T) H_1)^T = 0_{(p \wedge n) \times p}$ et $\text{tr}(I_p - H_1 H_1^T) = p - (p \wedge n)$. Donc

$$A_{ia} = \frac{1}{2} \sum_b^{p \wedge n} (H_1)_{ib} \sum_{r \neq b}^{p \wedge n} \frac{\psi_b}{l_b - l_r} (H_1^T)_{ba} + \frac{1}{2} (p - (p \wedge n)) \sum_b^{p \wedge n} (H_1)_{ib} \frac{\psi_b}{l_b} (H_1^T)_{ba} \quad (\text{B.18})$$

et

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

L'expression de B_{ia} peut être spécifiée en remarquant que

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

Donc

$$B_{ia} = -\frac{1}{2} \sum_b^{p \wedge n} (H_1)_{ib} \sum_{b \neq r}^{p \wedge n} \frac{\psi_r}{l_b - l_r} (H_1^T)_{ba} + \frac{1}{2} \text{tr}(L^{-1} \Psi) (I_p - H_1 H_1^T)_{ia}. \quad (\text{B.19})$$

Pour le terme C_{ia} dans (B.17), la règle de dérivation en chaîne donne

$$C_{ia} = \sum_j^p \sum_b^n (H_1)_{jb} (H_1)_{ab} \sum_q^n \frac{\partial \psi_b}{\partial l_q} d_{ij} l_q.$$

l'utilisation de (B.5) donne

$$C_{ia} = \sum_b^n (H_1)_{ab} \sum_q^n \frac{\partial \psi_b}{\partial l_q} (H_1)_{iq} (H_1 H_1^T)_{qb},$$

et donc ,

$$C_{ia} = \sum_b^{p \wedge n} (H_1)_{ib} \frac{\partial \psi_b}{\partial l_b} (H_1^T)_{ba}, \quad (\text{B.20})$$

puisque $H_1^T H_1 = I_{p \wedge n}$. Finalement, on rassemble les expressions (B.18), (B.19) et (B.20) pour avoir le résultat annoncé dans (B.11).

Maintenant, afin de prouver (B.12) (respectivement (B.13)) on applique (B.11) pour $\Psi = L$ (respectivement $\Psi = I_{p \wedge b}$), ce qui donne le résultat souhaité puisque $\Psi^{(1)} = (p+1)/2 I_{p \wedge n}$ (respectivement $\Psi^{(1)} = (p+1)/2 (p - p \wedge n) L^{-1}$).

Pour (B.14), nous avons de (A.1) que

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

Par conséquent, d'après (B.13), nous avons

$$(H_1 H_1^T \mathcal{D}_s)^T \{H_1 \Psi H_1^T\} = \mathcal{D}_s \{H_1 \Psi H_1\} - \frac{1}{2} (p - p \wedge n) H_1 L^{-1} H_1^T.$$

□

Lemme B.3

Dans le contexte de (2.6.7), supposons qu'il existe deux constantes positives c et b telles que pour tout $t \in \mathbb{R}^+$

$$c \leq \frac{F^*(t)}{f(t)} \leq b. \quad (\text{B.21})$$

Nous avons alors

$$\frac{1}{b^2} \leq \frac{1}{K^* \varphi_{\theta, \Sigma}^{**}(z, u)} \leq \frac{1}{c^2}, \quad (\text{B.22})$$

où K^* est donnée dans (2.6.2).

Preuve

Par définition de $F^{**}(\cdot)$ et $F^*(\cdot)$ dans (2.6.1) et (2.6.5), nous avons pour tout $t \in \mathbb{R}^+$

$$\frac{F^*(t)}{F^{**}(t)} = \left(\int_t^\infty F^*(x) dx \right)^{-1} \int_t^\infty \frac{f(x)}{F^*(x)} F^*(x) dx = E_t \left[\frac{f(X)}{F^*(X)} \right]$$

où E_t est l'espérance par rapport à la densité proportionnelle à $F^*(x) \mathbb{1}_{[t, \infty[}(x)$. Par conséquent, grâce à (2.6.7) et (B.21), avec $t = v$, nous avons

$$\frac{1}{b} \leq \frac{1}{\varphi_{\theta, \Sigma}^{**}(z, u)} = E_v \left[\frac{f(X)}{F^*(X)} \right] \leq \frac{1}{c}.$$

De plus, en posant $H(Z, U) \equiv 1$ dans (2.6.8), nous avons, selon (2.6.7) et par l'hypothèse (B.21) que

$$c \leq K^* \leq b.$$

Par conséquent,

$$1/b \leq K^* \leq 1/c$$

et (B.21) donnent (B.22).

□

ANNEXE **C**

ARTICLE SOUMIS

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

Dominique Fourdrinier * Anis M. Haddouche † Fatiha Mezoued ‡

June 4, 2019

Abstract

The problem of estimating the scale matrix Σ of an elliptically symmetric distribution is considered from a decision-theoretic point of view. As the natural estimators of the form $\hat{\Sigma}_a = aS$ (where S is the sample covariance matrix and a is a positive constant) perform poorly, we propose estimators of the form $\hat{\Sigma}_{a,G} = a(S + SS^+G(Z, S))$, where S^+ is the Moore-Penrose inverse of S and $G(Z, S)$ is a correction matrix and we provide conditions on $G(Z, S)$ such that $\hat{\Sigma}_{a,G}$ improve over $\hat{\Sigma}_a$ under the quadratic loss $L(\Sigma, \hat{\Sigma}) = \text{tr}(\hat{\Sigma}\Sigma^{-1} - I_p)^2$. 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.

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

Key words and phrases: elliptically symmetric distributions, high-dimensional statistics, quadratic loss, orthogonally invariant estimators, Stein-Haff type identities.

*Dominique Fourdrinier is Professor, Université de Normandie, UNIROUEN, UNIHAVRE, INSA Rouen, LITIS, avenue de l'Université, BP 12, 76801 Saint-Étienne-du-Rouvray, France. E-mail: Dominique.Fourdrinier@univ-rouen.fr.

†Anis M. Haddouche is Ph.D Student, École Nationale Supérieure de Statistique et d'Économie Appliquée (ENSSEA), LAMOPS, Tipaza, Algeria and Université de Normandie, UNIROUEN, UNIHAVRE, INSA Rouen, LITIS, avenue de l'Université, BP 12, 76801 Saint-Étienne-du-Rouvray, France. E-mail: mohamed-anis.haddouche@etu.univ-rouen.fr.

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

1 Introduction

Let Y be an observed $n \times p$ matrix with $p < n$. We consider the additive model

$$Y = M + \mathcal{E}, \quad \mathcal{E} \sim \mathcal{E}(0_{np}, I_n \otimes \Sigma), \quad (1.1)$$

where M denotes the unknown matrix of parameters, \mathcal{E} is elliptically symmetric distributed noise, with unknown covariance matrix proportional to $I_n \otimes \Sigma$ where Σ is a $p \times p$ invertible matrix and I_n is the n -dimensional identity matrix. The model (1.1) was considered by Canu and Fourdrinier [3] and is an extension of the Gaussian model in Candès et al. [2], where $\mathcal{E} \sim \mathcal{N}(0_{np}, I_n \otimes \Sigma)$. These authors were interested in estimating M while, here, we wish to estimate Σ . As in [2] and [3], we assume that M is of low-rank (or can be approximated by a low-rank matrix), that is

$$\text{rank}(M) = q < p \quad (1.2)$$

Although any estimator of Σ is a function of the $n \times p$ observation matrix Y (such as the usual estimator $a Y^T Y$, where a is a positive constant), this assumption of low-rank allows to essentially reduce the information in Y to an $(n - q) \times p$ matrix U (so that $a U^T U$ estimates Σ). Due to the low-rank assumption in (1.2), there exists Q_2 , a semi-orthogonal $n \times (n - q)$ matrix, such that

$$Q_2^T M = 0_{(n-q) \times p}. \quad (1.3)$$

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

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

with $\theta = Q_1^T M$, thanks to (1.3). The canonical form of the model (1.1) is

$$\begin{pmatrix} Z \\ U \end{pmatrix} = \begin{pmatrix} \theta \\ 0 \end{pmatrix} + Q^T \mathcal{E}, \quad (1.4)$$

where Z and U are respectively $q \times p$ and $(n - q) \times p$ matrices. Now, if we assume that $\mathcal{E} = Y - M$ has a density, it is necessarily of the form

$$\mathcal{E} \mapsto |\Sigma|^{-n/2} f\{\text{tr}(\mathcal{E} \Sigma^{-1} \mathcal{E}^T)\},$$

for some function f (see Canu and Fourdrinier [3]). Since Q is orthogonal, then $Q^T \mathcal{E}$ has density

$$\mathcal{E} \mapsto |\Sigma|^{-n/2} f\{\text{tr}(Q^T \mathcal{E} \Sigma^{-1} \mathcal{E}^T Q)\} = |\Sigma|^{-n/2} f\{\text{tr}(\mathcal{E} \Sigma^{-1} \mathcal{E}^T)\}. \quad (1.5)$$

It follows that $(Z^T U^T)^T = Q^T Y$ has an elliptically symmetric distribution around the matrix $(\theta^T 0^T)^T$. Hence, from (1.4) and (1.5), the joint density of Z and U is

$$\begin{aligned} (z, u) &\mapsto |\Sigma|^{-n/2} f\left[\text{tr}\left\{\begin{pmatrix} z - \theta \\ u \end{pmatrix} \Sigma^{-1} \begin{pmatrix} z - \theta \\ u \end{pmatrix}^T\right\}\right] \\ &= |\Sigma|^{-n/2} f\left[\text{tr}\{(z - \theta)\Sigma^{-1}(z - \theta)^T\} + \text{tr}\{\Sigma^{-1} u^T u\}\right]. \end{aligned} \quad (1.6)$$

Let $m = n - q$. We deal with estimating Σ for both cases $p \leq m$ and $p > m$, under the quadratic loss function

$$L(\Sigma, \hat{\Sigma}) = \text{tr}(\hat{\Sigma} \Sigma^{-1} - I_p)^2, \quad (1.7)$$

and its associated risk

$$R(\Sigma, \hat{\Sigma}) = E_{\theta, \Sigma}[L(\Sigma, \hat{\Sigma})], \quad (1.8)$$

where $\hat{\Sigma}$ is an estimator of Σ . Note that, as Σ^{-1} naturally intervenes in the parameterization of the density (1.6), estimating Σ is often more difficult than estimating Σ^{-1} .

The usual estimators of Σ are of the form $\hat{\Sigma}_a = aS$, where $a > 0$ and the sample covariance matrix $S = U^T U = Y^T Q_2 Q_2^T Y$. Two cases have to be distinguished: the case where S is invertible ($p \leq m$) and the case where S is singular ($p > m$). As pointed out by James and Stein [14], these estimators are inadmissible under a normal multivariate distribution. Since then, numerous authors have suggested improved estimators which improves over the unbiased estimator $S/(m-1)$ and over the maximum likelihood estimator S/m .

In the invertible case ($p \leq m$), the literature includes, Stein [20], [21], Efron and Morris [5], Haff [10], Takemura [22], Dey and Srinivasan [4], Sheena [19]. Also Kubokawa and Srivastava [17] showed that, under the Stein loss function $\text{tr}(\hat{\Sigma}\Sigma) - \log|\hat{\Sigma}\Sigma| - p$, the improved estimators proposed under the normality assumption remain robust in the sense that they are still improved estimators under the elliptically assumption (1.6). However, to our knowledge, a similar extension under the quadratic loss (1.7) has not yet been obtained.

More recently, attention has turned to the non-invertible case ($p > m$), often referred to as the high-dimensional setting. In this case, the usual estimators remain inadmissible. In the Gaussian setting, Konno [15] extends the result due to Haff [10] in the invertible case to the high-dimensional setting and proposed improved estimators under quadratic loss (1.7). However, as in the invertible case ($p \leq m$), no results have been yet established in the general elliptical framework under this loss.

In this paper, we give dominance results relative to the quadratic loss (1.7) in the elliptical setting (1.6) through a unified approach in the invertible and singular cases. For that purpose, we will denote by S^+ the Moore-Penrose pseudo inverse when S is singular, and the regular inverse S^{-1} when it is invertible. In Section 2, we consider alternative estimators of the form $\hat{\Sigma}_{a,G} = a(S + SS^+G(Z, S))$ where a is a positive constant. We give sufficient conditions on the correction matrix function $G(Z, S)$ for which $\hat{\Sigma}_{a,G}$ improves on $\hat{\Sigma}_a$, for a large subclass of elliptically symmetric distributions in (1.6). To do this, we derive a new version of the so-called Stein-Haff identity which distinguishes the invertible and the singular cases. In Section 3, we apply the results of Section 2 to the class of orthogonally invariant estimators. We also extend the estimators due to Haff [12], based on a Gaussian assumption, to the class of elliptically symmetric distributions. Examples illustrate the theory. Finally, an Appendix contains technical results and the proofs of some of the findings in this paper.

2 General case

Consider an alternative estimator for Σ of the form

$$\hat{\Sigma}_{a,G} = a(S + SS^+G(Z, S)), \quad (2.1)$$

where $a > 0$, S^+ denotes the Moore-Penrose pseudo-inverse matrix of S and $G(Z, S)$ is a $p \times p$ matrix function such that the correction factor $SS^+G(Z, S)$ is symmetric.

We develop conditions on $G := G(Z, S)$ in order that $\hat{\Sigma}_{a,G}$ improves on $\hat{\Sigma}_a$ under the quadratic risk (1.8), that is, such that the risk difference

$$\Delta(G) = R(\Sigma, a(S + SS^+G(Z, S))) - R(\Sigma, aS) \quad (2.2)$$

is non-positive. To this end, we give in the following proposition conditions to ensure the finiteness of the risks difference (2.2). The proof is deferred to [Appendix A.3](#).

Proposition 2.1 *Let $\|A\|_F = \sqrt{\text{tr}(A^T A)}$ be the Frobenius norm. Assume that $E_{\theta, \Sigma} [\|\Sigma^{-1} S\|_F^2] < \infty$ and $E_{\theta, \Sigma} [\|\Sigma^{-1} S S^+ G\|_F^2] < \infty$. Then the risks of $\hat{\Sigma}_a$ and $\hat{\Sigma}_{a, G}$ are finite. In that case, the risk difference in (2.2) is also finite and can be written as*

$$\Delta(G) = a^2 E_{\theta, \Sigma} [\text{tr}(\Sigma^{-1} S S^+ (2S + G) \Sigma^{-1} S S^+ G)] - 2a E_{\theta, \Sigma} [\text{tr}(\Sigma^{-1} S S^+ G)] . \quad (2.3)$$

The dependence in (2.3) 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 distribution. For this purpose, we define $E_{\theta, \Sigma}^*$ as the expectation with respect to the density

$$(z, u) \mapsto \frac{1}{K^*} |\Sigma|^{-n/2} F^* [\text{tr}\{(z - \theta) \Sigma^{-1} (z - \theta)^T\} + \text{tr}\{\Sigma^{-1} u^T u\}] \quad (2.4)$$

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

$$F^*(t) = \frac{1}{2} \int_t^\infty f(v) dv \quad (2.5)$$

and the normalizing constant

$$K^* = \int_{\mathbb{R}^{pm}} |\Sigma|^{-n/2} F^* [\text{tr}\{(z - \theta) \Sigma^{-1} (z - \theta)^T\} + \text{tr}\{\Sigma^{-1} u^T u\}] dz du \quad (2.6)$$

is assumed to be finite. Also, we will use the Haff operator \mathcal{D}_s whose generic element is

$$d_{ij}^s = \frac{1}{2} (1 + \delta_{ij}) \frac{\partial}{\partial S_{ij}}, \quad (2.7)$$

with $\delta_{ij} = 1$ when $i = j$ and $\delta_{ij} = 0$ when $i \neq j$. Denote, in the following, $p \wedge m = \min\{p, m\}$.

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 $E_{\theta, \Sigma} [\text{tr}(\Sigma^{-1} S S^+ G)] < \infty$. Then we have*

$$E_{\theta, \Sigma} [\text{tr}(\Sigma^{-1} S S^+ G)] = K^* E_{\theta, \Sigma}^* [\text{tr}(2 S S^+ \mathcal{D}_s \{S S^+ G\}^T + (m - (p \wedge m) - 1) S^+ G)] . \quad (2.8)$$

See [Appendix A.3](#) for the proof.

Note that, in the case where S is invertible ($p \leq m$), since $S^+ = S^{-1}$, Identity (2.8) corresponds to

$$E_{\theta, \Sigma} [\text{tr}(\Sigma^{-1} G)] = K^* E_{\theta, \Sigma}^* [\text{tr}(2 \mathcal{D}_s \{G\}^T + (m - p - 1) S^{-1} G)] ,$$

which is the identity given by Kubokawa and Srivastava [17]. In the singular case, Identity (2.8) becomes

$$E_{\theta, \Sigma} [\text{tr}(\Sigma^{-1} S S^+ G)] = K^* E_{\theta, \Sigma}^* [\text{tr}(2 S S^+ \mathcal{D}_s \{S S^+ G\}^T - S^+ G)] ,$$

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 of the right-hand side of (2.3), 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

$$(z, u) \mapsto \frac{1}{K^{**}} |\Sigma|^{-n/2} F^{**} [\text{tr}\{(z - \theta) \Sigma^{-1} (z - \theta)^T\} + \text{tr}\{\Sigma^{-1} u^T u\}]$$

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

$$F^{**}(t) = \frac{1}{2} \int_t^\infty F^*(v) dv \quad (2.9)$$

and the normalizing constant

$$K^{**} = \int_{\mathbb{R}^{pn}} |\Sigma|^{-n/2} F^{**} [\text{tr}\{(z - \theta) \Sigma^{-1} (z - \theta)^T\} + \text{tr}\{\Sigma^{-1} u^T u\}] dz du \quad (2.10)$$

is assumed to be finite.

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 . Assume that $SS^+V(Z, S)$ is symmetric such that $E_{\theta, \Sigma} [\text{tr}(\Sigma^{-1}SS^+V\Sigma^{-1}SS^+G)] < \infty$ and $E_{\theta, \Sigma}^* [\text{tr}(\Sigma^{-1}SS^+T^*)] < \infty$. Then we have*

$$E_{\theta, \Sigma} [\text{tr}(\Sigma^{-1}SS^+V\Sigma^{-1}SS^+G)] = K^* K^{**} E_{\theta, \Sigma}^* [\text{tr}(2SS^+\mathcal{D}_s\{SS^+T^*\}^T + (m - (p \wedge m) - 1)S^+T^*)] \quad (2.11)$$

with

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

See [Appendix A.3](#) for the proof.

Identity (2.11) parallels Identity (2.8) as the role of the density (1.6) (respectively, the role of the expectation $E_{\theta, \Sigma}$) is played by the density (2.4) (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 $E_{\theta, \Sigma}^* [\text{tr}(\Sigma^{-1}SS^+H)] < \infty$,

$$E_{\theta, \Sigma}^* [\text{tr}(\Sigma^{-1}SS^+H)] = K^{**} E_{\theta, \Sigma}^* [\text{tr}(2S^+S\mathcal{D}_s\{SS^+H\}^T + (m - (p \wedge m) - 1)S^+H)].$$

Now we can give an expression of the risk difference $\Delta(G)$ in (2.3) 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 $SS^+G(Z, S)$ is symmetric. Assume also that $E_{\theta, \Sigma} [\|\Sigma^{-1}S\|_F^2] < \infty$, $E_{\theta, \Sigma} [\|\Sigma^{-1}SS^+G\|_F^2] < \infty$ and $E_{\theta, \Sigma}^* [\text{tr}(\Sigma^{-1}SS^+T^*)] < \infty$. Then*

$$\begin{aligned} \Delta(G) &= a^2 K^* K^{**} E_{\theta, \Sigma}^* [\text{tr}(2SS^+\mathcal{D}_s\{SS^+T^*\}^T + (m - (p \wedge m) - 1)S^+T^*)] \\ &\quad - 2a K^* E_{\theta, \Sigma}^* [\text{tr}(2S^+S\mathcal{D}_s\{SS^+G\} + (m - (p \wedge m) - 1)S^+G)] \end{aligned} \quad (2.12)$$

where

$$T^* = 4(S + SS^+G)\mathcal{D}_s\{SS^+G\} + G(2mSS^+ - (p - m + 1)S^+G). \quad (2.13)$$

Proof

Recall that the conditions $E_{\theta,\Sigma} [\|\Sigma^{-1} S\|_F^2] < \infty$ and $E_{\theta,\Sigma} [\|\Sigma^{-1} S S^+ G\|_F^2] < \infty$ 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](#) $E_{\theta,\Sigma} [\|\text{tr}(\Sigma^{-1} S S^+ G)\|] < \infty$ since, by the Cauchy-Schwarz inequality

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

Similarly, note also that these conditions imply, with the condition $E_{\theta,\Sigma}^* [\|\text{tr}(\Sigma^{-1} S S^+ T^*)\|] < \infty$, the conditions in [Corollary 2.1](#) with T^* in [\(2.13\)](#). In particular, for $V = 2S + G$ we have $E_{\theta,\Sigma} [\|\text{tr}(\Sigma^{-1} S S^+ V \Sigma^{-1} S S^+ G)\|] < \infty$.

Then, applying [Lemma 2.1](#) to the second expectation in [\(2.3\)](#) and [Corollary 2.1](#) to the first one with $V = 2S + G$ gives [\(2.12\)](#), since

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

where we used [\(A.4.13\)](#), the fact that $SS^+(I_p - SS^+) = 0_{p \times p}$ and the symmetry of $SS^+ G$. \square

The risk difference in [\(2.12\)](#) involves expressions through the expectations $E_{\theta,\Sigma}^*$ and $E_{\theta,\Sigma}^{**}$. In order to compare these quantities, we reduce the distributional context in [\(1.6\)](#) to the subclass of densities such that, for all $t \in \mathbb{R}^+$,

$$c \leq \frac{F^*(t)}{f(t)} \leq b \tag{2.14}$$

where c and b are positive constants. The subclass of densities satisfying [Condition \(2.14\)](#) has been considered by Fourdrinier, Mezoued and Strawderman [\[7\]](#). This is a large subclass containing the multivariate normal distribution and the logistic type distribution given as an example in [Appendix A.1](#).

As we will see that 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 $a > 0$). The improvement over the class of $a S$'s will be shown through the improvement of $\hat{\Sigma}_{a_o,G} = a_o S + a_o SS^+ G(Z, S)$. The following proposition gives the optimal value of a_o , which holds whatever S is invertible or not. See [Appendix A.3](#) for the proof.

Proposition 2.3 *For the class of estimators $\hat{\Sigma}_a$, where $a > 0$, the best uniformly constant a which minimizes the risk in [\(1.8\)](#) is*

$$a_o = \frac{1}{K^{**} (p + m + 1)} , \tag{2.15}$$

where K^{**} is given in [\(2.10\)](#).

In the following theorem, we provide improvement conditions of $a_o(S + SS^+ G(Z, S))$ over $a_o S$. See [Appendix A.3](#) for the proof.

Theorem 2.1 *Consider a density as in [\(1.6\)](#) satisfying [\(2.14\)](#). Under the condition*

$$\text{tr} [2SS^+ \mathcal{D}_s\{SS^+ G\} + (m - (p \wedge m) - 1) S^+ G] \geq 0 , \tag{2.16}$$

the estimator $\hat{\Sigma}_{a_o, G} = a_o (S + SS^+ G(Z, S))$ improves over $\hat{\Sigma}_{a_o} = a_o S$ as soon as

$$\begin{aligned} & \text{tr} \left[2 S^+ S \mathcal{D}_s \{ SS^+ T^* \}^T + (m - p \wedge m - 1) S^+ T^* \right. \\ & \quad \left. - 2(p + m + 1) \frac{c^2}{b^2} (2 SS^+ \mathcal{D}_s \{ SS^+ G \} + (m - (p \wedge m) - 1) S^+ G) \right] \leq 0, \end{aligned} \quad (2.17)$$

where

$$T^* = 4(S + SS^+ G) \mathcal{D}_s \{ SS^+ G \} + G (2m SS^+ - (p - m + 1) S^+ G). \quad (2.18)$$

As is well known, the usual estimators of Σ are inadmissible. In fact, the largest (smallest) eigenvalues of $\hat{\Sigma}_a$ tends to be larger (smaller) than the corresponding eigenvalues of Σ . This fact suggest to shrink the eigenvalues of $\hat{\Sigma}_a$ toward a central value, which give rise to an orthogonally invariant estimators. **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 . First, we set some notation. Let \mathcal{O}_p be the group of $p \times p$ orthogonal matrices, and when $p \geq r$, let $\mathcal{L}_{p,r} = \{A \in \mathbb{R}^{p \times r} : A^T A = I_r\}$ 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 $\text{diag}(d_1, \dots, d_r)$ such that $d_1 > \dots > d_r > 0$.

Let $S = H_1 L H_1^T$, 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 $\hat{\Sigma}_{a, G} = a (S + S S^+ G(Z, S))$ in (2.1) that are orthogonally invariant: that is, such that

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

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

$$\hat{\Sigma}_{a, \Psi} = a H_1 (L + \Psi(L)) H_1^T. \quad (3.1)$$

In (3.1), the function $\Psi(L)$ modifies the eigenvalues in L . In the Gaussian context, Stein [21], Dey and Srinivasan [4], Haff [11], Kubokawa et al. [16], Ledoit and Wolf [18], Fisher and Sun [6] 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 (1.6) satisfying (2.14). Provided that

$$\sum_{i=1}^{p \wedge m} \left\{ (p + m - 2(p \wedge m) - 1) \frac{\psi_i}{l_i} + 2 \frac{\partial \psi_i}{\partial l_i} + \sum_{j \neq i}^{p \wedge m} \frac{\psi_i - \psi_j}{l_i - l_j} \right\} \geq 0, \quad (3.2)$$

the estimator $\hat{\Sigma}_{a_o, \Psi} = a_o H_1 \text{diag}(l_1 + \psi_1, l_2 + \psi_2, \dots, l_{p \wedge m} + \psi_{p \wedge m}) H_1^T$ improves over the optimal

estimator $\hat{\Sigma}_{a_o} = a_o H_1 \text{diag}(l_1, l_2, \dots, l_{p \wedge m}) H_1^T$ as soon as

$$\sum_{i=1}^{p \wedge m} \left\{ (p+m-2(p \wedge m)-1) \frac{\phi_i}{l_i} + 2 \frac{\partial \phi_i}{\partial l_i} + \sum_{j \neq i}^{p \wedge m} \frac{\phi_i - \phi_j}{l_i - l_j} - 2(p+m+1) \frac{c^2}{b^2} \left((p+m-2(p \wedge m)-1) \frac{\psi_i}{l_i} + 2 \frac{\partial \psi_i}{\partial l_i} + \sum_{j \neq i}^{p \wedge m} \frac{\psi_i - \psi_j}{l_i - l_j} \right) \right\} \leq 0, \quad (3.3)$$

where

$$\begin{aligned} \phi_i &= 2(p+m-p \wedge m) \psi_i + (p+m-2(p \wedge m)-1) \frac{\psi_i^2}{l_i} \\ &\quad + 4(\psi_i + l_i) \frac{\partial \psi_i}{\partial l_i} + 2(\psi_i + l_i) \sum_{j \neq i}^{p \wedge m} \frac{\psi_i - \psi_j}{l_i - l_j}. \end{aligned} \quad (3.4)$$

Proof Using the eigenvalue decomposition of S , it is clear that $S^+ = H_1 L^{-1} H_1^T$, $SS^+ = H_1^T H_1$, $SS^+G = G = H_1 \Psi H_1^T$ and $S^+G = H_1 L^{-1} \Psi H_1^T$. In this setting, Condition (2.16) in [Theorem 2.1](#) becomes

$$\text{tr} [2 H_1 H_1^T \mathcal{D}_s \{H_1 \Psi H_1^T\} (m - (p \wedge m) - 1) H_1 L^{-1} \Psi H_1^T] \geq 0.$$

Applying Identity (A.4.12) in [Lemma A.4.2](#) and using the fact that $H_1^T H_1 = I_{p \wedge m}$ and $H_1^T (I_p - H_1 H_1^T) = 0_{(p \wedge m) \times p}$, this can be written as

$$\text{tr} [2 \Psi^{(1)} + (m - (p \wedge m) - 1) L^{-1} \Psi] \geq 0 \quad (3.5)$$

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 (3.2).

As for Condition (2.17) in [Theorem 2.1](#), note that T^* in (2.18) is expressed as

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

Applying Identity (A.4.12) and also using the facts that $H_1 H_1^T = I_{p \wedge m}$ and $H_1^T (I_p - H_1 H_1^T) = 0_{(p \wedge m) \times (p \wedge m)}$, it follows that

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

where $\Phi \in \mathcal{D}_{p \wedge m}$ whose generic term is given in (3.4). Similarly to Condition (2.16), we apply Identity (A.4.12) with $G = T^*$ to obtain

$$\text{tr} [2 S^+ S \mathcal{D}_s \{S S^+ T^*\} + (m - (p \wedge m) - 1) S^+ T^*] = \text{tr} [2 \Phi^{(1)} + (m - (p \wedge m) - 1) L^{-1} \Phi], \quad (3.6)$$

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 (3.5) and (3.6). \square

3.1 An extension of on example of Haff

The following example is an extension of the one given by Haff [10], who considered the Gaussian context and the case where S is invertible. With this extension, we also deal with the elliptical context and the two cases where S is invertible and non-invertible.

Proposition 3.4 Consider a density as in (1.6) satisfying (2.14) and let p and m satisfy

$$\frac{p+m-(p \wedge m)+2}{p+m+1} \leq \frac{c^2}{b^2} \leq \frac{2p+2m-5(p \wedge m)-3}{p+m+1}. \quad (3.7)$$

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

$$\hat{\Sigma}_{a_o, \Psi} = a_o H_1 \text{diag}(l_1 + \nu t(\nu), l_2 + \nu t(\nu), \dots, l_{p \wedge m} + \nu t(\nu)) H_1^T,$$

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

- (i) $(p+m-2(p \wedge m)+1)t(\nu) + 2\nu t'(\nu) \geq 0$
- (ii) $0 \leq t(\nu) \leq \frac{2(p+m-2(p \wedge m)-1) \left((p+m+1)\frac{c^2}{b^2} - p - m + (p \wedge m) - 2 \right)}{(p+m-2(p \wedge m)+1)(p+m-2(p \wedge m)+3)}$
- (iii) $\left\{ 2(p+m-4(p \wedge m)+3)t(\nu) + 2\nu t'(\nu) + \left[2p+2m-5(p \wedge m)+5 - (p+m+1)\frac{c^2}{b^2} \right] \right\} t'(\nu) + 2 \{ t(\nu) + (p \wedge m)^2 \} \nu t''(\nu) \leq 0.$

Note that, when the function $t(\cdot)$ is constant, this is the estimator considered by Konno [15] in the Gaussian framework. In this setting, $c = b = 1$ and Condition (ii) suffices for $\hat{\Sigma}_{a_o, \Psi}$ to improve over $\hat{\Sigma}_{a_o}$.

Proof

The proof consists in proving Conditions (3.2) and (3.3) for $\Psi = \nu t(\nu) I_{p \wedge m}$, where $\nu = 1/\text{tr}(S^+)$ and $t(\cdot)$ is a twice differentiable non-increasing convex function. Note that Condition (3.7) holds if and only if $p+m-(p \wedge m)+2 \leq 2p+2m-5(p \wedge m)-3$ which is equivalent to

$$p+m \geq 4(p \wedge m) + 5.$$

Let $\rho_k = \nu^k \text{tr}(S^{+k})$. In this setting, Condition (3.2) is equivalent to

$$\sum_i^{p \wedge m} \left\{ (p+m-2(p \wedge m)-1) \frac{\nu t(\nu)}{l_i} + 2 \frac{\partial \{ \nu t(\nu) \}}{\partial l_i} \right\} \geq 0. \quad (3.8)$$

Using the fact that, for $r, k = 0, 1, 2, \dots$,

$$\frac{\partial}{\partial l_i} \left\{ \frac{\nu^k}{l_i^r} \right\} = k \frac{\nu^{k+1}}{l_i^{r+2}} - r \frac{\nu^k}{l_i^{r+1}} \quad \text{and} \quad \frac{\partial}{\partial l_i} \{ t(\nu) \} = \frac{\nu^2}{l_i^2} t'(\nu). \quad (3.9)$$

Condition (3.8) becomes

$$(p+m-2(p \wedge m)-1) t(\nu) + 2\rho_2 t(\nu) + 2\rho_2 \nu t'(\nu) \geq 0. \quad (3.10)$$

Thanks to the second Inequality in (A.2.3), this condition is satisfied as soon as Condition (i) holds.

Dealing with Condition (3.3), we also use (3.9) to express (3.4) as

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

Then, it follows that

$$\begin{aligned} \sum_i^{p \wedge m} \frac{\phi_i}{l_i} = & \{4 \rho_3 + (p+m-2(p \wedge m) - 1) \rho_2\} t^2(\nu) + 2 \{2 \rho_2 + (p+m-p \wedge m)\} t(\nu) \\ & + 4 \{\rho_3 t(\nu) + \rho_2\} \nu t'(\nu) \end{aligned} \quad (3.11)$$

and

$$\begin{aligned} \sum_i^{p \wedge m} \sum_{j \neq i}^{p \wedge m} \frac{\phi_i - \phi_j}{l_i - l_j} = & \left\{ 4 \nu^3 \sum_i^{p \wedge m} \sum_{j \neq i}^{p \wedge m} \frac{l_i^{-2} - l_j^{-2}}{l_i - l_j} + (p+m-2(p \wedge m) - 1) \nu^2 \sum_i^{p \wedge m} \sum_{j \neq i}^{p \wedge m} \frac{l_i^{-1} - l_j^{-1}}{l_i - l_j} \right\} t^2(\nu) \\ & + 4 \nu^2 \sum_i^{p \wedge m} \sum_{j \neq i}^{p \wedge m} \frac{l_i^{-1} - l_j^{-1}}{l_i - l_j} t(\nu) + 4 \left\{ \nu^3 t(\nu) \sum_i^{p \wedge m} \sum_{j \neq i}^{p \wedge m} \frac{l_i^{-2} - l_j^{-2}}{l_i - l_j} + \nu^2 \sum_i^{p \wedge m} \sum_{j \neq i}^{p \wedge m} \frac{l_i^{-1} - l_j^{-1}}{l_i - l_j} \right\} \nu t'(\nu). \end{aligned} \quad (3.12)$$

Moreover, using (3.9) repeatedly gives

$$\begin{aligned} \sum_i^{p \wedge m} \frac{\partial \phi_i}{\partial l_i} = & \{12 \rho_4 + 2(p+m-2(p \wedge m) - 5) \rho_3 - (p+m-2(p \wedge m) - 1) \rho_2\} t^2(\nu) \\ & + 2 \{4 \rho_3 + (p+m-(p \wedge m) - 2) \rho_2\} t(\nu) \\ & + 2 \{2 \rho_4 \nu t'(\nu) + 12 \rho_4 t(\nu) + (p+m-2(p \wedge m) - 5) \rho_3 t(\nu) \\ & \quad + 16 \rho_3 + (p+m-(p \wedge m) - 2) \rho_2\} \nu t'(\nu) \\ & + 4 \{\rho_4 t(\nu) + \rho_3\} \nu^2 t''(\nu). \end{aligned} \quad (3.13)$$

Substituting (3.11), (3.12), (3.13) and the left-hand side of (3.10) in the left-hand side of Condition (3.3), this last condition can be expressed as

$$H_1(z, u) + H_2(z, u) \leq 0 \quad (3.14)$$

where

$$\begin{aligned}
H_1(z, u) = & 2 \{12 \rho_4 + 2(p + m - 2(p \wedge m) - 5) \rho_3 - (p + m - 2(p \wedge m) - 1) \rho_2\} t^2(\nu) \\
& + (p + m - 2(p \wedge m) - 1) \{(p + m - 2(p \wedge m) - 1) \rho_2 + 4 \rho_3\} t^2(\nu) \\
& + 2 \left\{ (p + m - 2(p \wedge m) - 1) \left(p + m - (p \wedge m) + 2 \rho_2 - \frac{c^2}{b^2} (p + m + 1) \right) \right\} t(\nu) \\
& + 4 \left\{ 4 \rho_3 + (p + m - (p \wedge m) - 2) \rho_2 - \frac{c^2}{b^2} (p + m + 1) \rho_2 \right\} t(\nu) \\
& \nu^2 \left\{ 4 \nu \sum_i^{p \wedge m} \sum_{j \neq i}^{p \wedge m} \frac{l_i^{-2} - l_j^{-2}}{l_i - l_j} + (p + m + 2(p \wedge m) - 1) \sum_i^{p \wedge m} \sum_{j \neq i}^{p \wedge m} \frac{l_i^{-1} - l_j^{-1}}{l_i - l_j} \right\} t^2(\nu)
\end{aligned} \tag{3.15}$$

and

$$\begin{aligned}
H_2(z, u) = & \{48 \rho_4 t(\nu) + 8 \rho_4 \nu t'(\nu) + 8(p + m - 2(p \wedge m) - 3) \rho_3 t(\nu) + 32 \rho_3 \\
& + 4(2p + 2m - 3(p \wedge m) - 3) \rho_2 - 4 \frac{c^2}{b^2} (p + m + 1) \rho_2 \\
& + 4 \nu^3 t(\nu) \sum_i^{p \wedge m} \sum_{j \neq i}^{p \wedge m} \frac{l_i^{-2} - l_j^{-2}}{l_i - l_j} + 4 \nu^2 \sum_i^{p \wedge m} \sum_{j \neq i}^{p \wedge m} \frac{l_i^{-1} - l_j^{-1}}{l_i - l_j} \} \nu t'(\nu) \\
& + 8 \{ \rho_4 t(\nu) + \rho_3 \} \nu^2 t''(\nu).
\end{aligned} \tag{3.16}$$

Successive upper bounds will be developed for $H_1(z, u)$ in (3.15) and $H_2(z, u)$ in (3.16). First, it follows from the second inequalities in (A.2.5) and (A.2.6) in Lemma A.2.2 that the last term in the right-hand side of (3.15) is non-positive. Then, using the fact that $\rho_4 \leq \rho_3 \leq \rho_2$ (see Lemma A.2.1), we have

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

since $t(\cdot) \geq 0$. Therefore, it follows from the first inequality in (3.7) that the third term in the right-hand side of the last inequality is non-positive. Then, since $\rho_2 \leq 1$, an upper bound for (3.15) is given by

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

which is non-positive as soon as Condition (ii) holds.

As for the term $H_2(z, u)$ in (3.16), according to the first inequalities in (A.2.5) and (A.2.6) in Lemma A.2.2, the l_i 's terms between brackets satisfy

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

which implies that

$$\begin{aligned}
H_2(z, u) \leq & \{48 \rho_4 t(\nu) + 8 \rho_4 \nu t'(\nu) + 8(p + m - 4(p \wedge m) - 3) \rho_3 t(\nu) + 32 \rho_3 \\
& + 4 \left(2p + 2m - 5(p \wedge m) - 3 - (p + m + 1) \frac{c^2}{b^2} \right) \rho_2 \} \nu t'(\nu) \\
& + 8 \{ \rho_4 t(\nu) + \rho_3 \} \nu^2 t''(\nu),
\end{aligned} \tag{3.17}$$

since $t'(\cdot) \leq 0$. Next, according to the second inequality in (3.7) and thanks to Lemma A.2.1, it is clear that

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

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

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

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

$$\begin{aligned}
H_2(z, u) \leq & 4 \rho_4 \left\{ 2(p + m - 4(p \wedge m) + 3) t(\nu) + 2 \nu t'(\nu) \right. \\
& + \left. \left(2p + 2m - 5(p \wedge m) + 5 - (p + m + 1) \frac{c^2}{b^2} \right) \right\} \nu t'(\nu) \\
& + 8 \rho_4 \left\{ t(\nu) + \frac{\rho_3}{\rho_4} \right\} \nu^2 t''(\nu).
\end{aligned} \tag{3.18}$$

Now, it follows from the Cauchy-Schwarz inequality that

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

where for the second inequality we apply (A.2.3) in Lemma A.2.1 for $k = 2$. Therefore, we have

$$\frac{\rho_3}{\rho_4} = \frac{\text{tr}(S^+) \text{tr}^2(S^{+3})}{\text{tr}(S^{+4}) \text{tr}(S^{+3})} \leq \frac{1}{\rho_3} \leq (p \wedge m)^2 \tag{3.19}$$

according again to (A.2.3) in Lemma A.2.1. Therefore, thanks to (3.19) and using the fact that $t''(\cdot) \geq 0$, the upper bound for $H_2(z, u)$ in (3.18) is non-positive as soon as Condition (iii) in Proposition 3.4 holds. Since we have $H_1(z, u) \leq 0$, Condition (3.14) is satisfied, and hence, Condition (3.3) 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. \square

As an example illustrating [Proposition 3.4](#), let $t(\nu) = \alpha/(\beta + \nu)$, where $\alpha > 0$ and $\beta > 0$. Clearly $t'(\nu) = -1/(\beta + \nu)t(\nu) = -\alpha/(\beta + \nu)^2$ and $t''(\nu) = 2/(\beta + \nu)^2 t(\nu) = 2\alpha/(\beta + \nu)^3$. Therefore, Condition (i) in [Proposition 3.4](#) becomes

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

which is satisfied as soon as $p + m - 2(p \wedge m) \geq 1$ since

$$\frac{\nu}{\beta + \nu} \leq 1. \tag{3.20}$$

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

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

Now, Condition (iii) is equivalent to

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

According to (3.20), we have

$$\begin{aligned} & \left\{ -2(p + m - 4(p \wedge m))t(\nu) + 4(p \wedge m)^2 \right. \\ & \left. - \left(2p + 2m - 5(p \wedge m) + 5 - (p + m + 1)\frac{c^2}{b^2} \right) \right\} \frac{1}{\beta + \nu} t(\nu) \leq 0. \end{aligned}$$

Now, using the fact that $t(\nu) \geq 0$, the last condition is satisfied as soon as

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

which holds as soon as

$$\frac{\alpha}{\beta} \leq t_2(p, m),$$

since $t(\nu) \leq t(0) = \alpha/\beta$.

Finally, choosing $\alpha/\beta \leq t_1(p, m) \wedge t_2(p, m)$ implies that Conditions (ii) and (iii) in [Proposition 3.4](#) are both satisfied.

As mentioned by Haff [10], even in the Gaussian framework and in the non-singular case, calculations under the quadratic loss in (1.7) 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 $t(\cdot)$ functions than in Haff [10] and Konno [15].

4 Concluding remarks

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 aS (where S is the simple covariance matrix) 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 [12] and Konno [15].

As a natural extension of this work is to determine examples of improved estimators which are not necessarily orthogonally invariant. We have already obtained such examples when S is invertible. We hope to address to tackle this question when S is singular.

A Appendix

A.1 A subclass of elliptically symmetric distribution

The subclass of densities in (1.6) satisfying Condition (2.14) contains distributions such the multivariate normal and the logistic type distribution of the form

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

where $\beta > 0$ and $\gamma > 0$. Note that, for the logistic type distribution $c = 1/2\beta$ and $b = (1 + e^{-\gamma})/2\beta$. For 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 general elliptical setting, these expectations are related since, for any integrable function $H(Z, U)$, we have

$$K^* E_{\theta, \Sigma}^*[H(Z, U)] = E_{\theta, \Sigma}[\varphi_{\theta, \Sigma}^*(Z, U) H(Z, U)] \quad (\text{A.1.1})$$

and

$$\begin{aligned} K^{**} E_{\theta, \Sigma}^{**}[H(Z, U)] &= K^* E_{\theta, \Sigma}^*[\varphi_{\theta, \Sigma}^{**}(Z, U) H(Z, U)] \\ &= E_{\theta, \Sigma}[\varphi_{\theta, \Sigma}^{**}(Z, U) \varphi_{\theta, \Sigma}^*(Z, U) H(Z, U)] \end{aligned} \quad (\text{A.1.2})$$

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

$$\varphi_{\theta, \Sigma}^*(z, u) = \frac{F^*(v)}{f(v)} \quad \text{and} \quad \varphi_{\theta, \Sigma}^{**}(z, u) = \frac{F^{**}(v)}{F^*(v)}, \quad (\text{A.1.3})$$

with

$$v = \text{tr}\{(z - \theta) \Sigma^{-1} (z - \theta)^T\} + \text{tr}\{\Sigma^{-1} u^T u\}.$$

The following lemma will be helpful to deal with the dependence of the risk difference in (A.3.25) on the unknown parameters θ and Σ in $\varphi_{\theta, \Sigma}^*(z, u)$ and $\varphi_{\theta, \Sigma}^{**}(z, u)$.

Lemma A.1.1 *In the context of (A.1.3), assume that there exist two positives constant c and b such that for all $t \in \mathbb{R}^+$*

$$c \leq \frac{F^*(t)}{f(t)} \leq b. \quad (\text{A.1.4})$$

Then we have

$$\frac{1}{b^2} \leq \frac{1}{K^* \varphi_{\theta, \Sigma}^{**}(z, u)} \leq \frac{1}{c^2}, \quad (\text{A.1.5})$$

where K^* is given in (2.6).

Proof By definition of $F^{**}(\cdot)$ and $F^*(\cdot)$ in (2.5) and (2.9), we have for all $t \in \mathbb{R}^+$

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

where E_t is the expectation with respect to the density which is proportional to $F^*(x) \mathbb{1}_{[t, \infty[}(x)$. Hence, according to (A.1.3) and (A.1.4) with $t = v$, we have

$$\frac{1}{b} \leq \frac{1}{\varphi_{\theta, \Sigma}^{**}(z, u)} = E_v \left[\frac{f(X)}{F^*(X)} \right] \leq \frac{1}{c}. \quad (\text{A.1.6})$$

Moreover, setting $H(Z, U) \equiv 1$ in (A.1.1), we have, according to (A.1.3) and by assumption (A.1.4), $c \leq K^* \leq b$. Therefore $1/b \leq K^* \leq 1/c$ and (A.1.6) gives (A.1.5). \square

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 (2.7) and to the trace of S and S^+ .

Proposition A.2.1 *Let A , B and C be $p \times p$ matrix functions of S . Assuming that all partial derivatives and products exist as needed, we have*

$$\mathcal{D}_s \{AB\} = \mathcal{D}_s \{A\} B + (A^T \mathcal{D}_s)^T \{B\} \quad (\text{A.2.1})$$

$$\text{tr}[C (A^T \mathcal{D}_s)^T \{B\}] = \text{tr}[A^T (C \mathcal{D}_s)^T \{B\}^T] \quad (\text{A.2.2})$$

Proof The first identity was given by Haff [10, 9], so that we only give the proof of (A.2.2). We have

$$\text{tr}[C (A^T \mathcal{D}_s)^T \{B\}] = \sum_{i,j,k} C_{ij} (A^T \mathcal{D}_s)_{jk}^T B_{ki} = \sum_{i,j,k} C_{ij} (A^T \mathcal{D}_s)_{kj} B_{ik}^T.$$

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

$$\text{tr}[C (A^T \mathcal{D}_s)^T \{B\}] = \sum_{i,j,k,l} A_{kl}^T C_{ij} d_{jl} B_{ik}^T = \sum_{k,l,i} A_{kl}^T (C \mathcal{D}_s)_{li}^T B_{ik}^T = \text{tr}[A^T (C \mathcal{D}_s)^T \{B\}^T].$$

\square

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

Lemma A.2.1 Let $\rho_k = \text{tr}(S^{+k})/\text{tr}^k(S^+)$. Then, for $k = 2, 3, 4, \dots$,

$$(p \wedge m)^{1-k} \leq \rho_k \leq \rho_{k-1} \leq 1. \quad (\text{A.2.3})$$

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

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

Hence

$$\rho_k = \frac{\text{tr}(S^{+k})}{\text{tr}^k(S^+)} \leq \frac{\text{tr}(S^{+(k-1)})}{\text{tr}^{k-1}(S^+)} = \rho_{k-1}. \quad (\text{A.2.4})$$

Secondly, setting $k = 2$ in (A.2.4) gives $\rho_2 \leq 1$. Now, from the Hölder inequality, it follow that, for $k = 2, 3, 4, \dots$,

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

Then

$$\text{tr}^k(L^{-1}) \leq \text{tr}(L^{-k}) (p \wedge m)^{k-1},$$

and hence

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

Lemma A.2.2 The following inequalities hold:

$$-4(p \wedge m) \text{tr}(S^{+3}) < \sum_{i=1}^{p \wedge m} \sum_{j \neq i}^{p \wedge m} \frac{l_i^{-2} - l_j^{-2}}{l_i - l_j} < 0, \quad (\text{A.2.5})$$

and

$$-2(p \wedge m) \text{tr}(S^{+2}) < \sum_{i=1}^{p \wedge m} \sum_{j \neq i}^{p \wedge m} \frac{l_i^{-1} - l_j^{-1}}{l_i - l_j} < 0. \quad (\text{A.2.6})$$

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

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

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

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

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

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

which is (A.2.5).

Similarly, dealing with (A.2.6), we have

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

Hence, using the fact that $\sum_{i=1}^{p \wedge m} \sum_{j \leq i}^{p \wedge m} l_j^{-2} > 0$, it follows that

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

which is (A.2.6). □

A.3 Risk calculations and Stein-Haff type lemmas

In this appendix, first, we focused on risk calculations. Secondly following Fourdrinier, Mezoued and Wells [8], we give, in Lemma 2.1, a new version of the Stein-Haff type identity through the Haff operator \mathcal{D}_s in (2.7).

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

$$\operatorname{tr}^2(A^T B) \leq \|A\|_F^2 \|B\|_F^2. \quad (\text{A.3.1})$$

First assume that

$$E_{\theta, \Sigma} [\|\Sigma^{-1} S\|_F^2] < \infty. \quad (\text{A.3.2})$$

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

$$L(\Sigma, a S) = \operatorname{tr}(a S \Sigma^{-1} - I_p)^2 = a^2 \operatorname{tr}(\Sigma^{-1} S)^2 - 2a \operatorname{tr}(\Sigma^{-1} S) + p. \quad (\text{A.3.3})$$

Then, noticing that $\operatorname{tr}^2(\Sigma^{-1} S) = \operatorname{tr}^2(\Sigma^{-1/2} S \Sigma^{-1/2})$, applying (A.3.1) with $A = I_p$ and $B = \Sigma^{-1/2} S \Sigma^{-1/2}$ gives

$$\operatorname{tr}^2(\Sigma^{-1} S) \leq p \|\Sigma^{-1/2} S \Sigma^{-1/2}\|_F^2 = p \operatorname{tr}(\Sigma^{-1} S)^2. \quad (\text{A.3.4})$$

Now, with $A^T = B = \Sigma^{-1} S$, applying (A.3.1) gives

$$\operatorname{tr}(\Sigma^{-1} S)^2 \leq \operatorname{tr}((\Sigma^{-1} S)^T \Sigma^{-1} S) = \|\Sigma^{-1} S\|_F^2. \quad (\text{A.3.5})$$

Taking expectation in (A.3.3), it follows from (A.3.5) that, according to assumption (A.3.2), we have $E_{\theta,\Sigma} [\text{tr}(\Sigma^{-1}S)^2] < \infty$, and hence according to (A.3.4), $E_{\theta,\Sigma} [\text{tr}^2(\Sigma^{-1}S)] < \infty$, which implies that $E_{\theta,\Sigma} [\text{tr}(\Sigma^{-1}S)] < \infty$. This is the announced result.

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

$$E_{\theta,\Sigma} [\|\Sigma^{-1}SS^+G\|_F^2] < \infty. \quad (\text{A.3.6})$$

We will show that Conditions (A.3.2) and (A.3.6) insure that the risk of $\hat{\Sigma}_{a,G}$ is finite. Note that the loss of $\hat{\Sigma}_{a,G}$ is

$$L(\Sigma, \hat{\Sigma}_{a,G}) = L(\Sigma, aS) + a^2 \text{tr}(\Sigma^{-1}SS^+G)^2 + 2a \text{tr}[\Sigma^{-1}SS^+G(aS\Sigma^{-1} - I_p)]. \quad (\text{A.3.7})$$

As in (A.3.5), through the Cauchy-Schwarz inequality, we have

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

and hence,

$$E_{\theta,\Sigma} [\text{tr}(\Sigma^{-1}SS^+G)^2] \leq E_{\theta,\Sigma} [\|\Sigma^{-1}SS^+G\|_F^2] < \infty \quad (\text{A.3.8})$$

by assumption (A.3.6). Also by the Cauchy-Schwarz inequality associated to the inner product $\langle A, B \rangle = E_{\theta,\Sigma} [\text{tr}(A^T B)]$ it follows that

$$|E_{\theta,\Sigma} \{ \text{tr}[\Sigma^{-1}SS^+G(aS\Sigma^{-1} - I_p)] \}| \leq (E_{\theta,\Sigma} [\|\Sigma^{-1}SS^+G\|_F^2] E_{\theta,\Sigma} [\|aS\Sigma^{-1} - I_p\|_F^2])^{1/2}. \quad (\text{A.3.9})$$

By assumption (A.3.8), the first term in the right-hand side of (A.3.9) 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.3.7), (A.3.2) and (A.3.6) are sufficient conditions for the risk $\hat{\Sigma}_{a,G}$ to be finite.

Lastly, under Conditions (A.3.2) and (A.3.6), the risk difference $\Delta(G)$ in (2.2) between $\hat{\Sigma}_{a,G}$ and $\hat{\Sigma}_a$ is finite and can be expressed as

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

□

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

Lemma A.3.1 *For any $p \times p$ matrix function $G(z, s)$ weakly differentiable with respect to s , for any z , we have*

$$\text{tr}(U^T \nabla_U \{G^T S^+\}) = \text{tr}[2SS^+ \mathcal{D}_s \{SS^+ G\}^T - ((p \wedge m) + 1) S^+ G]. \quad (\text{A.3.10})$$

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

$$\text{tr}(U^T \nabla_U \{G^T S^+\}) = \sum_{i,j,l}^p \sum_k^m U_{ik}^T \frac{\partial}{\partial U_{kj}} \{(G^T SS^+)_{jl} S_{li}^+\} = R_1 + R_2$$

where

$$R_1 = \sum_{i,j,l}^p \sum_k^m S_{li}^+ U_{ik}^T \frac{\partial (G^T SS^+)_{jl}}{\partial U_{kj}} = \sum_{j,l}^p \sum_k^m (S^+ U^T)_{lk} \frac{\partial (G^T SS^+)_{jl}}{\partial U_{kj}} \quad (\text{A.3.11})$$

and

$$R_2 = \sum_{i,j,l}^p \sum_k^m U_{ik}^T (G^T SS^+)_{jl} \frac{\partial S_{li}^+}{\partial U_{kj}} = \sum_j^p \sum_k^m \left(G^T SS^+ \frac{\partial S^+}{\partial U_{kj}} U^T \right)_{jk}. \quad (\text{A.3.12})$$

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

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

Note also that

$$\frac{\partial S_{qr}}{\partial U_{kj}} = \frac{\partial}{\partial U_{kj}} \sum_{o=1}^m U_{qo}^T U_{or} = \sum_{o=1}^p \left(\frac{\partial U_{oq}}{\partial U_{kj}} U_{or} + U_{oq} \frac{\partial U_{or}}{\partial U_{kj}} \right) = U_{kr} \delta_{qj} + U_{kq} \delta_{rj}. \quad (\text{A.3.13})$$

Then

$$\begin{aligned} \frac{\partial (G^T SS^+)_{jl}}{\partial U_{kj}} &= \left(\sum_{r \leq j}^p U_{kr} \frac{\partial}{\partial S_{rj}} + \sum_{q \geq j}^p U_{kq} \frac{\partial}{\partial S_{qj}} \right) (G^T SS^+)_{jl} \\ &= \left(\sum_{r \leq j}^p U_{kr} \frac{\partial}{\partial S_{rj}} + \sum_{r \geq j}^p U_{kr} \frac{\partial}{\partial S_{rj}} \right) (G^T SS^+)_{jl} \\ &= \left(\sum_r^p U_{kr} \frac{\partial}{\partial S_{rj}} + U_{kj} \frac{\partial}{\partial S_{jj}} \right) (G^T SS^+)_{jl} \\ &= 2 \sum_r^p U_{kr} \frac{1}{2} (1 + \delta_{rj}) \frac{\partial}{\partial S_{rj}} (G^T SS^+)_{jl} \\ &= 2 \sum_r^p U_{kr} d_{rj}^s (G^T SS^+)_{jl}, \end{aligned}$$

by definition of the generic term d_{rj}^s of \mathcal{D}_s . Hence (A.3.11) becomes

$$R_1 = 2 \sum_{l,j}^p \sum_k^m (S^+ U^T)_{lk} \sum_r^p U_{kr} d_{rj}^s (G^T SS^+)_{jl} = 2 \text{tr}(SS^+ \mathcal{D}_s \{SS^+ G^T\}). \quad (\text{A.3.14})$$

As for R_2 in (A.3.12), using Proposition A.4.1 with $A = G^T S S^+$ and $B = U^T$, we have

$$\begin{aligned} R_2 &= \sum_j^p \sum_k^m ((G^T S S^+)(I - S S^+))_{jj} (U S S^+ U^T)_{kk} + \sum_j^p \sum_k^m ((I - S S^+) U^T)_{jk} (U (G^T S^+ S^+)^T)_{kj} \\ &\quad - \sum_j^p \sum_k^m ((G^T S^+)_{jj} (U S^+ U^T)_{kk} + (S^+ U^T)_{jk} (U S^+ G)_{kj}). \end{aligned}$$

Since $S S^+ (I_p - S S^+) = S (I_p - S S^+) = 0_{p \times p}$, $(I_p - S S^+) U^T = 0_{p \times (p \wedge m)}$, $\text{tr}(S S^+) = p \wedge m$, $S = U^T U$ and $S^+ = S^+ S S^+$. Thus R_2 simplifies in

$$R_2 = -(p \wedge m + 1) \text{tr}(S^+ G). \quad (\text{A.3.15})$$

Combining (A.3.14) and (A.3.15) gives the desired result. \square

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 [8] as follows:

$$E_{\theta, \Sigma} [\text{tr}(\Sigma^{-1} S H)] = K^* E_{\theta, \Sigma}^* [m \text{tr}(H) + \text{tr}(U^T \nabla_U H^T)].$$

These authors remark that an equivalent expression of this identity is

$$E_{\theta, \Sigma} [\text{tr}(\Sigma^{-1} S H)] = K^* E_{\theta, \Sigma}^* [m \text{tr}(S S^+ H) + \text{tr}(U^T \nabla_U H^T S S^+)],$$

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

$$E_{\theta, \Sigma} [\text{tr}(\Sigma^{-1} S S^+ G)] = K^* E_{\theta, \Sigma}^* [m \text{tr}(S^+ G) + \text{tr}(U^T \nabla_U G^T S^+)]. \quad (\text{A.3.16})$$

Therefore, we deduce the Stein-Haff type identity given in (2.8) through the Haff operator \mathcal{D}_s by replacing (A.3.10) in (A.3.16). \square

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

$$\begin{aligned} E_{\theta, \Sigma} [\text{tr}(\Sigma^{-1} S S^+ V \Sigma^{-1} S S^+ G)] &= K^* E_{\theta, \Sigma}^* [\text{tr}[2 S S^+ \mathcal{D}_s \{G^T S S^+ \Sigma^{-1} V^T S S^+\} \\ &\quad + (m - (p \wedge m) - 1) \Sigma^{-1} S S^+ G S^+ V]]. \end{aligned} \quad (\text{A.3.17})$$

Secondly, according to (A.2.1) in Proposition A.2.1, the first integrand term in the right-hand side of (A.3.17) can be rewritten as

$$\begin{aligned} \text{tr}[S S^+ \mathcal{D}_s \{G^T S S^+ \Sigma^{-1} V^T S S^+\}] &= \text{tr}[\Sigma^{-1} V^T S S^+ \mathcal{D}_s \{G^T S S^+\}] \\ &\quad + \text{tr}[S S^+ (\Sigma^{-1} S S^+ G \mathcal{D}_s)^T \{V^T S S^+\}]. \end{aligned} \quad (\text{A.3.18})$$

Now, applying (A.2.2) in Proposition A.2.1 to the second term in the right-hand side of (A.3.18) gives

$$\text{tr}[S S^+ (\Sigma^{-1} S S^+ G \mathcal{D}_s)^T \{V^T S S^+\}] = \text{tr}[\Sigma^{-1} S S^+ G (S S^+ \mathcal{D}_s)^T \{S S^+ V\}].$$

Furthermore, using the fact that $S S^+ = S S^+ S S^+$, it is clear that from (A.2.1) in Proposition A.2.1 that

$$(S S^+ \mathcal{D}_s)^T \{S S^+ V\} = \mathcal{D}_s \{S S^+ V\} - \mathcal{D}_s \{S S^+\} S S^+ V.$$

Hence, according to (A.4.14) and using the fact that $(I_p - SS^+)S = 0_{p \times p}$ we have

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

Therefore

$$\text{tr} \left[SS^+ (\Sigma^{-1} SS^+ G \mathcal{D}_s)^T \{V^T S^+ S\} \right] = \text{tr} \left[\Sigma^{-1} SS^+ G \left(\mathcal{D}_s \{SS^+ V\} - \frac{1}{2}(p - p \wedge m)S^+ V \right) \right]. \quad (\text{A.3.19})$$

Now, combining (A.3.17), (A.3.18) and (A.3.19) and using the symmetry of $SS^+ V$ gives

$$E_{\theta, \Sigma} [\text{tr} (\Sigma^{-1} SS^+ V \Sigma^{-1} SS^+ G)] = K^* E_{\theta, \Sigma}^* [\text{tr} (\Sigma^{-1} SS^+ T^*)], \quad (\text{A.3.20})$$

where

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

Finally, applying Lemma 2.1 to (A.3.20) with T^* instead of G gives

$$E_{\theta, \Sigma}^* [\text{tr} (\Sigma^{-1} SS^+ T^*)] = K^{**} E_{\theta, \Sigma}^{**} [\text{tr} (2S^+ S \mathcal{D}_s \{SS^+ T^*\}^T - (m - (p \wedge m) - 1)S^+ T^*)].$$

□

Proof of Proposition 2.3 Let $\hat{\Sigma}_a = aS$, where a is a positive constant. Then, according to (1.8), the risk at Σ of $\hat{\Sigma}_a$ is

$$\begin{aligned} R(\Sigma, \hat{\Sigma}_a) &= E_{\theta, \Sigma} [\text{tr} (aS \Sigma^{-1} - I_p)^2] \\ &= a^2 E_{\theta, \Sigma} [\text{tr} (\Sigma^{-1} SS^+ S)^2] - 2a E_{\theta, \Sigma} [\text{tr} (\Sigma^{-1} SS^+ S)] + p, \end{aligned} \quad (\text{A.3.21})$$

since $S = SS^+ S$. Setting $G = V = S$ in Corollary 2.1 to deal with the first term of the right-hand side of (A.3.21), and in Lemma 2.1 to deal with the second term, we have

$$\begin{aligned} R(\Sigma, \hat{\Sigma}_a) &= a^2 K^* K^{**} E_{\theta, \Sigma}^{**} [\text{tr} [2S^+ S \mathcal{D}_s \{SS^+ T^*\}^T + (m - (p \wedge m) - 1)S^+ T^*]] \\ &\quad - 2a K^* E_{\theta, \Sigma}^* [\text{tr} [2S^+ S \mathcal{D}_s \{S\} + (m - (p \wedge m) - 1)S^+ S]] + p, \end{aligned} \quad (\text{A.3.22})$$

where

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

Furthermore, thanks to (A.4.13) and (A.4.14) and using the fact that $S(I_p - SS^+) = 0_{p \times p}$, the integrand terms in the right-hand side of (A.3.22) are evaluated as

$$\text{tr} [2S^+ S \mathcal{D}_s \{SS^+ T^*\}^T + (m - (p \wedge m) - 1)S^+ T^*] = (p + m + 1)(p + m - p \wedge m) \text{tr}(SS^+) \quad (\text{A.3.23})$$

and

$$\text{tr} (2S^+ S \mathcal{D}_s \{S\} + (m - (p \wedge m) - 1)S^+ S) = (p + m - p \wedge m) \text{tr}(S^+ S) \quad (\text{A.3.24})$$

Substituting (A.3.23) and (A.3.24) in (A.3.22) gives

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

since $\text{tr}(S^+ S) = p \wedge m$. Clearly, choosing $a = 1/K^{**}(p + m + 1)$ is optimal under the risk (1.8). \square

Proof of Theorem 2.1 Expressing the risk difference in (2.12) only through the $E_{\theta, \Sigma}$ -expectation, thanks to (A.1.1) and (A.1.2), we have

$$\begin{aligned} \Delta(G) &= a_o^2 K^* E_{\theta, \Sigma} \left[\varphi_{\theta, \Sigma}^*(Z, U) \varphi_{\theta, \Sigma}^{**}(Z, U) \left(\text{tr} \left[2 S^+ S \mathcal{D}_s \{ S S^+ T^* \}^T - (m - (p \wedge m) - 1) S^+ T^* \right] \right. \right. \\ &\quad \left. \left. - \frac{2}{a_o K^* \varphi_{\theta, \Sigma}^{**}(Z, U)} \text{tr} \left[2 S^+ S \mathcal{D}_s \{ S S^+ G \} - (m - (p \wedge m) - 1) S^+ G \right] \right) \right], \end{aligned} \quad (\text{A.3.25})$$

where a_o is given in (2.15). Using Inequality (A.1.5) in Lemma A.1.1 and also noting that $K^{**} \geq c^2$, we have $K^{**}/K^* \varphi_{\theta, \Sigma}^*(Z, U) \geq c^2/b^2$. Therefore, under Condition (2.16),

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

Then, the estimator $\hat{\Sigma}_{a_o, G}$ dominates $\hat{\Sigma}_{a_o}$ as soon as Inequality (2.17) holds, since $\varphi_{\theta, \Sigma}^*(Z, U)$ and $\varphi_{\theta, \Sigma}^{**}(Z, U)$ are non-negative functions. \square

A.4 Differential expressions

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

Proposition A.4.1 *Recall that $S = U^T U$ 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*

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

Proof Applying a well known result about the derivatives of the Moore-Penrose inverses S^+ (see e.g. Harville [13]), we have for any $1 \leq k \leq m$ and any $1 \leq j \leq p$,

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

and then

$$\begin{aligned} \left(A \frac{\partial S^+}{\partial U_{kj}} B \right)_{li} &= - \left(A S^+ \frac{\partial S}{\partial U_{kj}} S^+ B \right)_{li} + \left(A (I_p - S S^+) \frac{\partial S}{\partial U_{kj}} S^+ S^+ B \right)_{li} \\ &\quad + \left(A S^+ S^+ \frac{\partial S}{\partial U_{kj}} (I_p - S S^+) B \right)_{li}. \end{aligned} \quad (\text{A.4.1})$$

Thanks to (A.3.13), the first term in the right-hand side of (A.4.1) becomes

$$\begin{aligned}
\left(A S^+ \frac{\partial S}{\partial U_{kj}} S^+ B \right)_{li} &= \sum_{q,r}^p (A S^+)_{lq} \left(\frac{\partial S}{\partial U_{kj}} \right)_{qr} (S^+ B)_{ri} \\
&= \sum_r^m (A S^+)_{lj} (S^+ B)_{ri} U_{kr} + \sum_{q=1}^m (A S^+)_{lq} (S^+ B)_{ji} U_{kq} \\
&= (A S^+)_{lj} (U S^+ B)_{ki} + (S^+ B)_{ji} (U S^+ A^T)_{kl}. \tag{A.4.2}
\end{aligned}$$

Similarly, the two other terms in the right-hand side of (A.4.1) become respectively

$$\begin{aligned}
\left(A (I_p - S S^+) \frac{\partial S}{\partial U_{kj}} S^+ S^+ B \right)_{li} &= (A (I_p - S S^+))_{lj} (U S^+ S^+ B)_{ki} \\
&\quad + (S^+ S^+ B)_{ji} (U (I_p - S S^+) A^T)_{kl} \tag{A.4.3}
\end{aligned}$$

and

$$\begin{aligned}
\left(A S^+ S^+ \frac{\partial S}{\partial U_{kj}} (I_p - S S^+) B \right)_{li} &= (A S^+ S^+)_{lj} (U (I_p - S S^+) B)_{ki} \\
&\quad + ((I_p - S S^+) B)_{ji} (U S^+ S^+ A^T)_{kl}. \tag{A.4.4}
\end{aligned}$$

Hence, replacing (A.4.2), (A.4.3) and (A.4.4) in (A.4.1) and using the fact that $U(I_p - S S^+) = 0_{(p \wedge m) \times p}$, give the desired result. \square

The following lemma is useful to evaluate the derivatives of eigenvalues and eigenvectors of S .

Lemma A.4.1 *Let $1 \leq i, j, a, r \leq p$ and $1 \leq k \leq p \wedge m$. Then*

$$d_{ij}^s l_k = (H_1)_{ik} (H_1)_{jk} \tag{A.4.5}$$

and

$$\begin{aligned}
(d_{ij}^s H_1)_{ak} &= \frac{1}{2} \sum_{r \neq k}^{p \wedge m} \frac{(H_1)_{ar}}{l_k - l_r} [(H_1)_{ir} (H_1)_{jk} + (H_1)_{jr} (H_1)_{ik}] \\
&\quad + \frac{1}{2 l_k} [(I_p - H_1 H_1^T)_{ai} (H_1)_{jk} + (I_p - H_1 H_1^T)_{aj} (H_1)_{ik}]. \tag{A.4.6}
\end{aligned}$$

Proof Recall that $S = H_1 L H_1^T$ 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 $H_2^T H_1 = 0_{(p - p \wedge m) \times (p \wedge m)}$ 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 completes H_1 .

For any differential operator of S , we have

$$dS = (dH_1) L H_1^T + H_1 (dL) H_1^T + H_1 L (dH_1^T),$$

which yields

$$H^T (dS) H_1 = \begin{bmatrix} H_1^T (dS) H_1 \\ H_2^T (dS) H_1 \end{bmatrix} = \begin{bmatrix} H_1^T (dH_1) L + H_1^T H_1 (dL) + H_1^T H_1 L (dH_1^T) H_1 \\ H_2^T (dH_1) L + H_2^T H_1 (dL) + H_2^T H_1 L (dH_1^T) H_1 \end{bmatrix}.$$

The differential expression of $H_1^T H_1 = I_{p \wedge m}$ gives that $(dH_1^T)H_1 = -H_1^T (dH_1)$. Then

$$\begin{bmatrix} H_1^T (dS) H_1 \\ H_2^T (dS) H_1 \end{bmatrix} = \begin{bmatrix} H_1^T (dH_1) L - L H_1^T (dH_1) + (dL) \\ H_2^T (dH_1) L \end{bmatrix}$$

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

$$\begin{aligned} (dL)_k &= \{H_1^T (dS) H_1\}_{kk} - \{H_1^T (dH_1) L\}_{kk} + \{L H_1^T (dH_1)\}_{kk} \\ &= \{H_1^T (dS) H_1\}_{kk}, \end{aligned} \quad (\text{A.4.7})$$

since $\{L H_1^T (dH_1)\}_{kk} - \{H_1^T (dH_1) L\}_{kk} = 0$.

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

$$\begin{aligned} \{H_1^T (dS) H_1\}_{rk} &= \{H_1^T (dH_1)\}_{rk} l_k - l_r \{H_1^T (dH_1)\}_{rk} \quad \text{for } r, k = 1, \dots, p \wedge m \\ \{H_2^T (dS) H_1\}_{rk} &= \{H_2^T (dH_1)\}_{rk} l_k \quad \text{for } r = p \wedge m + 1, \dots, p \text{ and } k = 1, \dots, p \wedge m. \end{aligned}$$

Therefore,

$$\begin{aligned} (H_1^T (dH_1))_{rk} &= \frac{1}{l_k - l_r} \{H_1^T (dS) H_1\}_{rk} \quad \text{for } r, k = 1, \dots, p \wedge m \\ (H_2^T (dH_1))_{rk} &= \frac{1}{l_k} \{H_2^T (dS) H_1\}_{rk} \quad \text{for } r = p \wedge m + 1, \dots, p \text{ and } k = 1, \dots, p \wedge m. \end{aligned} \quad (\text{A.4.8})$$

Now, if "d" is the Haff operator in (2.7), we have

$$(d_{ij}^s S)_{cd} = \frac{1}{2} (\delta_{ic} \delta_{jd} + \delta_{id} \delta_{jc}).$$

Then

$$\{H_1^T (d_{ij}^s S) H_1\}_{rk} = \frac{1}{2} [(H_1)_{ir} (H_1)_{jk} + (H_1)_{jr} (H_1)_{ik}] \quad (\text{A.4.9})$$

and

$$\{H_2^T (d_{ij}^s S) H_1\}_{rk} = \frac{1}{2} [(H_2)_{ir} (H_1)_{jk} + (H_2)_{jr} (H_1)_{ik}]. \quad (\text{A.4.10})$$

Hence, using (A.4.9) and (A.4.10), Equality (A.4.7) becomes

$$d_{ij}^s l_k = (H_1)_{ik} (H_1)_{jk},$$

which is (A.4.5).

Dealing with (A.4.6), note that, since $H H^T = I_p$

$$\begin{aligned} (d_{ij}^s H_1)_{ak} &= \{H H^T (d_{ij}^s H_1)\}_{ak} \\ &= \sum_{r \neq k}^{p \wedge m} (H_1)_{ar} \{H_1^T (d_{ij}^s H_1)\}_{rk} + \sum_{r=(p \wedge m)+1}^p (H_2)_{ar} \{H_2^T (d_{ij}^s H_1)\}_{rk}. \end{aligned} \quad (\text{A.4.11})$$

Combining (A.4.8), (A.4.9) and (A.4.10) we have, for $r \neq k$,

$$(H_1^T (d_{ij}^s H_1))_{rk} = \frac{(H_1)_{ir} (H_1)_{jk} + (H_1)_{jr} (H_1)_{ik}}{2(l_k - l_r)}$$

and

$$(H_2^T (d_{ij}^s H_1))_{rk} = \frac{(H_2)_{ir} (H_1)_{jk} + (H_2)_{jr} (H_1)_{ik}}{2l_k}.$$

Therefore (A.4.11) becomes

$$(d_{ij}^s H_1)_{ak} = \frac{1}{2} \sum_{r \neq k} \frac{(H_1)_{ar}}{l_k - l_r} [(H_1)_{ir} (H_1)_{jk} + (H_1)_{jr} (H_1)_{ik}] \\ + \frac{1}{2l_k} [(H_2 H_2^T)_{ai} (H_1)_{jk} + (H_2 H_2^T)_{aj} (H_1)_{ik}],$$

which gives the desired result in (A.4.6) since $H_2 H_2^T = (I_p - H_1 H_1^T)$. \square

The following lemma is a generalization of the results obtained by Haff [12] to both invertible and non-invertible cases.

Lemma A.4.2 *Under the notation of the proof of Lemma A.4.1, let $G = H_1 \Psi H_1^T$ where $\Psi \in \mathcal{D}_{p \wedge m}$. Then $SS^+ G = G$ and*

$$\mathcal{D}_s\{SS^+ G\} = \mathcal{D}_s\{H_1 \Psi H_1^T\} = H_1 \Psi^{(1)} H_1^T + \frac{1}{2} \text{tr}(L^{-1} \Psi) (I_p - H_1 H_1^T), \quad (\text{A.4.12})$$

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

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

Furthermore

$$\mathcal{D}_s\{S\} = \mathcal{D}_s\{H_1 L H_1^T\} = \frac{1}{2} (p+1) H_1 H_1^T + \frac{1}{2} (p \wedge m) (I_p - H_1 H_1^T), \quad (\text{A.4.13})$$

$$\mathcal{D}_s\{SS^+\} = \mathcal{D}_s\{H_1 H_1^T\} = \frac{1}{2} (p - p \wedge m) H_1 L^{-1} H_1^T + \frac{1}{2} \text{tr}(L^{-1}) (I_p - H_1 H_1^T) \quad (\text{A.4.14})$$

and

$$(SS^+ \mathcal{D}_s)^T \{SS^+ G\} = (H_1 H_1^T \mathcal{D}_s)^T \{H_1 \Psi H_1^T\} = \mathcal{D}_s\{H_1 \Psi H_1\} - \frac{1}{2} (p - m) (H_1 L^{-1} \Psi H_1^T). \quad (\text{A.4.15})$$

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

$$(\mathcal{D}_s\{H_1 \Psi H_1^T\})_{ia} = \sum_j^p \sum_b^{p \wedge m} d_{ij}^s \{(H_1)_{jb} \psi_b (H_1)_{ab}\} = A_{ia} + B_{ia} + C_{ia},$$

where

$$A_{ia} = \sum_j^p \sum_b^{p \wedge m} \psi_b (H_1)_{ab} (d_{ij}^s H_1)_{jb}, \quad (\text{A.4.16})$$

$$B_{ia} = \sum_j^p \sum_b^{p \wedge m} \psi_b (H_1)_{jb} (d_{ij}^s H_1)_{ab} \quad (\text{A.4.17})$$

and

$$C_{ia} = \sum_j^p \sum_b^{p \wedge m} (H_1)_{jb} (H_1)_{ab} (d_{ij}^s \psi)_b. \quad (\text{A.4.18})$$

First, we deal with (A.4.16) and (A.4.17). Thanks to (A.4.6), we obtain

$$\begin{aligned} A_{ia} &= \frac{1}{2} \sum_j^p \sum_b^{p \wedge m} \sum_{r \neq b}^{p \wedge m} \frac{\psi_b (H_1)_{ab} (H_1)_{jr}}{l_b - l_r} [(H_1)_{ir} (H_1)_{jb} + (H_1)_{jr} (H_1)_{ib}] \\ &\quad + \frac{1}{2} \sum_j^p \sum_b^{p \wedge m} \frac{\psi_b (H_1)_{ab}}{l_b} [(I_p - H_1 H_1^T)_{ji} (H_1)_{jb} + (I_p - H_1 H_1^T)_{jj} (H_1)_{ib}] \end{aligned}$$

and

$$\begin{aligned} B_{ia} &= \frac{1}{2} \sum_j^p \sum_b^{p \wedge m} \sum_{r \neq b}^{p \wedge m} \frac{\psi_b (H_1)_{jb} (H_1)_{ar}}{l_b - l_r} [(H_1)_{ir} (H_1)_{jb} + (H_1)_{jr} (H_1)_{ib}] \\ &\quad + \frac{1}{2} \sum_j^p \sum_b^{p \wedge m} \frac{\psi_b (H_1)_{jb}}{l_b} [(I_p - H_1 H_1^T)_{ai} (H_1)_{jb} + (I_p - H_1 H_1^T)_{aj} (H_1)_{ik}]. \end{aligned}$$

Summing on j , we have

$$\begin{aligned} A_{ia} &= \frac{1}{2} \sum_b^{p \wedge m} \sum_{r \neq b}^{p \wedge m} \frac{\psi_b (H_1)_{ab}}{l_b - l_r} [(H_1)_{ir} (H_1^T H_1)_{rb} + (H_1^T H_1)_{rr} (H_1)_{ib}] \\ &\quad + \frac{1}{2} \sum_b^{p \wedge m} \frac{\psi_b (H_1)_{ab}}{l_b} [(H_1^T (I_p - H_1 H_1^T))_{bi} + \text{tr}(I_p - H_1 H_1^T) (H_1)_{ib}] \end{aligned}$$

and

$$\begin{aligned} B_{ia} &= \frac{1}{2} \sum_b^{p \wedge m} \sum_{r \neq b}^{p \wedge m} \frac{\psi_b (H_1)_{ar}}{l_b - l_r} [(H_1)_{ir} (H_1^T H_1)_{bb} + (H_1^T H_1)_{br} (H_1)_{ib}] \\ &\quad + \frac{1}{2} \sum_b^{p \wedge m} \frac{\psi_b}{l_b} [(I_p - H_1 H_1^T)_{ai} (H_1^T H_1)_{bb} + ((I_p - H_1 H_1^T) H_1)_{ab} (H_1)_{ik}]. \end{aligned}$$

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

$$A_{ia} = \frac{1}{2} \sum_b^{p \wedge m} \sum_{r \neq b}^{p \wedge m} \frac{\psi_b (H_1)_{ab}}{l_b - l_r} (H_1)_{ib} + \frac{1}{2} \sum_b^{p \wedge m} \frac{\psi_b (H_1)_{ab}}{l_b} [(H_1^T (I_p - H_1 H_1^T))_{bi} + \text{tr}(I_p - H_1 H_1^T) (H_1)_{ib}]$$

and

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

Since $H_1^T H_1 = I_{p \wedge m}$, we have $H_1^T (I_p - H_1 H_1^T) = ((I_p - H_1 H_1^T) H_1)^T = 0_{(p \wedge m) \times p}$ and $\text{tr}(I_p - H_1 H_1^T) = p - (p \wedge m)$. Hence

$$A_{ia} = \frac{1}{2} \sum_b^{p \wedge m} (H_1)_{ib} \sum_{r \neq b}^{p \wedge m} \frac{\psi_b}{l_b - l_r} (H_1^T)_{ba} + \frac{1}{2} (p - (p \wedge m)) \sum_b^{p \wedge m} (H_1)_{ib} \frac{\psi_b}{l_b} (H_1^T)_{ba} \quad (\text{A.4.19})$$

and

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

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,

$$B_{ia} = -\frac{1}{2} \sum_b^{p \wedge m} (H_1)_{ib} \sum_{b \neq r}^{p \wedge m} \frac{\psi_r}{l_b - l_r} (H_1^T)_{ba} + \frac{1}{2} \text{tr}(L^{-1} \Psi) (I_p - H_1 H_1^T)_{ia}. \quad (\text{A.4.20})$$

Dealing with C_{ia} in (A.4.18), 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.4.5), we have

$$C_{ia} = \sum_b^{p \wedge m} (H_1)_{ab} \sum_q^{p \wedge m} \frac{\partial \psi_b}{\partial l_q} (H_1)_{iq} (H_1 H_1^T)_{qb},$$

and hence,

$$C_{ia} = \sum_b^{p \wedge m} (H_1)_{ib} \frac{\partial \psi_b}{\partial l_b} (H_1^T)_{ba}, \quad (\text{A.4.21})$$

since $H_1^T H_1 = I_{p \wedge m}$. Finally, gathering expressions (A.4.19), (A.4.20) and (A.4.21) gives the result in (A.4.12).

Now, in order to prove (A.4.13) (respectively (A.4.14)) we apply (A.4.12) 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.4.15), we have from (A.2.1)

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

Then, according to (A.4.14), we have

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

□

References

- [1] J. O. Berger. Minimax estimation of location vectors for a wide class of densities. *Annals of Statistics*, 3:1318–1328, 1975.
- [2] E.J. Candès, C.A. Sing-Long, and J.D. Trzasko. Unbiased risk estimates for singular value thresholding and spectral estimators. *IEEE Transactions on signal processing*, 61:4643–4657, 2013.
- [3] S. Canu and D. Fourdrinier. Unbiased risk estimates for matrix estimation in the elliptical case. *Journal of Multivariate Analysis*, 158:60–72, 2017.
- [4] D. K. Dey and C. Srinivasan. Estimation of a covariance matrix under Stein’s loss. 13:1581–1591, 1985.
- [5] B. Efron and C. Morris. Multivariate empirical Bayes and estimation of covariance matrices. *The Annals of Statistics*, 4:22–32, 1976.
- [6] T.J. Fisher and X. Sun. Improved Stein-type shrinkage estimators for the high-dimensional multivariate normal covariance matrix. *Computational Statistics and Data Analysis*, 55:1909–1918, 2011.
- [7] 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.
- [8] D. Fourdrinier, F. Mezoued, and M.T. Wells. Estimation of the inverse scatter matrix of an elliptically symmetric distribution. *Journal of Multivariate Analysis*, 143:32–55, 2016.
- [9] L.R. Haff. An identity for the Wishart distribution with applications. *Journal of Multivariate Analysis*, 9:531–544, 1979.
- [10] L.R. Haff. Empirical Bayes estimation of the multivariate normal covariance matrix. *The Annals of Statistics*, 8:586–597, 1980.
- [11] L.R. Haff. On linear log-odds and estimation of discriminant coefficients. *Communications in Statistics-Theory and Methods*, 15:2131–2144, 1986.
- [12] L.R. Haff. The variational form of certain Bayes estimators. *The Annals of Statistics*, 19:1163–1190, 1991.
- [13] D.A. Harville. *Matrix Algebra from a statistician’s perspective*. Springer, 2008.
- [14] 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, California, 1961. University of California Press.
- [15] 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.
- [16] T. Kubokawa, C. Robert, and A.M.E. Saleh. Empirical Bayes estimation of the covariance matrix of a normal distribution with unknown mean under an entropy loss. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 402–410, 1992.

- [17] 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.
- [18] O. Ledoit and M. Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88:365 – 411, 2004.
- [19] Y. Sheena. Unbiased estimator of risk for an orthogonally invariant estimator of a covariance matrix. *Journal of Japan Statistical Society*, 25:35–48, 1995.
- [20] C. Stein. Estimation of a covariance matrix. In *Rietz Lecture, 39th Annual Meeting of the IMS*. Atlanta, Georgia, 1975.
- [21] C. Stein. Lectures on the theory of estimation of many parameters. *Journal of Soviet Mathematics*, 34:1373–1403, 1986.
- [22] A. Takemura. An orthogonally invariant minimax estimator of the covariance matrix of a multivariate normal population. *Tsukuba Journal of Mathematics*, 8:367–376, 1984.

BIBLIOGRAPHIE

- [And03] T.W. Anderson. *Estimation of the mean vector and the covariance matrix. in :An Introduction to multivariate statistical analysis (Wiley series in probability and statistics)*. Wiley-Interscience, 3 edition, 2003. 3
- [ANS14] I. Alvarez, J. Niemi, and M. Simpson. Bayesian inference for a covariance matrix. *arXiv preprint arXiv :1408.4050*, 2014. 6
- [Ber75] J. Berger. Minimax estimation of location vectors for a wide class of densities. *The Annals of Statistics*, 3(6) :1318–1328, 1975. 21
- [Bil95] P. Billingsley. *Probability and measure*. Wiley Series in Probability and Statistics. John Wiley & Sons Inc, 2 sub edition, 1995. 16
- [BST05] O. Berger, W. E. Strawderman, and D. Tang. Posterior propriety and admissibility of hyperpriors in normal hierarchical models. *The Annals of Statistics*, 33(2) :606–646, 04 2005. 6
- [CF17] S. Canu and D. Fourdrinier. Unbiased risk estimates for matrix estimation in the elliptical case. *Journal of Multivariate Analysis*, 158(C) :60–72, 2017. 9, 18, 20
- [CFR89] D. Cellier, D. Fourdrinier, and C. Robert. Robust shrinkage estimators of the location parameter for elliptically symmetric distributions. *Journal of Multivariate Analysis*, 29(1) :39–52, 1989. 20
- [Chm81] M.A. Chmielewski. Elliptically symmetric distributions : A review and bibliography. *International Statistical Review*, 49(1) :67–74, 1981. 11
- [CST13] E.J. Candès, C.A. Sing-Long, and J.D. Trzasko. Unbiased risk estimates for singular value thresholding and spectral estimators. *IEEE Transactions on signal processing*, 61 :4643–4657, 2013. 9, 18
- [CW16] D. Chételat and M.T. Wells. Improved second order estimation in the singular multivariate normal model. *Journal of Multivariate Analysis*, 147(C), 2016. 8, 66
- [CWH09] Y. Chen, A. Wiesel, and A.O. Hero. Shrinkage estimation of high dimensional covariance matrices. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2937–2940, 2009. 7
- [Dev86] L. Devroye. *Non-uniform random variate generation*. Springer-Verlag, New York, NY, USA, 1986. 59

- [DGLSG02] J.A. Díaz-García, V. Leiva-Sánchez, and M. Galea. Singular elliptical distribution : Density and applications. *Communications in Statistics - Theory and Methods*, 31(5) :665–681, 2002. 16
- [DRT81] A.P. Dempster, D.B. Rubin, and R.K. Tsutakawa. Estimation in covariance components models. *Journal of the American Statistical Association*, 76(374) :341–353, 1981. 6
- [DS85] D.K. Dey and C. Srinivasan. Estimation of a covariance matrix under Stein’s loss. *The Annals of Statistics*, 13(4) :1581–1591, 1985. 6, 36
- [EM76] B. Efron and C. Morris. Multivariate empirical bayes and estimation of covariance matrices. *Annals of Statistics*, 4(1) :22–32, 1976. 45
- [Fis19] R.A. Fisher. The correlation between relatives on the supposition of mendelian inheritance. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, 52(2) :399–433, 1919. 1
- [Fis36] R.A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(7) :179–188, 1936. 2
- [FKN90] K.T. Fang, S. Kotz, and K.W. Ng. *Symmetric multivariate and related distributions*. Springer, 1990. 12, 13, 15, 16, 59
- [FMS12] 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. 21
- [FMW16] D. Fourdrinier, F. Mezoued, and M.T. Wells. Estimation of the inverse scatter matrix of an elliptically symmetric distribution. *Journal of Multivariate Analysis*, 143(C) :32–55, 2016. 21, 22, 26
- [FS08] D. Fourdrinier and W.E. Strawderman. A unified and generalized set of shrinkage bounds on minimax stein estimates. *Journal of Multivariate Analysis*, 99(10) :2221–2233, 2008. 21
- [FS10] D. Fourdrinier and W. E. Strawderman. Robust generalized bayes minimax estimators of location vectors for spherically symmetric distributions with unknown scale. In *Borrowing strength : theory powering applications—A Festschrift for Lawrence D. Brown*, volume 6, pages 249–262. Institute of Mathematical Statistics, 2010. 20
- [FS11] T.J. Fisher and X. Sun. Improved Stein-type shrinkage estimators for the high-dimensional multivariate normal covariance matrix. *Computational Statistics and Data Analysis*, 55(5) :1909–1918, 2011. 7, 36
- [FS15] D. Fourdrinier and W.E. Strawderman. Robust minimax Stein estimation under invariant data-based loss for spherically and elliptically symmetric distributions. *Metrika*, 78(4) :461–484, 2015. 45
- [FSW18] D. Fourdrinier, W.E. Strawderman, and M.T. Wells. *Shrinkage Estimation*. Springer, 2018. 12, 13, 15, 16
- [FZ90] K.T. Fang and Y.T. Zhang. *Generalized multivariate analysis. 1990*. Science Press, Springer-Verlag, Beijing, 1990. 17
- [GVB13] A.K. Gupta, T. Varga, and T. Bodnar. *Elliptically contoured models in statistics and Portfolio theory*. Springer-Verlag New York, 2 edition, 2013. 11
- [GVL12] G.H. Golub and C.F. Van Loan. *Matrix computations*, volume 3. Johns Hopkins University Press, 2012. 19
- [Haf79] L.R. Haff. An identity for the Wishart distribution with applications. *Journal of Multivariate Analysis*, 9(4) :531–544, 1979. 24

- [Haf80] L.R. Haff. Empirical bayes estimation of the multivariate normal covariance matrix. *The Annals of Statistics*, 8(3) :586–597, 1980. 6, 28, 34, 38, 43, 44, 66, 69
- [Haf86] L.R. Haff. On linear log-odds and estimation of discriminant coefficients. *Communications in Statistics-Theory and Methods*, 15(7) :2131–2144, 1986. 36
- [Haf91] L.R. Haff. The variational form of certain Bayes estimators. *The Annals of Statistics*, 19(3) :1163–1190, 1991. 6, 75
- [Han82] L. Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4) :1029–54, 1982. 2
- [Har08] D.A. Harville. *Matrix Algebra from statistician's perspective*. Springer, 2008. 72
- [HV08] M. Hamada and E. A. Valdez. Capm and option pricing with elliptically contoured distributions. *Journal of Risk and Insurance*, 75(2) :387–409, 2008. 58
- [IKS16] Y. Ikeda, T. Kubokawa, and M.S. Srivastava. Comparison of linear shrinkage estimators of a large covariance matrix in normal and non-normal distributions. *Computational Statistics and Data Analysis*, 95(C) :95 – 108, 2016. 7
- [JLSX10a] H. Ji, C. Liu, Z. Shen, and Y. Xu. Robust video denoising using low rank matrix completion. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1791–1798. IEEE, 2010. 9
- [JLSX10b] H. Ji, C. Liu, Z. Shen, and Y. Xu. Robust video denoising using low rank matrix completion. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1791–1798. IEEE, 2010. 18
- [Jol02] I. Jolliffe. *Principal component analysis*. Springer Verlag, New York, 2002. 1
- [Jor96] P. Jorion. Risk 2 : Measuring the risk in value at risk. *Financial Analysts Journal*, 52(6) :47–56, 1996. 1
- [JS61] 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. 3, 4
- [Kel70] D. Kelker. Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā : The Indian Journal of Statistics, Series A (1961-2002)*, 32(4) :419–430, 1970. 12
- [Kon07] Y. Konno. Improving on the sample covariance matrix for a complex elliptically contoured distribution. *Journal of Statistical Planning and Inference*, 137(7) :2475 – 2486, 2007. 8
- [Kon09] 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. 7, 34, 43, 66
- [KRS92] T. Kubokawa, C. Robert, and AK. Md. E. Saleh. Empirical Bayes estimation of the covariance matrix of a normal distribution with unknown mean under an entropy loss. *Sankhyā : The Indian Journal of Statistics, Series A*, 54(2) :402–410, 1992. 6, 36
- [KS99] 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. 8, 24, 26
- [KS01] T. Kubokawa and M.S. Srivastava. Robust improvement in estimation of a mean matrix in an elliptically contoured distribution. *Journal of Multivariate Analysis*, 76(1) :138–152, 2001. 20

- [KS03] T. Kubokawa and M.S. Srivastava. Estimating the covariance matrix : a new approach. *Journal of Multivariate Analysis*, 86(1) :28–47, 2003. 8
- [KS08] T. Kubokawa and M.S. Srivastava. Estimation of the precision matrix of a singular Wishart distribution and its application in high-dimensional data. *Journal of Multivariate Analysis*, 99(9) :1906–1928, 2008. 45
- [KT06] T. Kubokawa and M.T. Tsai. Estimation of covariance matrices in fixed and mixed effects linear models. *Journal of multivariate analysis*, 97(10) :2242–2261, 2006. 8
- [LP85] S. Lin and M. Perlman. A Monte Carlo comparison of four estimators of a covariance matrix. *Multivariate Analysis*, pages 411–429, 1985. 6
- [LW03] O. Ledoit and M. Wolf. Honey, I shrunk the sample covariance matrix. *UPF Economics and Business Working Paper*, 30(691), 2003. 57
- [LW04] O. Ledoit and M. Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2) :365 – 411, 2004. 7, 36
- [Mah36] P. C. Mahalanobis. On the generalised distance in statistics. In *Proceedings National Institute of Science, India*, volume 12, pages 49–55, 1936. 16
- [Man63] B. Mandelbrot. The variation of certain speculative prices. *The Journal of Business*, 36(4) :394–419, 1963. 58
- [Mar52] H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1) :77–91, 1952. 1, 2, 11, 56
- [Mar03] Y. Maruyama. A robust generalized Bayes estimator improving on the James-Stein estimator for spherically symmetric distributions. *Statistics and Decisions International Mathematical Journal for Stochastic Methods and Models*, 21(1) :69–78, 2003. 20
- [MS05] Y. Maruyama and W.E. Strawderman. A new class of generalized Bayes minimax ridge regression estimators. *The Annals of Statistics*, 33(4) :1753–1770, 2005. 20
- [Pou13] M. Pourahmadi. *Covariance matrices. In : High-dimensional covariance estimation with high-dimensional data*. Wiley, 1 edition, 2013. 3, 6
- [S.64] William F. S. Capital asset price : A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3) :425–442, 1964. 58
- [Sri03] M.S. Srivastava. Singular Wishart and multivariate Beta distributions. *The Annals of Statistics*, 31(5) :1537–1560, 2003. 16
- [SS05] J. Schafer and K. Strimmer. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology*, 4(1) :1175–1189, 2005. 7
- [Ste56] C. Stein. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1 : Contributions to the Theory of Statistics*, pages 197–206, Berkeley, Calif., 1956. University of California Press. 3
- [Ste75] C. Stein. Estimation of a covariance matrix. In *Rietz Lecture, 39th Annual Meeting of the IMS*. Atlanta, Georgia, 1975. 5, 8, 24
- [Ste77] C. Stein. Lectures on the theory of estimation of many parameters. In *URSS*, 1977. 5, 8
- [Ste86] C. Stein. Lectures on the theory of estimation of many parameters. *Journal of Soviet Mathematics*, 34(1) :1373–1403, 1986. 5, 8, 36

- [TK15] H. Tsukuma and T. Kubokawa. A unified approach to estimating a normal mean matrix in high and low dimensions. *Journal of Multivariate Analysis*, 139 :312 – 328, 2015. [45](#)
- [TK16] H. Tsukuma and T. Kubokawa. Unified improvements in estimation of a normal covariance matrix in high and low dimensions. *Journal of Multivariate Analysis*, 100(143) :233–248, 2016. [8](#), [66](#)
- [Uhl94] H. Uhlig. On singular Wishart and singular multivariate beta distributions. *The Annals of Statistics*, 22(1) :395–405, 1994. [16](#)
- [YB94] R. Yang and J.O. Berger. Estimation of a covariance matrix using the reference prior. *The Annals of Statistics*, 22(3) :1195–1211, 09 1994. [6](#)
- [YB18] R. Yang and J.O. Berger. Bayesian analysis of the covariance matrix of normal distribution with a new class of prior. *Unpublished paper*, 2018. [6](#)