université PARIS-SACLAY

NNT : 2018SACLC091

Thèse de doctorat

Théorie des matrices aléatoires pour l'imagerie hyperspectrale

Thèse de doctorat de l'Université Paris-Saclay
préparée à CentraleSupélec

École doctorale n°580 : Sciences et Technologies de l'Information et de la Communication (STIC)

Spécialité de doctorat : Automatique, traitement du signal, traitement des images, robotique (pôle 1)

Thèse présentée et soutenue à Gif-sur-Yvette le 18 décembre 2018 par

Eugénie TERREAUX

Composition du jury :

Julie Delon Professeur Université, MAP5, Université Paris Descartes	Présidente
Jean-Yves Tourneret Professeur Université, INP-ENSEIHHT, Toulouse	Rapporteur
Jocelyn Chanussot Professeur Université, Gipsa-Lab, Grenoble	Rapporteur
Sylvie Marcos DR CNRS, L2S CentraleSupelec, Gif-sur-Yvette	Examinateur
Pascal Vallet Maître de conférence, IPB - Institut Polytechnique de Bordeaux	Examinateur
Jean-Philippe OVARLEZ Maître de recherche 2, ONERA Palaiseau et CentraleSupelec SONDRA (DEMR)	Directeur de thèse
Frédéric PASCAL Professeur, CentraleSupelec L2S	Codirecteur de thèse

Laboratoire SONDRA – CentraleSupélec
3 rue Joliot Curie, 91192 Gif-sur-Yvette cedex

Remerciements

Cette thèse a été menée au laboratoire SONDRA de CentraleSupelec, avec le support de la DGA. Je remercie ainsi les acteurs de ces organismes pour avoir permis à cette thèse d'être menée à bien.

Je remercie mon directeur de thèse Jean-Philippe Ovarlez pour son encadrement fructueux, sa patience et son support durant ces trois ans de thèse. Je remercie également mon co-directeur de thèse Frédéric Pascal pour ses nombreuses idées et son expertise scientifique. Ces trois années ont été riches en enseignement grâce à vous.

Je remercie les membres de mon jury de thèse, pour l'intérêt qu'ils ont montré à mon travail. Je remercie Jean-Yves Tourneret et Jocelyn Chanussot d'avoir rapporté mon manuscrit. Je remercie Sylvie Marcos, Pascal Vallet et Julie Delon d'avoir accepté de faire partie de mon jury de thèse. Je remercie également Romain Couillet pour son aide scientifique. Enfin, merci à Emmanuelle Jay pour son apport non négligeable à cette thèse, et pour m'avoir initiée à certaines problématiques de la finance. Merci aussi à Chengfang Ren pour sa grande disponibilité.

Je remercie également les membres du laboratoire SONDRA pour leur accueil. Je remercie particulièrement Orian Couderc, Pedro Mendes-Ruiz, Israel Inostroza, Uyhour Tan, Chengfang Ren et Thierry Letertre pour tous ces moments partagés. Merci également à Virginie Bouvier, Sylvain Azarian, Laetitia Thirion-Lefevre et Régis Guinvarc'H pour leur bonne humeur. Je souhaite bon courage et bonne chance à Ammar Mian et Bruno Meriaux pour la suite de leurs travaux. Merci à Alice Combernoux, Lucien Bacharach et Virginie Ollier pour leur soutien. Je remercie également Alexandre Renaux, Nabil El Korso et les membres du L2S et de l'ONERA pour leur sympathie. Je remercie tous ceux qui m'ont apporté une aide quelconque durant ces trois années.

Pour finir, un grand merci à mes amies et à ma famille, en particulier mes sœurs Claire et Anne-Sophie ainsi que mes parents, merci pour votre soutien et votre implication. Et enfin merci Clément pour ces trois années riches en évènement, pour ton soutien inébranlable et pour tout le reste. Cette thèse est aussi pour toi, Maxime, que tu connaisses la vraie valeur du travail.

Merci à tous,
Eugénie.

Table des matières

Abréviations **ix**

Notations **xi**

1 Problèmes et méthodes en imagerie hyperspectrale **3**
1.1 L'imagerie hyperspectrale . . . 3
1.2 Quelques méthodes classiques pour le problème de l'estimation du nombre d'*endmembers* . . . 5
1.2.1 Des méthodes qui reposent sur un critère d'information théorique . . . 6
1.2.2 Des méthodes statistiques plus générales et la méthode Hysime . . . 7
1.2.3 Des méthodes géométriques . . . 9
1.3 Quelques méthodes classiques pour le problème du démélange . . . 9
1.3.1 Des méthodes statistiques . . . 10
1.3.2 Des méthodes géométriques . . . 11
1.4 Modélisation CES d'une image hyperspectrale . . . 13
1.4.1 Les distributions elliptiques complexes et les matrices de Toeplitz . . . 13
1.4.2 Récapitulatif sur le modèle et les hypothèses . . . 17
1.5 Limite des méthodes proposées et motivations pour la théorie des matrices aléatoires . . . 18
1.5.1 Méthodes d'estimation du nombre d'*endmembers* qui reposent sur un critère d'information théorique . . . 19
1.5.2 Méthodes statistiques d'estimation du nombre d'*endmembers* . . . 19
1.5.3 Méthodes de démélange géométriques . . . 20
1.5.4 Motivations pour la théorie des matrices aléatoires . . . 21
1.6 La théorie des matrices aléatoires . . . 23
1.6.1 Introduction à la théorie des matrices aléatoires . . . 23
1.6.2 Quelques éléments de base en théorie des matrices aléatoires . . . 24
1.6.3 Distribution des valeurs propres pour un modèle plus général . . . 29
1.6.4 Distribution des valeurs propres pour le modèle *Spiked* . . . 32
1.7 Quelques méthodes RMT pour l'imagerie hyperspectrale . . . 35
1.8 Introduction à l'estimation robuste . . . 36
1.9 Synthèse . . . 39

2 Sélection d'ordre pour le démélange d'images hyperspectrales **41**
2.1 Rappel des principales hypothèses . . . 41
2.2 Estimation du nombre d'*endmembers* pour un bruit non corrélé, d'entrées indépendantes et identiquement distribuées. . . . 42
2.3 Estimation du nombre d'*endmembers* pour un bruit corrélé, d'entrées indépendantes et identiquement distribuées . . . 43
2.4 Estimation du nombre d'*endmembers* pour un bruit CES . . . 47
2.4.1 Bruit CES "blanc" . . . 47
2.4.2 Approche non-robuste pour un bruit CES corrélé . . . 50

2.4.3 Approche robuste pour un bruit CES corrélé . . . 56
2.5 Récapitulatif des algorithmes proposés pour l'estimation du nombre d'end-members . . . 66
2.5.1 Bruit gaussien blanc . . . 67
2.5.2 Bruit corrélé i.i.d. de moyenne nulle . . . 67
2.5.3 Bruit CES blanc . . . 68
2.5.4 Approche RMT non-robuste . . . 68
2.5.5 Approche RMT robuste cas général . . . 69
2.5.6 Approche robuste RMT : cas de l'estimation de Tyler . . . 70
2.6 Synthèse . . . 71

3 Applications **73**
3.1 Imagerie hyperspectrale . . . 73
3.1.1 Approche RMT non CES . . . 73
3.1.2 Comparaison des méthodes adaptés au bruit CES . . . 75
3.2 Allocation de portefeuilles en finance . . . 82
3.2.1 Petite introduction du problème . . . 82
3.2.2 Méthode suivie . . . 84
3.2.3 Comparaison des résultats . . . 84
3.3 Synthèse . . . 87

4 Le problème du démélange spectral **89**
4.1 Démélange spectral dans le cas où $p \ll m$ . . . 90
4.1.1 Problème et méthodes . . . 90
4.1.2 Nouvelles méthodes de démélange d'images hyperspectrales . . . 91
4.1.3 Applications . . . 94
4.2 Limite du démélange spectral dans le cas où $p < m$ . . . 111
4.2.1 Méthode . . . 111
4.3 Synthèse . . . 112

Conclusion **115**

A Généralités **119**
A.1 Définitions . . . 119
A.2 Lemmes et théorèmes . . . 120

B Preuves des théorèmes **121**
B.1 Preuve du Théorème 2.4.1 . . . 121
B.1.1 Analyse de sup $\left|\mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda)\right] - \mathbb{E}(\tau)\,\gamma_m(\lambda)\right|$ pour $\lambda \in [0, 2\pi)$ . . . 122
B.1.2 Analyse de sup $\left|\hat{\gamma}_m^{noise}(\lambda) - \mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda)\right]\right|$ pour $\lambda \in [0, 2\pi)$ . . . 123
B.2 Preuve du Théorème 2.4.2 . . . 125
B.3 Preuve du Théorème 2.4.3 . . . 126
B.4 Preuve du Théorème 2.4.5 . . . 129
B.5 Preuve du Théorème 2.4.6 . . . 129

C Publications **131**
C.1 Conférences . . . 131
C.1.1 Soumis . . . 131
C.2 Journaux . . . 131

Table des figures

1.1 Image hyperspectrale issue de la base de données du DSO National Laboratories. 4
1.2 Spectres et abondances pour une scène urbaine [106]. 5
1.3 Valeurs des critères AIC, MDL et de la méthode [94] pour $N = 1000$,$m = 10$, $p = 4$, RSB $= 20dB$ 8
1.4 Simplex obtenu avec $p = 3$, $m = 900$ et $N = 1000$, données issues de [60]. . . 13
1.5 $J(\mathbf{M}, \mathbf{S})$ pour différentes méthodes, avec : $N = 1000$,$m = 96$, sur images simulées avec des spectres hyperspectraux de [60] et un bruit additif gaussien. 14
1.6 $N = 1000$, $m = 10$, $p = 4$, bruit CES. 19
1.7 Performances de la méthode Hysime pour différents cas de figure. 20
1.8 Représentation des données dans un simplex, $N = 1000$, $m = 900$, données corrélées. 21
1.9 $J(\mathbf{M}, \mathbf{S})$ pour différentes méthodes, avec $N = 1000$ et $m = 960$, sur images simulées avec des spectres hyperspectraux de [60] et un bruit additif CES . . 22
1.10 Comparaison entre les différentes méthodes de $\left\|\hat{\mathbf{M}} - \mathbf{M}\right\|_F$. 22
1.11 Distribution des valeurs propres de la SCM d'un bruit blanc, à gauche $c_N = 0.01$ et à droite $c_N = 0.57$. 25
1.12 Distribution des valeurs propres de $\frac{1}{N}\mathbf{Y}\mathbf{Y}^H$ pour différentes valeurs de c . . . 26
1.13 Loi de Marchenko-Pastur et histogramme des valeurs propres pour un bruit blanc gaussien, de variance unité et de moyenne nulle, $c = 0,90$, avec, à droite $N = 100$ et $m = 90$, et, à gauche, $m = 900$ et $N = 1000$. 27
1.14 Loi de Tracy-Widom et histogramme des valeurs propres, $c_N = 0.57$. 28
1.15 Distribution des valeurs propres pour $c = 0.09$, $\nu = \frac{1}{3}\delta_1 + \frac{1}{3}\delta_5 + \frac{1}{3}\delta_{10}$. 30
1.16 Zones correspondants au support complémentaire pour la distribution des valeurs propres de la matrice $\mathbf{W}$. 32
1.17 Distribution des valeurs propres de l'estimation empirique de la matrice de covariance pour un bruit gaussien et deux sources indépendantes. 34
1.18 Distribution des valeurs propres de l'estimation empirique de la matrice de covariance pour un bruit gaussien et quatre sources indépendantes. 34
1.19 Fonctions de poids de certains M-estimateurs pour $m = 10$ et différentes valeurs de q pour l'estimateur de Huber. 39

2.1 à gauche : PFA (en $\log_{10}$) en fonction du seuil ζ pour 3000 itérations et à droite : probabilité de détection en fonction du RSB, pour 500 itérations, $m = 200$, $N = 1000$ et une PFA de 0.01, en vert les barres d'erreurs. 46
2.2 Illustration du Théorème 2.4.1 pour $c = 0.7$ et pour des $\{\tau_i\}_{i\in[\![1,N]\!]}$ suivant une loi inverse gamma ($N \in [\![20, 6000]\!]$) 52
2.3 Comparaison de la convergence entre $\check{\mathbf{C}}_{SCM}$ et la SCM, pour $c = 0.2$, $\rho = 0.7$, échelle logarithmique. 52
2.4 Valeurs propres des matrices $\check{\mathbf{\Sigma}}_{SCM}$ et $\hat{\mathbf{S}}$ lorsque le signal $\mathbf{Y}$ a été blanchi par $\check{\mathbf{C}}_{SCM}$ et le seuil correspondant t pour $\rho = 0.7$, $m = 900$, $N = 2000$, $\tau =$ inverse gamma, $\alpha = 0.1$, $p = 0$. 55

2.5 Valeurs propres des matrices $\check{\mathbf{\Sigma}}$ et $\hat{\mathbf{S}}$ lorsque le signal $\mathbf{Y}$ n'a pas été blanchi par $\check{\mathbf{C}}_{SCM}$, et le seuil correspondant t pour $\rho = 0.7$, $m = 900$, $N = 2000$, $\tau =$ inverse gamma, $\alpha = 0.1$, $p = 0$. 55
2.6 Valeurs propres des matrices $\check{\mathbf{\Sigma}}_{SCM}$ et $\hat{\mathbf{S}}$ et le seuil t, pour $\rho = 0.7$, $m = 900$, $N = 2000$, $\tau = t^2 - student$, $\alpha = 0.1$, $p = 0$. 56
2.7 Illustration de la consistance du Théorème 2.4.3 pour $c = 0.7$ et pour des $\{\tau_i\}_{i\in[\![1,N]\!]}$ suivant une loi inverse gamma ($N \in [\![20, 6000]\!]$). 58
2.8 Distribution des valeurs propres des matrices $\hat{\mathbf{S}}$ et $\check{\mathbf{\Sigma}}_{FP}$ ainsi que le seuil t pour $\rho = 0.7$, $m = 900$, $N = 2000$, $\tau \sim$ inverse gamma. 59
2.9 Illustration de la robustesse au choix de u. Distribution des valeurs propres des matrices $\hat{\mathbf{S}}$ et $\check{\mathbf{\Sigma}}_{FP}$ et le seuil t pour $\rho = 0.7$, $m = 900$, $N = 2000$, $\tau = t^2$, $t \sim$ Student 60
2.10 Distribution des valeurs propres des matrices $\hat{\mathbf{S}}$ et $\tilde{\mathbf{\Sigma}}_{FP}$ pour un signal non blanchi ainsi que le seuil t pour $\rho = 0.7$, $m = 900$, $N = 2000$, $\tau = t^2$, $t \sim$ Student. 60
2.11 Distribution des valeurs propres des matrices $\hat{\mathbf{S}}$ et de la matrice représentant la SCM de $\check{\mathbf{Y}}_{wFP}\check{\mathbf{Y}}_{wFP}^H/N$ et le seuil t pour $\rho = 0.7$, $m = 900$, $N = 2000$, $\tau \sim$ inverse gamma. 61
2.12 Distribution des valeurs propres des matrices $\hat{\mathbf{S}}$ et $\check{\mathbf{\Sigma}}_{FP}$ lorsqu'un endmembers est présent dans la signal, $RSB = 10dB$ ainsi que le seuil t, pour $\rho = 0.7$, $m = 900$, $N = 2000$, $\tau \sim$ inverse gamma. 62
2.13 63
2.14 Distribution de Marchenko-Pastur et distribution des valeurs propres de $\check{\mathbf{\Sigma}}_{TYL}$ pour 4 *endmembers* de RSB=10 dB, $N = 2000$, $m = 900$, $c = 0.7$, les $\{\tau\}_{i\in[\![1,N]\!]}$ suivant une loi inverse gamma. 65
2.15 Distribution de Marchenko-Pastur et distribution des valeurs propres de $\check{\mathbf{\Sigma}}_{TYL}$ pour 4 *endmembers* de RSB=30 dB, $N = 2000$, $m = 900$, $c = 0.7$, les $\{\tau\}_{i\in[\![1,N]\!]}$ suivant une loi t^2, t distribué selon Student-t. 65
2.16 Distribution de Marchenko-Pastur et distribution des valeurs propres de l'estimation de la matrice de covariance avec un blanchiment FP $u(x) = \frac{1+0.1}{x+0.1}$, pour 4 *endmembers* de $RSB = 10dB$, $N = 2000$, $m = 900$, $c = 0.7$, les $\{\tau\}_{i\in[\![1,N]\!]}$ suivant une loi inverse gamma. 66

3.1 Image hyperspectrale utilisée, avec et sans la voiture, de taille $m = 167$, $N = 6561$. 74
3.2 Courbe de la PFA en fonction du seuil, $m = 167$, $N = 21 \times 21$. 74
3.3 Différentes distributions de bruit, à gauche lorsque les $\{\tau_i\}_{i\in[\![1,N]\!]}$ suivent une loi du t^2 (carré d'une loi de Student) et à droite une loi inverse gamma. . . . 76
3.4 Estimation du nombre d'*endmembers* $\hat{p}$ pour un bruit CES corrélé, $m = 400$, $c = 0.2$, $p = 4$, $\rho = 0.7$, $\{\tau\} \sim$ inverse gamma, en fonction du RSB. 77
3.5 Estimation du nombre d'*endmembers* $\hat{p}$ pour un bruit CES corrélé, $m = 400$, $c = 0.2$, $p = 4$, $\rho = 0.7$, $\{\tau\} \sim t^2$ avec $t \sim$ Student-t , en fonction du RSB. . . 77
3.6 Estimation du nombre d'*endmembers* $\hat{p}$ pour un bruit CES corrélé, $m = 900$, $c = 0.45$, $p = 4$,$\rho = 0.7$, $\{\tau\} \sim$ inverse gamma, en fonction du RSB. 78
3.7 Estimation du nombre d'*endmembers* $\hat{p}$ pour un bruit CES corrélé, $m = 900$, $c = 0.45$, $p = 4$,$\rho = 0.7$, $\{\tau\} \sim t^2$ avec $t \sim Student - t$, en fonction du RSB. . 78
3.8 Spectres réels utilisés pour la figure 3.9. 79
3.9 Estimation du nombre d'*endmembers* $\hat{p}$ pour un bruit CES corrélé, $m = 240$, $c = 0.24$, $p = 2$,$\rho = 0.7$, $\{\tau\} \sim$ inverse gamma , en fonction du RSB. 80
3.10 Une bande spectrale des images Indian Pines (à gauche) et PaviaU (à droite) 81
3.11 Une bande spectrale des images "Cube" (à gauche) et SalinasA (à droite) . . 81
3.12 richesse du portefeuille, débutant à 100 pour la première période, sur le début du mois de Janvier. 86

3.13 "*Turnover*" entre les différentes périodes consécutives. 86

4.1 5 premières composantes de la PCA sur l'image Indian Pines [35]. 95
4.2 Vérité terrain pour la figure Indian Pines. 96
4.3 5 premières composantes de la PCA sur l'image Pavia U [35]. 96
4.4 Vérité terrain pour la figure Indian Pines. 97
4.5 5 premières composantes de la PCA sur l'image Salinas A [35]. 97
4.6 Vérité terrain pour la figure Indian Pines, une couleur correspondant à un *endmember*. 98
4.7 5 premières composantes de la PCA sur l'image contenant les voitures. 98
4.8 Critères SID et SAD pour des images avec des spectres hyperspectraux réels, $m = 960$,$N = 1200$, pour différents RSB et différentes méthodes. 102
4.9 Critères SID et AID pour des images avec des spectres hyperspectraux réels, $m = 960$,$N = 1200$, pour différents RSB et différentes méthodes. 102
4.10 Distance euclidienne entre $\mathbf{S}$ et $\hat{\mathbf{S}}$ pour $m = 960$,$N = 1200$ et différentes valeurs de RSB. 103
4.11 Une bande spectrale d'une scène urbaine [128], à gauche $\mathbf{Y} - \mathbf{MS}$, au milieu $\mathbf{Y}$ projeté sur le sous espace signal et à droite $\mathbf{MS}$. 104
4.12 Vérité terrain des spectres de l'image Cuprite 104
4.13 Vérité terrain des abondances de l'image Cuprite , chaque image correspondant à un *endmember*. 105
4.14 Abondances obtenues après application sur les données Cuprite de la méthode projetée. 106
4.15 Spectres obtenus après application sur les données Cuprite de la méthode projetée. 106
4.16 Comparaison entre les spectres de l'Alunite et de la Kaolinite issus de la vérité terrain et de la méthode projetée, à gauche selon la méthode projetée proposée, à droite les vérités terrains. 107
4.17 Spectres issus de la méthode robuste appliquée sur les données Cuprite. Beaucoup de réflectances sont à zéro. 107
4.18 Abondances issues de la méthode robuste appliquée sur les données Cuprite. . 108
4.19 Résultats de la méthode MVC-NMF sur les données Cuprite, résultats similaires pour tous les *endmembers*. 108
4.20 Spectres issus de la méthode PGNMF appliquée sur les données Cuprite. . . . 109
4.21 Abondances issues de la méthode PGNMF appliquée sur les données Cuprite. 109
4.22 Spectres issus de la méthode NMF appliquée sur les données Cuprite. 110
4.23 Abondances issues de la méthode NMF appliquée sur les données Cuprite. . . 110

Abréviations

RSB	*Rapport Signal sur Bruit (Signal to Noise Ratio SNR)*
SCM	Estimation empirique de la matrice de covariance (Sample Covariance Matrix)
MLE	Méthode des moindres carrés (Minimum Least Squares Error)
EMV	Estimateur du Maximum de Vraisemblance
PFA	Probabilité de Fausse Alarme
CES	Bruit symétrique elliptique complexe (Complex Elliptically Symmetric)
AID	Information de divergence entre les abondances (Abundance Information Divergence)
SAD	Distance angulaire entre les spectres (Spectral Angle Distance)
SID	Information de divergence entre les spectres (Spectral Information Divergence)
AIC	Critère d'information d'Akaike (Akaike Information Criterion)
NMF	Factorisation non-négative pour une matrice (Non-Négative Matrix Factorization)
MVC-NMF	NMF avec contrainte de volume minimum (Minimum Volume Constrained NMF)
PGNMF	NMF avec gradient projeté (Projected Gradient NMF)
p.s.	Convergence presque sûre
e.s.d.	Distribution spectrale empirique

Notations

$\mathbb{N}$	Ensemble des entiers naturels
$\mathbb{R}$	Ensemble des réels
$\mathbb{R}^\star$	Ensemble des réels différents de zéro
$\mathbb{C}$	Ensemble des nombres complexes
Im	Partie Imaginaire d'un élément complexe
Re	Partie Réelle d'un élément complexe
i	Complexe tel que $i^2 = -1$
$[\![\cdot]\!]$	Intervalle d'entiers naturels
$(\cdot)^+$	maximum entre 0 et $\cdot$
$\mathbf{v}$	Vecteurs notés en gras minuscule
$\star$	Opérateur conjugué
$\dagger$	Opérateur transposé conjugué
$\mathcal{L}$	Opérateur de Toeplitz (1.4.2)
$\mathbf{M}$	Matrices notées en gras majuscule
$\mathbf{I}_a$	Matrice identité de dimensions $a \times a$
$\mathbf{1}_N$	Vecteur colonne de 1, de taille $N \times 1$
$\mathbf{0}_{m\times m}$	Matrice de 0, de taille $m \times m$
$[\mathbf{M}]_{i,j}$	Élément i, j de la matrice $\mathbf{M}$
Tr	Opérateur de la trace
$.^T$	Opérateur de la transposée
$.^H$	Opérateur de la matrice adjointe
$\mathcal{T}$	Opérateur de "Toeplitzification" (1.4.2)
$\mathbf{M}^{1/2}$	Racine carrée hermitienne d'une matrice $\mathbf{M}$
$\|\cdot\|_F$	Norme de Frobenius
$\|\cdot\|$	Norme spectrale
$vect(\cdot)$	Opérateur de vectorisation d'une matrice
$\mathbf{M} \preceq \mathbf{N}$	La matrice $\mathbf{N} - \mathbf{M}$ est définie positive
$\sharp$	Cardinal d'un ensemble
$Supp$	Support d'une distribution ou d'une fonction
exp	Fonction exponentielle
log	Fonction logarithme
Γ	La fonction gamma
$.'$	Dérivée d'une fonction
$\mathbb{E}$	Espérance mathématique
$\mathbb{P}$	Probabilité

Introduction

Les chaînes de traitement du signal se ressemblent souvent : il s'agit d'abord d'acquérir des données, puis de séparer la partie informative de la partie inintéressante pour l'application en question. Cette partie non informative, sans intérêt, qui est ainsi appelée bruit du signal, gêne généralement le traitement des données, car le manque d'information à son égard empêche bien souvent son élimination avant traitement des informations. De même, la grande dimension des données peut être un frein à leur analyse, à la fois parce que le temps nécessaire à leur analyse peut être plus grand, mais aussi par manque de méthodes adaptées. Or, de plus en plus, les données à traiter dans différents domaines sont de volume de plus en plus important, et ceci, pour différentes raisons dont le développement d'instrumentation plus précise qui permet l'acquisition de plus de données, le besoin de transmettre plus d'informations plus rapidement, le besoin de donner des résultats de plus en plus fiables et donc de prendre en compte plus de paramètres et de données.

L'objectif de cette thèse est de proposer des méthodes permettant d'améliorer le traitement des images hyperspectrales, et ceci, sans aucun a priori sur les images en question. Cette thèse montre également la polyvalence des applications possibles des méthodes proposées. Le problème assez général de cette thèse est ainsi d'abord de séparer correctement la partie bruit de la partie intéressante d'un signal de type image hyperspectrale, puis de traiter correctement l'information véhiculée.

Les méthodes développées en théorie des matrices aléatoires, théorie particulièrement adaptée aux grand flux de données et donc aux grandes dimensions, sont adaptées au cas de l'imagerie hyperspectrale et au cas du modèle plus général considéré ici pour cette thèse. L'imagerie hyperspectrale ainsi que la théorie des matrices aléatoires sont des domaines récents en traitement du signal ; cette thèse s'applique donc à améliorer des méthodes existantes et proposer des méthodes innovantes dans un domaine en plein développement.

Cette thèse est présentée en quatre chapitres, et s'attèle à deux problèmes distincts qui peuvent être traités séparément. Le premier problème est la sélection d'ordre du modèle (notion développée dans le premier chapitre), le second le problème est celui du démélange spectral (également défini dans le premier chapitre).

Le premier chapitre dresse l'état de l'art des méthodes déjà existantes en imagerie hyperspectrale pour ces deux problèmes, et les compare entre elles. Certaines méthodes seront définitivement abandonnées pour le problème abordé dans cette thèse. Ce chapitre introduit la théorie des matrices aléatoires et donne aussi, sans entrer dans les détails, quelques éléments de cette théorie qui seront utiles pour les chapitres suivants. Il présente aussi quelques méthodes issues de cette théorie et adapté au problème de cette thèse. Il pose également le modèle choisi pour les images hyperspectrales. Le second chapitre propose des méthodes pour l'estimation d'ordre du modèle, méthodes reposant à la fois sur des éléments de la théorie des matrices aléatoires et sur des méthodes de traitement des images hyperspectrales. Un retour sur le modèle choisi sera nécessaire pour montrer que dans certains cas particuliers, il est possible d'utiliser des méthodes jusqu'alors utilisées dans d'autres domaines mais adaptables au cas de l'imagerie hyperspectrale. Trois principales méthodes seront rete-

nues pour la suite, valables pour le modèle général développé dans le premier chapitre. Des applications au cas images simulées et réelles sont regroupées dans un troisième chapitre. Enfin, un quatrième chapitre expose de nouvelles méthodes pour le problème du démélange spectral, tout en rappelant certains résultats obtenus avec des méthodes "classiques" pour une meilleure évaluation des performances des solutions proposées. De ces chapitres découle la conclusion que les méthodes proposées sont parfois meilleures que les méthodes déjà existantes, d'autant plus si certaines hypothèses sur les images hyperspectrales sont vérifiées. En annexe sont proposés des rappels mathématiques et les preuves des principaux théorèmes développés dans cette thèse.

Chapitre 1

Problèmes et méthodes en imagerie hyperspectrale

Dans ce chapitre sont présentées l'imagerie hyperspectrale, ses principales problématiques ainsi que certaines méthodes classiques de résolution associées à ces problématiques. Les limites de ces méthodes sont également montrées. Un type de modélisation particulier est développé, qui sera repris par la suite.

1.1 L'imagerie hyperspectrale

L'imagerie hyperspectrale ou spectroscopie hyperspectrale [39] est une technique d'imagerie passive qui renvoie, pour une scène spatiale considérée, une centaine d'images représentant des mesures de radiance, chacune des images étant dans une longueur d'onde différente comprise en général entre 0.4 μ m et 3 μ m (visible et proche infrarouge). Une image hyperspectrale peut donc être assimilée à un cube de données, les données étant des radiances mesurées pour chaque longueur d'onde choisie et pour chaque zone ou pixel de l'image. Les longueurs d'onde doivent, en général, être presque contiguës, souvent prises avec un pas de 10 nm, pour permettre une exploitation plus facile des données [43]. Elle se différencie de l'imagerie multispectrale ou superspectrale par le nombre de longueurs d'onde ou spectres considérés. Les images proviennent de spectromètres hyperspectraux, mesurant des valeurs de radiance issues de la réflexion du soleil, et souvent embarqués dans des engins survolant la scène étudiée, tels que des avions ou des satellites. Certaines longueurs d'onde ne peuvent ainsi pas être considérées car elles sont situées dans la zone d'absorption de l'atmosphère [70]. La résolution spatiale est, quant à elle, en général de l'ordre de la dizaine de mètre. Cette résolution dépend du champ de vision du spectromètre, lui même fonction de la taille des détecteurs, de leur résolution spatiale mais aussi de l'altitude de laquelle est prise l'image. La figure 1.1 montre un exemple d'image hyperspectrale, où x et y représentent les dimensions spatiales, tandis que sur l'axe vertical se trouvent les différentes longueurs d'ondes considérées. Sur cette image d'une zone rurale, seulement 10 bandes spectrales sont représentées pour plus de visibilité. Les dimensions spatiales de l'image sont 400×316 soit 126400 pixels.

Dans une image hyperspectrale peuvent être présents plusieurs types de matériaux, sources ou objets, appelés *endmembers*. La notion d'*endmember* reste assez subjective puisque selon la résolution du spectromètre utilisé, un objet peut être considéré comme plusieurs *endmembers* s'il est constitué de différents matériaux ou un seul. De plus, les matériaux sont soumis à une assez grande variabilité spectrale [100] : un même matériau, peut, selon les conditions sous lesquelles l'image a été acquise, renvoyer des spectres différents, ce qui est le cas par exemple de certains matériaux lorsqu'ils sont mouillés par la pluie. Cette variabilité

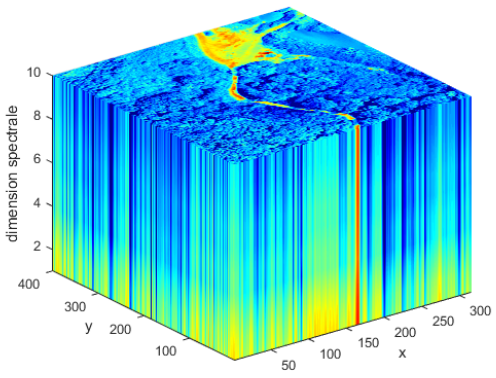

FIGURE 1.1 – Image hyperspectrale issue de la base de données du DSO National Laboratories.

spectrale est une difficulté supplémentaire pour l'interprétation des données. Mais aussi, selon la résolution spatiale du détecteur utilisé, plusieurs *endmembers* peuvent renvoyer leur radiance sur un même pixel. Il s'agit de pixels mélangés, c'est à dire que plusieurs matériaux sont présents sur un même pixel. Dans ce cas, les *endmembers* sont mélangés linéairement [58]. Il peut aussi être supposé, mais ce ne sera pas le cas ici, que plusieurs réflexions de la lumière du soleil ont lieu à l'intérieur d'un même matériau ou entre couches de matériaux avant d'être renvoyés vers le spectromètre. Il s'agit d'un modèle dit de mélange non linéaire [38], [20]. En général, en plus d'*endmembers*, un bruit s'ajoute aux données, apparu par exemple par l'action du spectromètre. Ce bruit peut être corrélé, pas forcément gaussien, de puissance plus ou moins importante par rapport au signal. Comme dans beaucoup de domaines en traitement du signal, ce bruit gène l'amélioration des performances pour beaucoup de méthodes, car souvent, aucun a priori ne peut être obtenu. La figure 1.2 présente les spectres de 6 *endmembers* présents dans une image représentant une scène urbaine ainsi que leurs abondances, c'est à dire leur répartition sur l'image. Les données de l'image sont issues de la base de données constituée dans [106].

Les images hyperspectrales sont donc des images qui contiennent un grand nombre de données et qui nécessitent des outils de traitement adaptés. Les domaines d'utilisation de ces images sont assez variés, des applications militaires (pour la surveillance de zones d'intêret par exemple) aux applications médicales en passant par l'astronomie (pour la détection et la détermination de la composition de corps) et l'agronomie (pour l'évaluation par exemple de la composition des sols) [70]. Les problèmes couramment rencontrés pour l'imagerie hyperspectrale sont l'estimation du nombre d'*endmembers* présents sur une image [62], la détection d'anomalies [83], [126], l'estimation du spectre des *endmembers* présents ou l'identification des *endmembers* ainsi que le démélange [13], qui consiste à retrouver les *endmembers* de pixels mélangés. Tous ces problèmes peuvent faire appel à de nombreuses méthodes et algorithmes de traitement du signal tels que des algorithmes de classification [21], de détection [69], de problèmes inverses [81], des méthodes de reconnaissance de forme [92], de champs de Markov [104], de machine learning [92], des méthodes adaptés à la parcimonie [122], [125], des méthodes propres aux images ou plutôt générales telles que des méthodes de traitement statistique des données [102] ou autre.

Les problèmes considérés dans ce travail sont ceux de l'estimation du nombre d'*endmembers* ou estimation d'ordre de modèle, ainsi que le problème du démélange spectral. Ces problèmes sont plutôt difficiles dans la mesure ou, pour cette thèse, aucun a priori n'est disponible en

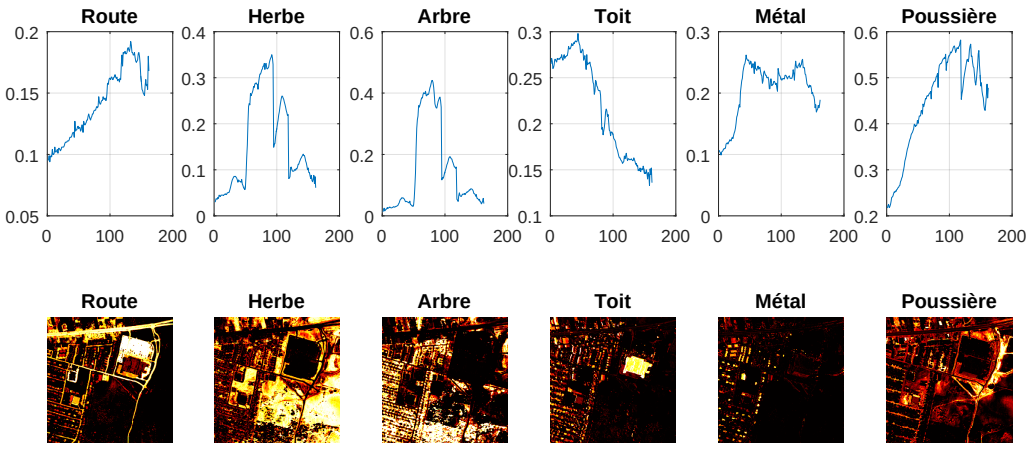

FIGURE 1.2 – Spectres et abondances pour une scène urbaine [106].

dehors des données reçues par les capteurs.

1.2 Quelques méthodes classiques pour le problème de l'estimation du nombre d'*endmembers*

Dans la littérature, le problème de l'estimation du nombre d'*endmembers* se retrouve sous le nom d'estimation d'ordre de modèle, ce qui revient souvent à l'estimation de la dimension du sous espace signal. Ce problème assez général se pose dans de nombreux domaines, tels que les communications sans fil [117], le traitement d'antenne [4], le traitement radar [25] ou bien d'autres problèmes [79] en traitement du signal ou faisant appel à des méthodes d'analyse des données, telle que la finance ou encore l'agriculture. Mais ici, il sera question de l'estimation du nombre d'*endmembers*, même si toutes les méthodes développées par la suite peuvent s'adapter à ces autres domaines.

Avant de présenter quelques méthodes utilisés fréquemment pour le problème de l'estimation du nombre d'*endmembers* en imagerie hyperspectrale, il reste à définir ce qu'est cette quantité. Dans la littérature, il est souvent fait référence à la dimension intrinsèque (*Intrinsic Dimension* ID) [19] pour décrire le nombre d'*endmembers*, qui se définit comme suit : soient N le norme de pixels de l'image hyperspectrale considérée (sur la figure 1.1, $N = \mathrm{x} \times \mathrm{y}$), $(\mathbf{y}_i)_{1 \leq i \leq N}$ les données recueillies c'est à dire les spectres obtenus pour les longueurs d'ondes considérées sur les N pixels, où, pour chaque donnée $\mathbf{y}_i$, $\mathbf{y}_i = \mathbf{d}_i + \mathbf{b}_i$ avec $\mathbf{d}_i$ les informations contenues dans le pixel i et $\mathbf{b}_i$ le bruit contenu dans ce pixel. Alors l'ID des données est égal à p, la dimension du sous-espace engendré par les $\mathbf{d}_i$, $1 \leq i \leq N$. Ceci correspond à la dimension du sous-espace signal. D'autres définitions peuvent être proposées pour l'ID, notamment [22], qui définit cette grandeur par le nombre minimal de paramètres nécessaires pour décrire les données. Comme expliqué précédemment, le nombre d'*endmembers* étant une notion plutôt subjective, une autre définition intéressante peut aussi être prise en compte, il s'agit de la dimension virtuelle (*Virtual Dimension* VD) qui se définit comme le nombre d'*endmembers* nécessaires pour proposer une solution au problème du démélange proche de la vérité. La dimension effective (*Effective Dimensionality* ED), quant à elle, est la dimension du sous-espace affine permettant de décrire correctement tous les pixels. Encore une fois, ces définitions ne reposent pas sur des critères vraiment objectifs. Ici, le nombre

d'*endmembers* est associé à l'ID et est noté p.

Il existe un grand nombre de méthodes pour estimer p. Celles-ci reposent sur des principes assez différents. Les sections suivantes présentent quelques unes des plus utilisées. Ces méthodes sont des méthodes dites "globales", approches utilisées dans un cas de mélange linéaire. Dans le cas de mélange non linéaire, qui ne sera pas étudié ici, d'autres méthodes sont envisageables, méthodes dites alors "locales".

1.2.1 Des méthodes qui reposent sur un critère d'information théorique

La méthode proposée par Akaike [2], le critère d'information d'Akaike (*Akaike Information Criterion* AIC), celle de Rissanen [95] et de Schwarz [99], le critère de la taille minimale de description (*Minimum Description Length* MDL) comptent parmi les méthodes populaires pour des signaux contaminés par un bruit blanc gaussien. Ces deux méthodes sont assez identiques dans leur principe et se déclinent comme suit. Soit $\mathbf{Y} \in \mathbb{R}^{m \times N}$ la matrice image des observations, contenant les spectres de dimensions m, c'est à dire que m bandes spectrales sont considérées, obtenus sur les N différents pixels. Alors $\mathbf{Y}$ peut se noter $[\mathbf{y}_1, \ldots, \mathbf{y}_N]$ où chaque $(\mathbf{y}_i)_{1 \leq i \leq N}$ correspond à un spectre bruité acquis. Ces vecteurs se décomposent alors en une partie contenant les *endmembers* et du bruit, supposé blanc gaussien dans un premier temps. Soit également $\mathbf{\Theta}$ un vecteur contenant les paramètres de l'image c'est à dire la matrice de covariance des données $\mathbf{Y}$, ses valeurs propres et ses vecteurs propres, dont l'estimation par maximum de vraisemblance est notée $\hat{\mathbf{\Theta}}$. Soient $\lambda_0 > \lambda_1 > \ldots > \lambda_{m-1}$ les valeurs propres de l'estimation empirique de la matrice de covariance des données, appelée SCM, alors la méthode AIC consiste à sélectionner le modèle qui correspond le plus aux données c'est à dire à estimer p de la façon suivante dans le cas où les données sont réelles :

$$\hat{p}_{AIC} = \underset{k}{\operatorname{argmin}}\, AIC(k) = \underset{k}{\operatorname{argmin}} \left[-2 \log f(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N \mid \hat{\mathbf{\Theta}}_k) + 2\,k \right], \tag{1.1}$$

où $\hat{p}_{AIC}$ est l'estimation de p, k est le nombre de degrés de liberté, f la densité de probabilité du modèle, et $\mathbf{\Theta}_k^H = \left(\lambda_0, ..., \lambda_{k-1}, \sigma^2, \mathbf{V}_0^H, ..., \mathbf{V}_{k-1}^H\right)$ avec σ^2 la variance du signal et $\left(\mathbf{V}_0^H, ..., \mathbf{V}_{k-1}^H\right)$ les vecteurs propres correspondant aux valeurs propres associées. Ainsi,

$$\hat{p}_{AIC} = \underset{k \in [0, m-1]}{\operatorname{argmin}} \left[-2\,N\,(m-k)\,\log \left(\frac{\displaystyle\prod_{i=k}^{m-1} \hat{\lambda}_i^{1/(m-k-1)}}{\dfrac{1}{m-k} \displaystyle\sum_{i=k}^{m-1} \hat{\lambda}_i} \right) + 2\,k \right]. \tag{1.2}$$

Ce critère est présenté ici dans le cas des données réelles, mais il s'adapte au cas des images dont les données sont complexes, c'est dire les images pour lesquelles $\mathbf{Y} \in \mathbb{C}^{m \times N}$ en modifiant le nombre de degrés de liberté k par $k\,(2\,m-k)+1$. Le critère est alors le même à part le dernier terme qui devient $2\,k\,(2\,m-k)$ au lieu de $2\,k$, voir [120] pour plus de précisions.

Seulement ce critère s'est montré inconsistant. Le critère MDL a été conçu pour pallier ce problème. Il diffère donc peu du critère AIC et se construit comme suit :

$$\hat{p}_{MDL} = \underset{k \in [0, m-1]}{\operatorname{argmin}} \left[-N\,(m-k)\,\log \left(\frac{\displaystyle\prod_{i=k}^{m-1} \hat{\lambda}_i^{1/(m-k-1)}}{\dfrac{1}{m-k} \displaystyle\sum_{i=k}^{m-1} \hat{\lambda}_i} \right) + \frac{1}{2} k \log N \right]. \tag{1.3}$$

Ce critère, peut, de même, s'adapter au cas complexe en changeant le dernier terme en $\frac{1}{2}k\,(2\,m-k)\log N$.

$\hat{p}_{MDL}$ représentant aussi l'estimation de p, par le critère MDL. La différence principale avec le critère AIC étant la présence du terme en $\log N$, donnant un terme de pénalité un peu plus précis.

Le dernier critère qu'il peut être intéressant de considérer est celui proposé par [94], noté RAD, particulièrement robuste au cas où le bruit serait corrélé. Ce critère ressemble aux deux autres mais recherche cette fois-ci le dernier k pour lequel il est positif et s'écrit comme suit :

$$\hat{p}_{RAD} = \underset{k\in[1,m-1]}{\operatorname{argmax}} \left\{ \frac{\lambda_{k+1}}{\sum_{i=2}^{m}\lambda_i} - \frac{\xi_k}{\sum_{i=1}^{m-1}\xi_i} > 0 \right\} . \tag{1.4}$$

avec $\xi_k = 1-\alpha\frac{(\lambda_k-\mu_k)}{\mu_k}$, $\mu_k = \frac{1}{m-k}\sum_{i=k+1}^{m}\lambda_i$ et α tel que $\alpha\max_k \frac{\lambda_k-\mu_k}{\mu_k} = 1$.

Ces méthodes donnent de bons résultats, notamment pour des images ayant un bruit gaussien, comme c'est le cas pour la figure 1.3, où les trois critères sont tracés, pour $N = 1000$ et $m = 10$, $p = 4$ avec, pour chaque *endmember*, un RSB de $20\,dB$ défini par $20\,log_{10}\left(\frac{P_d}{P_b}\right)$, P_d et P_b les puissances des *endmembers* et du bruit respectivement. Le bruit n'est pas corrélé ici, mais la méthode [94] donne de bons résultats pour un bruit gaussien non blanc. Les valeurs du critère de la méthode proposée dans [94] sont multipliées par 100 pour plus de visibilité. Pour rappel, les critères AIC et MDL renvoient pour $\hat{p}$ le plus petit $\hat{p}$ atteint, soit ici 4 sur la figure 1.3. Quant au critère RAD, il s'agit de trouver le dernier $\hat{p}$ pour lequel le critère est positif, soit ici, 4 également. Ces méthodes trouvent toutes trois le bon nombre d'*endmembers*, c'est à dire 4.

1.2.2 Des méthodes statistiques plus générales et la méthode Hysime

Une autre façon d'estimer l'ID est de considérer la matrice de corrélation ou, voire et, de covariance pour en déduire p sans passer par la minimisation d'un critère. C'est par exemple le cas de l'analyse en composantes principales (*Principal Component Analysis* PCA) [97] ou de la décomposition en valeurs singulières (*Singular Value Decomposition* SVD) où l'ID sera le nombre de valeurs propres de la matrice de covariance supérieures à un seuil. Ce seuil reste à fixer et il n'est pas toujours facile, selon la nature du bruit, de le déterminer. Ces méthodes ne fonctionnent systématiquement bien que pour des distributions de bruit connues ou pour des images sans bruit. En général, ces techniques sont utilisées pour réduire la dimension des images en projetant les données sur le sous espace signal.

Cependant, d'autres méthodes plus efficaces et robustes qui se fondent sur les mêmes principes ont été avancées. Dans cette partie sont rassemblées ces principales méthodes sans donner trop de détails pour certaines car elles ne seront pas réutilisées par la suite. Elles sont tout de même nommées et rapidement présentées ici car souvent utilisées en traitement des images hyperspectrales. De plus, cela permet au lecteur de se rendre compte de la grande diversité des méthodes existantes, et de l'intérêt d'en développer de nouvelles, si les premières se montrent inefficaces dans certaines situations.

La méthode *Maximum Noise Fraction* (MNF) a été introduite dans l'article [42] et se présente comme une alternative à la méthode PCA. Cette méthode consiste à rechercher

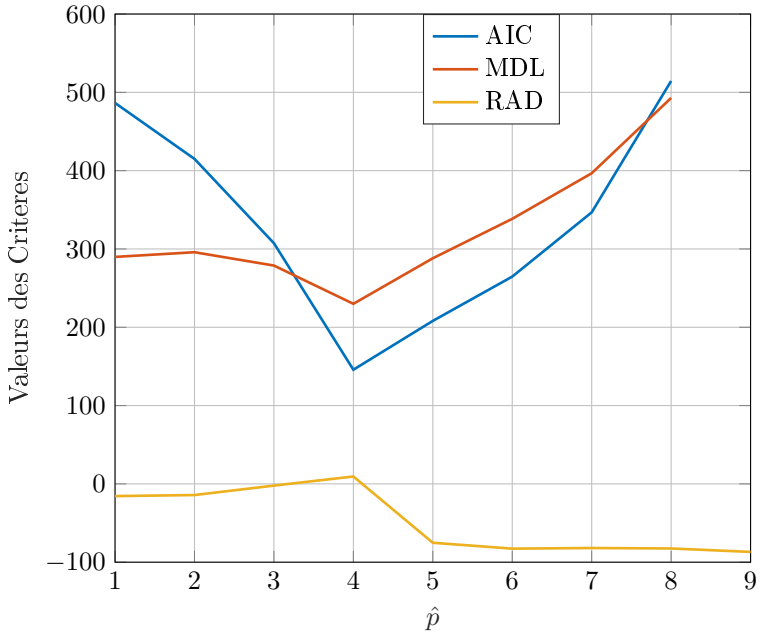

FIGURE 1.3 – Valeurs des critères AIC, MDL et de la méthode [94] pour $N = 1000$,$m = 10$, $p = 4$, RSB $= 20dB$

les vecteurs de l'image qui représentent au mieux les spectres des *endmembers* présents en cherchant à maximiser le RSB. Mais les limites de la MNF sont vite atteintes puisqu'elle nécessite la connaissance de la matrice de covariance du bruit et celle du signal.

La méthode proposée par Harsany, Farrand et Chang (HFC) [22] quant à elle, propose de considérer la matrice de covariance estimée des données ainsi que celle du bruit. Cette méthode repose sur le fait que, pour un bruit blanc de moyenne nulle, les p plus grandes valeurs propres de la matrice de corrélation sont supérieures à celles de la matrice de covariance tandis que les suivantes sont égales. Il suffit donc de choisir un seuil pour la différence entre deux valeurs propres à partir duquel ces deux valeurs propres ne sont plus considérées comme égales. Pour améliorer les performances de détection, une méthode bruit-blanchi-HFC (Noise-Whitened-HFC notée NWHFC) a ensuite été introduite. NWHFC est donc la version blanchie de HFC, où la matrice des données est blanchie.

Enfin, une méthode a été proposée par Bioucas-Dias et Nascimento [11], nommée Hysime, qui estime au sens des moindres carrés, les valeurs propres qui représentent le mieux la moyenne des données $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N]$. Un peu plus en détail, la méthode estime d'abord les matrices de covariance du bruit et du signal, notées $\mathbf{R}_d$ et $\mathbf{R}_b$, puis effectue une SVD sur $\mathbf{R}_d$ pour en déduire les vecteurs propres singuliers. $\mathbf{P}_k$, la matrice de projection sur le sous espace des vecteurs propres singuliers correspondant aux k plus grandes valeurs propres est calculée, pour $1 \leq k \leq m$. $\mathbf{P}_k^{\perp}$ représente la matrice de projection sur le sous espace orthogonal à celui associé à $\mathbf{P_k}$. p est ensuite estimé par l'équation suivante :

$$\hat{p} = \underset{1 \leq k \leq m}{\operatorname{argmin}} \; \bar{\mathbf{y}}^T \mathbf{P}_k^{\perp} \mathbf{y} + 2Tr(\mathbf{P}_k \mathbf{R}_b / N) \,, \tag{1.5}$$

où $\bar{\mathbf{y}}$ représente la moyenne des données soit $\bar{\mathbf{y}} = \sum_{i=1}^{N} \mathbf{y}_i$.

Un comparatif de certaines de ces méthodes assez populaires est proposé dans l'article [96]. Sur ce comparatif il est visible que les méthodes présentées ici ne donnent pas de bons

résultats dans le cas ou l'ID est supérieur à 20. Aucune méthode ne trouve systématiquement le bon ordre p, pour n'importe quel bruit de RSB raisonnable et n'importe quelle valeur de p. En effet, si le bruit est inconnu, les méthodes nécessitent une estimation de ses caractéristiques, estimation qui peut alors poser problème et dégrader les performances de la méthode si la méthode d'estimation choisie n'est pas adaptée. Cependant, la méthode Hysime donne de relativement bons résultats par rapport aux autres méthodes proposées ici, tandis que les méthodes HFC et NWHFC ne donnent pas de résultats constants en fonction de la forme du bruit considéré.

La méthode Hysime sera donc considérée par la suite dans certaines simulations pour comparer les résultats proposés.

1.2.3 Des méthodes géométriques

A titre informatif, il existe des méthodes qui se fondent sur des études de distance ou de volume pour estimer l'ID. Parmi elles, la méthode function d'indication empirique (*Empirical Indication Function* EIF) [68] construite à partir de l'étude des valeurs propres et l'ACP mais ne peut s'appliquer aux données corrélées.

La méthode de [49] propose d'étudier la distance entre plus proches voisins de spectres appartenant à différents pixels. Mais le plus gros inconvénient de cette méthode est que, encore une fois, la connaissance de la matrice du bruit est absolument nécessaire.

1.3 Quelques méthodes classiques pour le problème du démélange

Le démélange spectral est un problème qui vient de la structure de l'image, c'est à dire un problème lié à la façon dont est construite une matrice représentant une image hyperspectrale. Une image hyperspectrale, dont les spectres recueillis $[\mathbf{y}_1, \ldots, \mathbf{y}_N]$ sont regroupés dans une matrice image $\mathbf{Y}$, peut s'écrire, dans le cadre d'un modèle de mélange linéaire [58], [57] :

$$\mathbf{Y} = \mathbf{M}\,\mathbf{S} + \mathbf{B}\,, \tag{1.6}$$

où $\mathbf{M} \in \mathbb{R}^{m\times p}$ contient les spectres des *endmembers* et est appelée matrice de mélange, considérée déterministe, $\mathbf{S} \in \mathbb{R}^{p\times N}$ contient la proportion des *endmembers* présents dans chaque pixel ou abondances (la matrice $\mathbf{S}$ est ainsi la matrice des abondances), également considérée comme déterministe seulement dans un premier temps, et où $\mathbf{B} \in \mathbb{C}^{m\times N}$, $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_N]$, représente le bruit additif obtenu sur l'image recueillie. Le modèle choisi pour ce bruit sera plus développé par la suite. Autrement écrit, pour chaque pixel $\mathbf{y}_i$:

$$\mathbf{y}_i = \sum_{j=1}^{p} s_{i,j}\mathbf{m}_j + \mathbf{b}_i\,, \tag{1.7}$$

où $s_{i,j} = [\mathbf{S}]_{i,j}$ et $m_{i,j} = (\mathbf{m}_j)_i = [\mathbf{M}]_{i,j}$. Chaque $m_{i,j}$ correspond à une mesure de spectre et est de ce fait, positif.

Connaissant $\mathbf{Y}$, les méthodes de démélange spectral permettent de retrouver $\mathbf{M}$ et $\mathbf{S}$. Le problème posé comme tel n'admet pas de solution unique, puisque, pour toute matrice inversible positive $\mathbf{U}$, si $\mathbf{M}$ et $\mathbf{S}$ sont solutions, $\mathbf{M}\,\mathbf{U}$ et $\mathbf{U}^{-1}\,\mathbf{S}$ ou $\mathbf{M}\,\mathbf{U}^{-1}$ et $\mathbf{U}\,\mathbf{S}$ sont aussi solutions de l'équation en $\mathbf{M}$ et $\mathbf{S}$ (1.6). Il est donc nécessaire de rajouter un certain nombre de contraintes sur la matrice de mélange et celle des abondances, même si l'unicité n'est

pas forcément assurée. Les contraintes qui seront toujours considérées par la suite sont les suivantes :

$$\begin{aligned} \text{Non-négativité} &: \forall i \in [1,p]\,,\ \forall j \in [1,N]\ s_{i,j} \geq 0\,, \\ \text{Somme égale à un} &: \forall j \in [1,N] \sum_{i=1}^{p} s_{i,j} = 1\,. \end{aligned} \tag{1.8}$$

Ces contraintes sont en effet en accord avec l'interprétation physique des données choisie ici, c'est à dire une matrice $\mathbf{M}$ contenant les spectres positifs des *endmembers* et une matrice $\mathbf{S}$ contenant la proportion entre 0 et 1 de chaque *endmember* dans chaque pixel i considéré.

Ce problème est assez général et beaucoup de méthodes non spécifiques aux images hyperspectrales peuvent être utilisées. Il s'agit du problème de séparation aveugle de sources, sous contraintes, c'est à dire un problème d'optimisation sous contraintes. Pour plus de visibilité, l'ensemble des méthodes présentées dans ce travail a été divisé en deux groupes, celles utilisant des méthodes plus "statistiques" c'est à dire exploitant l'estimation de certains paramètres pour estimer $\mathbf{M}$ et $\mathbf{S}$ et celles utilisant des outils plutôt relatifs à des calculs de volumes géométriques. L'article [13] regroupe un bon nombre des méthodes possibles et certaines sont reprises ici.

La plupart de ces méthodes proposent en amont une réduction des données par projection sur un sous espace identifié, et ce, afin de réduire le temps de calcul, leur complexité et augmenter le RSB du signal reçu. La méthode Hysime, citée ci-dessus, permet par exemple ce genre de manipulation et est d'ailleurs souvent utilisée pour cette étape. Cependant, cette étape n'améliore pas les résultats en terme de consistance pour l'estimation de $\mathbf{M}$ et $\mathbf{S}$ et nécessite l'estimation correcte du sous-espace signal et du sous espace bruit, pour permettre la projection des données sur ces espaces.

Les méthodes qui reposent sur une hypothèse de parcimonie ne sont pas présentées car le modèle choisi ne comprend pas cet a priori.

1.3.1 Des méthodes statistiques

Une des méthodes les plus connues pour ce genre de problème est l'analyse en composantes indépendantes (*Independant Component Ananlysis*, ICA) [28] ; par exemple [24] propose une application en imagerie hyperspectrale. Le problème de cette méthode est qu'elle repose sur l'hypothèse forte de l'indépendance des sources. Or ici, la contrainte de la somme égale à un implique une dépendance statistique des abondances. L'article [81] traite d'ailleurs ce problème. Cette méthode est donc assez restrictive et difficilement applicable au cas de l'imagerie hyperspectrale.

Une autre famille de méthodes utilisées est celle des méthodes bayésiennes. Ces méthodes recherchent un *Maximum A Posteriori* (MAP) [10] à partir de lois à priori nécessaires sur les abondances et sur la matrice de mélange. Le bruit est aussi souvent soumis à de fortes hypothèses, comme par exemple des hypothèses de gaussianité. C'est l'inconvénient de ces méthodes. A titre d'exemple, l'article [37] introduit un algorithme bayesien d'estimation de $\mathbf{M}$ et $\mathbf{S}$ en réduisant les données en amont. D'autres articles proposent d'autres dérivées de méthodes bayésiennes, comme détaillé dans [13], qui dresse un état de l'art des approches bayésiennes pour l'imagerie hyperspectrale. Ici, il n'est fait aucun a priori, ni sur la matrice des abondances, ni sur la matrice de mélange, ce qui explique pourquoi ces méthodes ne seront pas considérées dans cette thèse.

1.3.2 Des méthodes géométriques

Les méthodes géométriques présentent, sous certains aspects, plus de flexibilité pour l'application en imagerie hyperspectrale. Il est possible de les classer en différentes catégories : des méthodes qui optimisent le problème par réduction d'une erreur quadratique, et les méthodes qui optimisent un calcul de volume sur les données.

Le premier type de méthode recherche la solution au problème général suivant :

$$\underset{\mathbf{M},\mathbf{S}}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{M}\,\mathbf{S}\|_F^2 \text{ , sous contraintes (1.8).} \tag{1.9}$$

Soit $J(\mathbf{M},\mathbf{S}) = \|\mathbf{Y} - \mathbf{M}\,\mathbf{S}\|_F^2$. Ce problème se règle en général avec une méthode de descente type descente de gradient. La méthode de factorisation par matrice non-négative (*Nonnegative Matrix Factorization* NMF) est une méthode de ce type dont la convergence a été montrée dans [63]. L'algorithme de cette méthode est décrit juste après. L'article [26] propose d'ailleurs une étude complète de cette méthode, non appliquée au cas de l'imagerie hyperspectrale. Cela dit, cette méthode est particulièrement adaptée au cas de l'imagerie hyperspectrale [77] puisqu'elle ne nécessite pas d'hypothèses particulières sur les images comme par exemple la présence de pixel non-mélangés ou dit "purs", c'est à dire constitués d'un seul *endmember*, hypothèse malheureusement souvent nécessaire pour d'autres types de méthodes. Par contre, cette méthode de descente n'étant pas convexe, elle peut converger vers des minimas locaux de la fonction à optimiser et n'assure pas la consistance de l'estimation obtenue pour les matrices de mélange et des abondances. Il est donc nécessaire de rajouter un certain nombre de contraintes comme (1.8).

Ci dessous est présenté l'algorithme général, en prenant en compte les contraintes (1.8). Les matrices $\mathbf{M}$ et $\mathbf{S}$ sont initialisées avec une valeur choisie respectant (1.8), pour donner $\mathbf{M}_0$ et $\mathbf{S}_0$. Les gradients de $J(\mathbf{M},\mathbf{S})$ sont d'abord calculés comme suit à chaque itération k, pour $0 < k \leq k_{max}$, k_{max} le nombre maximum d'itération (ce point sera développé plus tard) :

$$\frac{\partial J\,(\mathbf{M},\mathbf{S})}{\partial \mathbf{M}} = (\mathbf{M}\,\mathbf{S} - \mathbf{Y})\,\mathbf{S}^T\,, \tag{1.10}$$

$$\frac{\partial J\,(\bar{\mathbf{M}},\mathbf{S})}{\partial \mathbf{S}} = \bar{\mathbf{M}}^T\left(\bar{\mathbf{M}}\,\mathbf{S} - \bar{\mathbf{Y}}\right)\,. \tag{1.11}$$

avec, en notant $\mathbf{1}_N$ et $\mathbf{1}_p$ deux vecteurs de dimension N et p respectivement, constitués de 1, et :

$$\bar{\mathbf{M}} = \begin{pmatrix}\mathbf{M}\\ \mathbf{1}_N\end{pmatrix} \text{ et } \bar{\mathbf{Y}} = \begin{pmatrix}\mathbf{Y}\\ \mathbf{1}_p\end{pmatrix}\,. \tag{1.12}$$

Cette manipulation permet de respecter (1.8), comme utilisé dans [48], [77] ou encore [65]. Puis, pour chaque itération k chacune des matrices $\mathbf{M}$ et $\mathbf{S}$ est modifiée comme suit :

$$\mathbf{M}_k \leftarrow \mathbf{M}_{k-1} + \mu_{\mathbf{M}_{k-1}} \frac{\partial J(\mathbf{M}_{k-1},\mathbf{S}_{k-1})}{\partial \mathbf{M}}\,, \tag{1.13}$$

$$\mathbf{S}_k \leftarrow \mathbf{S}_{k-1} + \mu_{\mathbf{S}_{k-1}} \frac{\partial J(\bar{\mathbf{M}}_{\mathbf{k-1}},\mathbf{S}_{k-1})}{\partial \mathbf{S}}\,. \tag{1.14}$$

où $\mu_{\mathbf{M}}$ et $\mu_{\mathbf{S}}$ représentent les pas de descente. Plusieurs solutions sont possibles pour calculer ces pas de descente, qui doivent, pour chaque itération, permettre de diminuer $J(\mathbf{M},\mathbf{S})$. Parmi ceux possibles, l'article [63] propose des pas de descente assez classiques mais de meilleurs résultats sont obtenus en utilisant des pas qui vérifient la règle d'Armijo (voir l'annexe A.2) comme proposé dans [77]. Le nombre k_{max} peut être fixé arbitrairement ou être atteint une fois que $J(\mathbf{M},\mathbf{S})$ vaut une valeur minimale fixée arbitrairement. Pour améliorer encore les performances de cet algorithme, certains auteurs proposent d'ajouter plusieurs

autres contraintes par exemple dans les articles [65] et [51]. Même si cela améliore l'algorithme général, les temps de calculs peuvent devenir assez longs, des lois à priori sont nécessaires pour la matrice des abondances et de mélange et au final de meilleurs résultats sont obtenus pour d'autres méthodes, comme montré dans la figure 1.5.

Si une contrainte de volume sur la matrice de mélange est ajoutée à l'algorithme NMF, il est possible d'améliorer sensiblement la méthode. La nouvelle méthode est alors appelée *Minimum Volume Constraint*-NMF (MVC-NMF). Il s'agit alors non plus de minimiser $J(\mathbf{M},\mathbf{S})$ mais :

$$f(\mathbf{M},\mathbf{S}) = \frac{1}{2}J(\mathbf{M},\mathbf{S}) + \lambda\, K(\mathbf{M}) \text{ vérifiant (1.8)}, \tag{1.15}$$

où $K(\mathbf{M})$ est une fonction de pénalité sur le volume du simplex de la matrice de l'estimation de $\mathbf{M}$, $\lambda \in \mathbb{R}$ un paramètre de régularisation à fixer arbitrairement (ce qui est un point faible de la méthode). Le calcul du volume du simplex représentant la matrice $\mathbf{M}$ peut se faire comme pour [77], c'est à dire un calcul de déterminant sur la matrice de mélange réduite grâce à une ACP. Ainsi, $K(\mathbf{M}) = \dfrac{1}{2(p-1)!}\left|\begin{pmatrix}\mathbf{1}_p^T \\ \mathbf{M}\end{pmatrix}\right|^2$, avec $\tilde{\mathbf{M}} = \mathbf{U}^T\left(\mathbf{M} - \boldsymbol{\nu}\,\mathbf{1}_p^T\right)$ où $\boldsymbol{\nu}$ est un vecteur colonne contenant la moyenne sur les lignes de $\mathbf{Y}$, et où $\mathbf{U}$, de dimension $m \times (p-1)$ est une matrice qui contient les $p-1$ données les plus importantes de $\mathbf{Y}$ obtenues par PCA. L'article [77] donne le calcul complet et l'expression finale de ce volume qui n'est qu'en fonction de la matrice $\mathbf{M}$ uniquement.

Enfin, toujours pour améliorer les performances de cette méthode, l'article [64] introduit une méthode de gradient projeté pour la NMF, notée PGNMF (*Projected Gradient Methods for Non-negative Matrix Factorization*). Cette méthode a pour intérêt de converger plus rapidement que les autres méthodes, en projetant, avant chaque itération, les $\mathbf{M}_{k-1} + \mu_{\mathbf{M}_{k-1}} \frac{\partial J(\mathbf{M}_{k-1},\mathbf{S}_{k-1})}{\partial \mathbf{M}}$ et $\mathbf{S}_{k-1} + \mu_{\mathbf{S}_{k-1}} \frac{\partial J(\mathbf{M}_{\mathbf{k-1}},\mathbf{S}_{k-1})}{\partial \mathbf{S}}$ sur un espace choisi.

Le second type de méthode assez utilisé est celui des méthodes de calcul de volume. La méthode N-FINDR [123] est une des plus connues même si elle nécessite la présence de pixels purs pour chaque *endmember*. L'algorithme recherche alors les pixels qui permettent d'obtenir le volume de simplex le plus grand. Ces pixels correspondent au spectre des *endmembers* (et sont des pixels purs). Ici, cette méthode ne peut pas être utilisée telle quelle puisqu'il n'y a pas forcément de pixel purs dans les données considérées. La figure 1.4 illustre ce principe de minimisation de volume de simplex. Y sont disposés plusieurs pixels, dont trois correspondent à des pixels purs, contenant trois *endmembers* différents. Le volume maximum du simplex contenu à l'intèrieur des données est ainsi obtenu en considérant comme points extrêmes les trois vrais *endmembers*. Pour cette figure 1.4, les spectres des trois *endmembers* utilisés ont été obtenus dans [60], auxquels ont été ajoutés un bruit additif gaussien. Le nombre de bandes spectrales considéré est $m = 900$ et le nombre de pixels est 1000. De nombreuses méthodes reposent sur un calcul de volume de simplex, et améliorent la méthode N-FINDR. C'est le cas de la méthode *Vertex Component Analysis* (VCA), proposée dans [82]. Cette méthode étudie la projection des données sur le sous espace orthogonal aux *endmembers* déjà trouvés, le pixel le plus extrême après projection sera alors pris comme nouvel *endmember*.

La figure 1.5 regroupe ces méthodes et présente les résultats en termes de l'erreur quadratique moyenne $J(\mathbf{M},\mathbf{S})$. Sur cette figure, les données reçues dans la matrice image $\mathbf{Y}$ sont obtenues à l'aide d'un bruit additif gaussien et de 4 *endmembers* dont les spectres ont été récupérés sur [60]. Les RSB ne sont pas tout à fait définis comme précédemment pour des raisons de visibilité des résultats : le RSB est ici défini comme l'inverse de la puissance du bruit. Sur la figure 1.5, il varie de 1 à $\frac{1}{0.001}$. Les méthodes NMF1 et NMF2 correspondent respectivement aux méthodes proposées dans [51] et [65] la différence entre les deux étant les

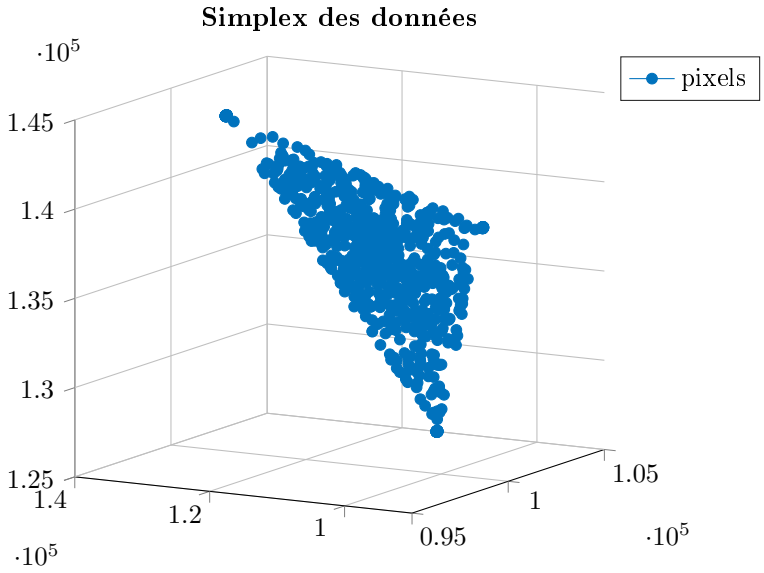

FIGURE 1.4 – Simplex obtenu avec $p = 3$, $m = 900$ et $N = 1000$, données issues de [60].

contraintes rajoutées dans le critère de minimisation $f(\mathbf{M}, \mathbf{S})$. Cette figure permet de voir que les résultats sont du même ordre de grandeur, et qu'ils dépendent du critère de minimisation choisi (visible notamment entre NMF1 et NMF2). Pour ces simulations, $N = 1000$ et $m = 96$, ce qui correspond à une petite image, ce choix étant fait pour des raisons de durée de calculs pour certaines méthodes. Pour les autres, prendre N et m plus grands ne change pas les résultats. Les résultats obtenus ont été moyennés à partir de 10 simulations.

1.4 Modélisation CES d'une image hyperspectrale

Une image hyperspectrale peut se modéliser de différentes façons. Ici, comme déjà signalé, le modèle choisi est un modèle de mélange linéaire [58], [57], c'est à dire des données mélangées linéairement et un bruit additif. Les données contenues dans $\mathbf{M}\,\mathbf{S}$ sont des données de terrain, toujours constituées des spectres des *endmembers*, rassemblés dans la matrice de mélange $\mathbf{M}$ et leur proportion dans les différents pixels dans la matrice des abondances $\mathbf{S}$. Quand au bruit noté $\mathbf{B}$ jusqu'ici, il peut être dû à différents phénomènes et donc modélisé de différentes façons. Comme aucun apriori n'est fait pour le bruit, l'approximer par un bruit gaussien pourrait sembler un peu trop restrictif. La section suivante définit ainsi un cas un peu plus général qui sera considéré par la suite.

1.4.1 Les distributions elliptiques complexes et les matrices de Toeplitz

Les distributions elliptiques ont été introduites par Kelker en 1970 dans [55], puis étudiées dans [3] et [16]. Elles sont utilisées notamment dans les applications pour des données issues de radars, mais il faudra attendre 2004 pour qu'Ollila [85] élargisse le cas de ces distributions au cas complexe, et qu'elles puissent ainsi être utilisées pour des applications plus diversifiées [86]. Les définitions suivantes permettent d'introduire d'abord les distributions elliptiques réelles avant la généralisation au cas complexe. Ces définitions plus approfondies peuvent être retrouvées dans la thèse [67]. Mais avant, voici un rappel sur les distributions gaussiennes multivariées :

Definition 1.4.1. *Distribution réelle gaussienne multivariée*

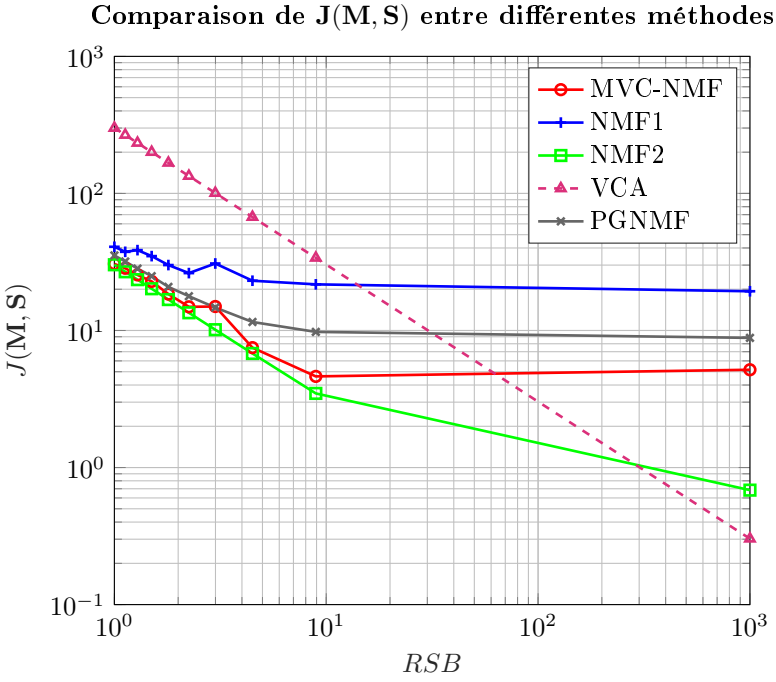

FIGURE 1.5 – $J(\mathbf{M},\mathbf{S})$ pour différentes méthodes, avec : $N = 1000$,$m = 96$, sur images simulées avec des spectres hyperspectraux de [60] et un bruit additif gaussien.

Soit $\mathbf{x} \in \mathbb{R}^m$ *un vecteur aléatoire de matrice de covariance* $\mathbf{C}$, $\mathbf{C}$ *une matrice hermitienne et d'espérance* $\boldsymbol{\mu}$. *Alors il est dit gaussien réel multivarié si sa densité de probabilité peut s'écrire :*

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{m/2}|\mathbf{C}|^{1/2}} \exp\left(-\frac{(\mathbf{x}-\boldsymbol{\mu})^T\ \mathbf{C}^{-1}\ (\mathbf{x}-\boldsymbol{\mu})}{2}\right). \tag{1.16}$$

Il est alors noté : $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$

La généralisation au cas complexe se fait comme suit, en reprenant les mêmes notations :

Definition 1.4.2. *Distribution complexe gaussienne multivariée*
Soit $\mathbf{z} \in \mathbb{C}^m$ *un vecteur aléatoire de matrice de covariance* $\mathbf{C}$, $\mathbf{C}$ *une matrice complexe hermitienne, définie positive, et d'espérance* $\boldsymbol{\mu}$. *Alors il est dit gaussien complexe multivarié si sa densité de probabilité peut s'écrire :*

$$f(\mathbf{z}) = \frac{1}{(\pi)^m|\mathbf{C}|} \exp\left(-(\mathbf{z}-\boldsymbol{\mu})^H\ \mathbf{C}^{-1}\ ((\mathbf{z}-\boldsymbol{\mu})\right). \tag{1.17}$$

Il est alors noté : $\mathbf{z} \sim \mathcal{CN}(\boldsymbol{\mu}, \mathbf{C})$

Les distributions elliptiques réelles sont des distributions qui généralisent les distributions gaussiennes multivariées réelles.

Definition 1.4.3. *Distribution elliptique réelle*
Soit $\mathbf{x} \in \mathbb{R}^m$ *un vecteur aléatoire. Alors il est dit elliptiquement distribué s'il existe un vecteur* $\boldsymbol{\mu} \in \mathbb{R}^m$, *une matrice symétrique réelle, semi-définie positive* $\mathbf{C} \in \mathbb{R}^{m\times m}$ *et une fonction* $\Psi : \mathbb{R}^+ \to \mathbb{R}$ *telle que la fonction caractéristique* $\Phi_{\mathbf{x}-\boldsymbol{\mu}}$ *de* $\mathbf{x}-\boldsymbol{\mu}$ *corresponde à* $\mathbf{t} \to \Psi(\mathbf{t}^T\,\mathbf{C}\,\mathbf{t})$ *pour tout* $\mathbf{t} \in \mathbb{R}^m$.

Il est alors noté : $\mathbf{x} \sim \mathcal{E}_m(\boldsymbol{\mu}, \mathbf{C})$.

L'article [16] apporte une définition plus appliquée de ces distributions : si $\mathbf{x} \sim \mathcal{E}_m(\boldsymbol{\mu}, \mathbf{C})$, tel que $\mathbf{C}$ est de rang $k \leq m$ alors $\mathbf{x}$ peut se décomposer, par définition, de la manière suivante :

$$\mathbf{x} = \boldsymbol{\mu} + r\boldsymbol{\Delta}\mathbf{u}^{(k)}\,, \tag{1.18}$$

où $\boldsymbol{\Delta} \in \mathbb{R}^{m\times k}$, de rang k, est telle que $\mathbf{C} = \boldsymbol{\Delta}\boldsymbol{\Delta}^T$, $\mathbf{u}^{(k)} \in \mathbb{R}^k$ est un vecteur aléatoire de distribution uniforme sur l'hypersphère unité $\mathcal{S}^{k-1}$ définie comme suit :

$$\mathcal{S}^{k-1} = \{\mathbf{x} \in \mathbb{R}^k \backslash\, \|\mathbf{x}\| = 1\}\,, \tag{1.19}$$

et $r \in \mathbb{R}$ une variable aléatoire positive indépendante de $\mathbf{u}^{(k)}$. Il est également possible de caractériser ces distributions par leur densité de probabilité, lorsqu'elle existe, c'est à dire lorsque $\mathbf{C}$ est une matrice définie positive et lorsque r a une fonction de répartition continue. Cette densité de probabilité s'écrit alors, pour tout $\mathbf{x} \neq \boldsymbol{\mu}$ et pour g, une fonction positive nommée génératrice de densité :

$$f(\mathbf{x}) = c_{m,g}\,|\mathbf{C}|^{-1/2}\, g\left((\mathbf{x}-\boldsymbol{\mu})^T\ \mathbf{C}^{-1}\ (\mathbf{x}-\boldsymbol{\mu})\right)\,, \tag{1.20}$$

où $c_{m,g}$ est une constante de normalisation qui a pour expression exacte :

$$c_{m,g} = \left(\frac{(2\,\pi)^{m/2}}{\Gamma\left(\frac{m}{2}\right)} \int_0^\infty t^{m-1}\, g\left(t^2\right)\, dt\right)^{-1}\,. \tag{1.21}$$

Les équations (1.20) et (1.16) sont assez similaires, avec, au lieu de l'exponentielle, la fonction g. Le cas complexe se généralise de la façon suivante :

Definition 1.4.4. *Distribution elliptique complexe généralisée [85]*
Soit $\mathbf{z} \in \mathbb{C}$ un vecteur aléatoire tel que $\mathbf{z} = \mathbf{x} + i\,\mathbf{y}$, $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{y} \in \mathbb{R}^m$, i tel que $i^2 = -1$. Alors $\mathbf{z}$ a une distribution elliptique complexe généralisée si et seulement si $\tilde{\mathbf{z}} = \mathrm{vect}(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{2m\times 1}$ a une distribution elliptique réelle.

La densité de probabilité de $\tilde{\mathbf{z}}$ s'écrit, pour $\mathbf{z} \neq \boldsymbol{\mu}$:

$$f_{\tilde{\mathbf{z}}}(\mathbf{z}) = c_{m,g}\,|\mathbf{C}|^{-1}\, g\left((\mathbf{z}-\boldsymbol{\mu})^H\ \mathbf{C}^{-1}\ (\mathbf{z}-\boldsymbol{\mu})\right)\,, \tag{1.22}$$

avec, cette fois ci, la matrice de covariance $\mathbf{C}$ définie de la même façon que dans [67], c'est à dire $\mathbf{C} = \begin{pmatrix}\mathbf{C}_{xx} & \mathbf{C}_{xy}\\ \mathbf{C}_{yx} & \mathbf{C}_{yy}\end{pmatrix}$ et $\boldsymbol{\mu} = \begin{pmatrix}\boldsymbol{\mu}_x\\ \boldsymbol{\mu}_y\end{pmatrix}$. Les matrices $\mathbf{C}_{xx}$, $\mathbf{C}_{yy}$, $\mathbf{C}_{xy}$ et $\mathbf{C}_{yx}$ sont des matrices réelles, et telles que $\mathbf{C}_{xy} = \mathbf{C}_{yx}^H$. Soient : $\boldsymbol{\mu} = \boldsymbol{\mu}_x + i\boldsymbol{\mu}_y$, $\boldsymbol{\Gamma} = \mathbf{C}_{xx} + \mathbf{C}_{yy} + i\,(\mathbf{C}_{yx} - \mathbf{C}_{xy})$ et $\boldsymbol{\Sigma} = \mathbf{C}_{xx} - \mathbf{C}_{yy} + i\,(\mathbf{C}_{yx} + \mathbf{C}_{xy})$, $\boldsymbol{\Gamma}$ la matrice de dispersion et $\boldsymbol{\Sigma}$ la matrice de pseudo-dispersion.

Alors la notation usuelle est : $\mathbf{z} \sim \mathcal{GCE}(\boldsymbol{\mu}_z, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}, g)$.

Voici trois exemples de ce type de distribution :

— la distribution normale multivariée, avec

$$g(t) = \exp(-t)\,, \tag{1.23}$$

et $c_{m,g} = \pi^{-m}$,

— la t-distribution multivariée, à μ degrés de libertés, où

$$g(t) = \left(1 + \frac{2t}{\mu}\right)^{-\frac{2m+\mu}{2}}\,, \tag{1.24}$$

et $c_{m,g} = \dfrac{2^m\,\Gamma\left(\frac{2\,m+\mu}{2}\right)}{(\pi\,\mu)^m\,\Gamma(\mu/2)}$. Pour cette distribution, prendre $\mu = 1$ renvoie à la distribution de Cauchy tandis que $\mu \to \infty$ correspond à la distribution gaussienne multivariée.

— la K-distribution multivariée de paramètre $\mu > 0$:

$$g(t) = t^{(\mu-m)/2}\, K_{\mu-m}\left(2\sqrt{\mu t}\right) , \tag{1.25}$$

$K_{\mu-m}(.)$ la fonction de Bessel modifiée de second type d'ordre $\mu - m$, et $c_{m,g} = \frac{2\mu^{(\mu+m)/2}}{\Gamma(\mu)\pi^m}$.

Il est à noter qu'un théorème établi dans l'article [85] montre la proportionalité entre la matrice de dispersion $\mathbf{\Gamma}$ et la matrice de covariance $\mathbf{C}$ de $\mathbf{z}$. La densité de probabilité de $\mathbf{z}$ peut se déduire de celle de $\tilde{\mathbf{z}}$ comme proposé dans l'article [67]. Une sous-classe de ces distributions est proposée dans l'article [61], qui est souvent utilisée dans les applications en traitement du signal. Il s'agit des distributions elliptiques complexes du second ordre ou circulaires, dont un cas particulier sera ici utilisé par la suite.

Definition 1.4.5. *Distribution elliptique complexe du second ordre [61]*
Ce sont les distributions telles que $\mathbf{z} \sim \mathcal{GCE}(\boldsymbol{\mu}_z, \mathbf{\Gamma}, \mathbf{0}_{m\times m}, g)$. *Dans ce cas, il est possible d'écrire simplement la densité de probabilité de* $\mathbf{z}$ *:*

$$f(\mathbf{z}) = c_{2m,g}\, 2^m\, |\mathbf{\Sigma}|^{-1}\, g\left[2(\mathbf{z}-\boldsymbol{\mu})^H \mathbf{C}^{-1}(\mathbf{z}-\boldsymbol{\mu})\right] . \tag{1.26}$$

Enfin, une dernière définition importante pour la suite est la définition d'un cas particulier qui caractérise les distributions elliptiques complexes du second ordre. Il s'agit des distributions gaussiennes composées. Plusieurs résultats sur ces distributions sont donnés dans [52].

Definition 1.4.6. *Distribution complexe gaussienne composée*
Soit $\mathbf{z} \in \mathbb{C}^m$ *un vecteur aléatoire qui suit une distribution elliptique complexe du second ordre, de moyenne nulle, de matrice de covariance* $\mathbf{C}$, $\mathbf{C}$ *une matrice définie positive et* τ *une variable aléatoire positive de densité de probabilité* p_τ. *Alors* $\mathbf{b} = \sqrt{\tau}\mathbf{z}$ *a une distribution gaussienne composée. La densité de probabilité de* $\mathbf{b}$ *peut s'écrire :*

$$f_{\mathbf{b}}(\mathbf{b}) = \frac{1}{\pi^m\, |\mathbf{C}|}\, h_m\left(\mathbf{b}^H\, \mathbf{C}^{-1}\, \mathbf{b}\right) , \tag{1.27}$$

où $h_m\left(\mathbf{b}^H\, \mathbf{C}^{-1}\, \mathbf{b}\right) = \displaystyle\int_0^{+\infty} t^{-m}\, \exp\left(-\frac{\mathbf{b}^H\, \mathbf{C}^{-1}\, \mathbf{b}}{t}\right) p_\tau(t)\, dt$, $\mathbf{z}$ *est appelé speckle et* τ *texture.*

Il est intéressant de noter que d'autres distributions peuvent être obtenues à partir de cette famille de distributions, comme par exemple :

— la distribution gaussienne multivariée complexe : il suffit de choisir pour δ la fonction de Dirac :

$$p_\tau(t) = \delta(t-1) , \tag{1.28}$$

— la distribution de Laplace de paramètre $b > 0$, obtenue pour

$$p_\tau(t) = \frac{b^2}{2}\, \exp\left(-\frac{b^2 t}{2}\right) , \tag{1.29}$$

c'est à dire obtenue lorsque τ suit une loi exponentielle de paramètre $\dfrac{b^2}{2}$.

— la distribution de Cauchy, obtenue pour

$$p_\tau(t) = \frac{b^2}{2}\, t^{-2}\, \exp\left(-\frac{b^2 t}{2}\right) . \tag{1.30}$$

D'autres exemples sont donnés dans [52]. Certaines distributions peuvent également être déduites de ces distributions gaussiennes composées sans pour autant que la fonction p_τ n'ait une expression explicite. Ces distributions gaussiennes composées présentent donc l'avantage de regrouper un certain nombre de distributions connues et utilisées pour modéliser différents types de bruits. Elles seront ainsi utilisées par la suite.

Pour une variable aléatoire $\mathbf{b}$ suivant une distribution gaussienne composée complexe, il est donc possible de noter $\mathbf{b} = \sqrt{\tau}\,\mathbf{z}$ où $\mathbf{z}$ est un vecteur complexe de matrice de covariance $\mathbf{C}$. Par la suite, $\mathbf{z}$ sera pris de moyenne nulle et de matrice de covariance égale à l'identité. La corrélation des données contenues dans $\mathbf{b}$ sera ainsi traduite comme suit : $\mathbf{b} = \sqrt{\tau}\,\mathbf{C}^{1/2}\mathbf{z}$. Une image hyperspectrale sera alors modélisée avec une matrice de bruit $\mathbf{B}$ de une distribution CES. Le bruit $\mathbf{B}$ peut donc s'écrire :

$$\mathbf{B} = \mathbf{C}^{1/2}\,\mathbf{X}\,\mathbf{T}^{1/2}\,, \tag{1.31}$$

avec quelques hypothèses et définitions supplémentaires que voici :

Hypothèse 1

- $\mathbf{T} \in \mathbb{R}^{N\times N}$ est une matrice diagonale, contenant sur sa diagonale, pour tout $i \in [\![1, N]\!]$, les τ_i correspondant aux τ de la distribution gaussienne composée pour chaque pixel.
- $\mathbf{X} \in \mathbb{R}^{m\times N}$ est un bruit d'entrées indépendantes et identiquement distribuées, de moyennes zéro et de variances unité.

Hypothèse 2

- $\mathbf{C} \in \mathbb{C}^{m\times m}$ est une matrice de Toeplitz (voir [41] pour plus de détails sur les matrices de Toeplitz et ci-après).

Ce dernier point est développé par la suite. Cette hypothèse permet d'obtenir dans les chapitres suivants une estimation plus consistante de la matrice $\mathbf{C}$ et ne dégrade pas les résultats obtenus sur signaux réels.

Les matrices de Toeplitz sont des matrices dont les éléments sur une même diagonale sont tous identiques. Leurs propriétés ont été largement exploitées (voir [41] pour une étude globale de ces matrices, [40] pour l'étude de la distribution de leurs valeurs propres) , et sont utilisées pour de nombreuses applications en traitement du signal, telles par exemple que des applications pour les séries temporelles [14], pour l'estimation de canaux parcimonieux [47], pour la détection de sources [119] ou encore pour l'estimation d'angles d'arrivées [1]. Ici, $\mathbf{C}$ sera supposée de Toeplitz, hypothèse discutable mais déjà utilisée pour des images hyperspectrales par exemple dans la thèse [93], mais cette hypothèse ne sera pas toujours nécessaire dans les différents chapitres et sections. Cette hypothèse vient de l'utilisation possible de certains résultats théoriques non utilisables sinon, et d'une certaine réalité physique, comme vont le montrer les simulations sur images réelles, simulations qui ne sont généralement pas affectées par cette hypothèse qui conduit parfois à une amélioration des résultats.

1.4.2 Récapitulatif sur le modèle et les hypothèses

Voici quelques notations supplémentaires : pour tout vecteur $\mathbf{x}$ de taille m, $\mathcal{L} : \mathbf{x} \to \mathcal{L}(\mathbf{x})$ est la matrice $m \times m$ définie comme l'opérateur de Toeplitz : $\forall i, j \in [\![1, m]\!]$, $([\mathcal{L}(\mathbf{x})]_{i,j})_{i\leq j} = x_{i-j}$ et $([\mathcal{L}(\mathbf{x})]_{i,j})_{i>j} = x^\star_{i-j}$. Pour toute matrice $\mathbf{A}$ de taille $m \times m$, $\mathcal{T}(\mathbf{A})$ représente la matrice $\mathcal{L}(\check{\mathbf{a}})$ où $\check{\mathbf{a}}$ est le vecteur dont chaque composant $\check{\mathbf{a}}_{i,\,1\leq i\leq m}$ contient la somme de la

$i-$ième diagonale de $\mathbf{A}$ divisée par m. C'est à dire :

$$\mathcal{T}(\mathbf{A}) = \frac{1}{m} \begin{pmatrix} \sum_{i=j} [\mathbf{A}]_{i,j} & \sum_{j=i+1} [\mathbf{A}]_{i,j} & \dots & [\mathbf{A}]_{1,m} \\ \sum_{i=j+1} [\mathbf{A}]_{i,j} & \sum_{i=j} [\mathbf{A}]_{i,j} & \ddots & \\ & \ddots & \ddots & \vdots \\ & & & \sum_{i=j} [\mathbf{A}]_{i,j} \end{pmatrix} . \tag{1.32}$$

Cet opérateur $\mathcal{T}$ peut être vu comme une opération de "toeplitzification" (rectification) consistant à rendre une matrice $\mathbf{A}$ de Toeplitz.

Pour des données hyperspectrales $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_N]$ de taille $m \times N$ contenant N observations $\{\mathbf{y}_i\}_{i \in [\![1,N]\!]}$ de taille m constituées de p *endmembers* mélangés auxquels s'ajoute un bruit additif, le modèle choisi est donc le modèle suivant :

$$\mathbf{y}_i = \sum_{j=1}^{p} s_{i,j} \mathbf{m}_j + \sqrt{\tau_i}\, \mathbf{C}^{1/2}\, \mathbf{x}_i \,, \quad i \in [\![1, N]\!] \,, \tag{1.33}$$

où, autrement écrit :

$$\mathbf{Y} = \mathbf{M}\,\mathbf{S} + \mathbf{C}^{1/2}\,\mathbf{X}\,\mathbf{T}^{1/2} \,. \tag{1.34}$$

La matrice $\mathbf{C}$, supposée hermitienne, non-négative et de Toeplitz, est telle que : $\mathbf{C} = \mathcal{L}\left([c_1, \dots, c_m]^T\right)$, où encore :

$$\mathbf{C} = \begin{pmatrix} c_1 & c_2 & \dots & c_m \\ c_2^\star & c_1 & \dots & c_{m-1} \\ & & \ddots & \\ c_m^\star & c_{m-1}^\star & \dots & c_1 \end{pmatrix} .$$

Et ce, sous les hypothèses **1** et **2** (1.4.1), auxquelles est rajoutée l'hypothèse suivante :

Hypothèse 3 :

- m et N sont supposés grands, et même tels que $N \to \infty$, $m \to \infty$ avec $c_N = \dfrac{m}{N} \underset{N \to \infty}{\to} c > 0$.

Le modèle choisi est donc celui d'un bruit CES corrélé par une matrice $\mathbf{C}$, dans le cadre du régime des grandes dimensions (hypothèse **3**). Ce régime des grandes dimensions signifie en pratique que m et N sont grands, et du même ordre de grandeur. Ce modèle se justifie plutôt bien dans la pratique et englobe plus de cas que les modèles reposant sur un bruit gaussien, blanc, connu ou encore les modèles qui ne prennent pas en compte le bruit. La plus grande difficulté est ici que seule la matrice $\mathbf{Y}$ est supposée connue, et que ni le bruit, ni p, ni les matrices $\mathbf{M}$, $\mathbf{S}$ et leurs dimensions ne sont connus.

1.5 Limite des méthodes proposées et motivations pour la théorie des matrices aléatoires

Certaines des hypothèses faites sur le modèle choisi ne permettent pas d'utiliser des méthodes classiques présentées précédemment, puisque les conditions d'application ne sont pas toutes vérifiées.

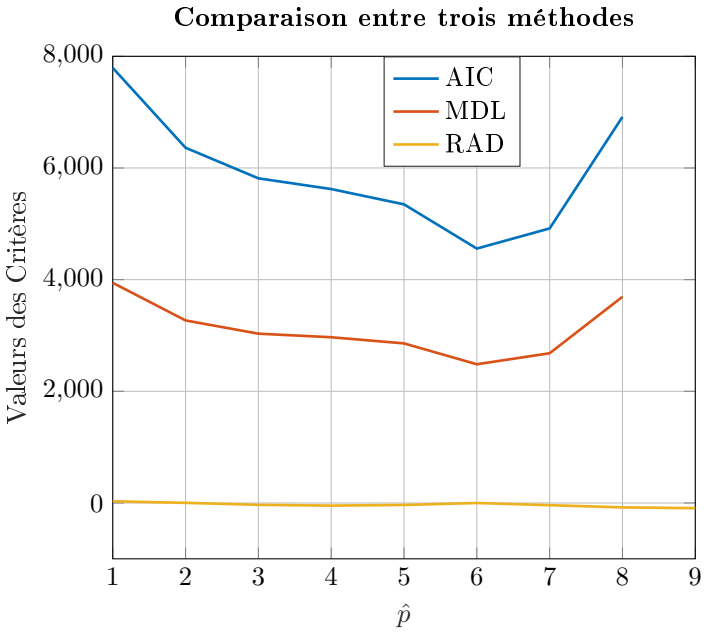

FIGURE 1.6 – $N = 1000$, $m = 10$, $p = 4$, bruit CES.

1.5.1 Méthodes d'estimation du nombre d'*endmembers* qui reposent sur un critère d'information théorique

Les méthodes d'estimation du nombre d'*endmembers* qui reposent sur un critère d'information théorique ne peuvent pas être utilisées dans le cas d'un signal corrélé en grandes dimensions. En effet, la méthode AIC tend à surestimer le nombre d'*endmembers* dans le cas des grandes dimensions, et les trois méthodes n'estiment plus correctement le nombre d'*endmembers* pour le cas d'un bruit CES, même pour des faibles dimensions. La figure 1.6 montre que pour les mêmes données que sur la figure 1.3, l'ajout d'un bruit CES à la place d'un bruit gaussien fausse les résultats : la méthode RAD [94] renvoie un nombre d'*endmembers* égal à 1, tandis que les méthodes AIC et MDL surestiment cet ordre à $\hat{p} = 6$.

Les auteurs de [114] ont travaillé sur ce problème et proposent de rajouter un terme de pénalité à fixer pour rendre la méthode AIC consistante dans le cadre de l'hypothèse 3 (1.4.2). En effet, la consistance des estimateurs qui utilisent un critère d'information théorique provient du fait que, lorsque $N \to \infty$ avec m constant, la SCM converge presque surement vers la matrice de covariance réelle en $\sqrt{\frac{\log(\log(N))}{N}}$. Or ceci n'est plus vrai sous l'hypothèse 3 (1.4.2) comme présenté dans la partie introduisant les matrices aléatoires. Les méthodes AIC, MDL et RAD ne sont donc plus applicables en grandes dimensions sans ajouter un terme de pénalité assurant leur consistance. Mais ce terme ne permet pas d'obtenir de bons résultats pour un bruit CES. Quand à la méthode RAD, même si elle est plus robuste au cas de la corrélation du bruit, elle ne donne pas de bons résultats pour le bruit CES de la figure 1.6, et ne s'applique pas non plus au régime des grandes dimensions.

1.5.2 Méthodes statistiques d'estimation du nombre d'*endmembers*

La méthode Hysime [12], qui donne de bons résultats dans le cas où $N \to \infty$ pour m fixé, ne permet pas de retrouver le bon nombre d'*endmembers* sous l'hypothèse 3 (1.4.2) lorsque le bruit est CES. La simulation de la figure 1.7 montre les performances de la méthode Hysime

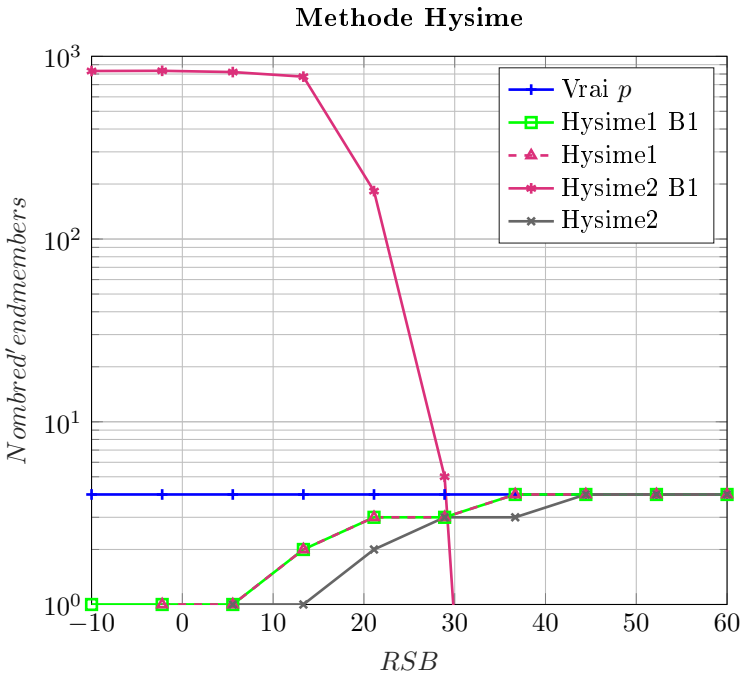

FIGURE 1.7 – Performances de la méthode Hysime pour différents cas de figure.

pour les cas suivant : Hysime1 est un cas de faible dimension, défini par $N = 1000$, $m = 96$ avec un bruit gaussien, Hysime2 est un cas de dimension plus importante avec un bruit CES, pour $N = 1000$ et $m = 960$. Le bruit est soit estimé comme proposé par J.Nascimento et J. Bioucas-Dias dans [12], dans ce cas la méthode est annotée d'un "B1", soit le vrai bruit est mis en entrée de l'algorithme Hysime. Le vrai p est indiqué en bleu.

Que le vrai bruit soit fourni en entrée de l'algorithme ou non, les résultats sont bien plus mauvais en grande dimension avec un bruit CES (Hysime2). De plus, le bruit n'est en général pas disponible et la méthode de [12] ne permet pas de l'estimer correctement dans le cas CES de grande dimension, comme visible sur la courbe de la méthode Hysime2 B1 qui trouve $\hat{p} = 0$ pour des forts RSB, alors que pour la méthode Hysime1, la figure montre bien que le bruit est correctement estimé et que la méthode est consistante pour des RSB plus faibles que Hysime2. Les courbes Hysime1 et Hysime1 B1 se superposent, montrant que le bruit est correctement estimé. Dans un cas pratique, sur une image réelle, il n'est donc pas possible d'estimer correctement le nombre d'*endmembers* avec cette méthode si l'image réelle se rapproche plus du modèle proposé ici que du modèle gaussien en faible dimension. En effet, dès que le bruit n'est pas suffisamment bien estimé ou si les RSB sont trop faibles, la méthode ne trouvera pas le bon nombre d'*endmembers*.

1.5.3 Méthodes de démélange géométriques

Les méthodes de démélange qui reposent sur un principe de minimisation de volume ne peuvent pas être utilisées telles quelles. En effet, dans le cas où l'image hyperspectrale est corrélée, alors le simplex représentant les différents pixels de l'image ne permet pas de localiser les *endmembers*. Ceci est visible sur la figure 1.8, sur laquelle, à droite, sont représentées les mêmes données que sur la figure 1.4 mis à part le bruit qui est corrélé fortement. La figure de gauche montre un cas moins extrême pour lequel le bruit est corrélé avec la matrice $\mathbf{W} = \mathcal{L}(\rho_{\mathbf{0}}, \ldots, \rho_{\mathbf{m-1}})$, pour $\rho_i = 0.7^i$, $i \in [\![0, m-1]\!]$. Sur ces figures, les pixels sont beaucoup plus éparpillés et les méthodes de maximisation de volume échouent.

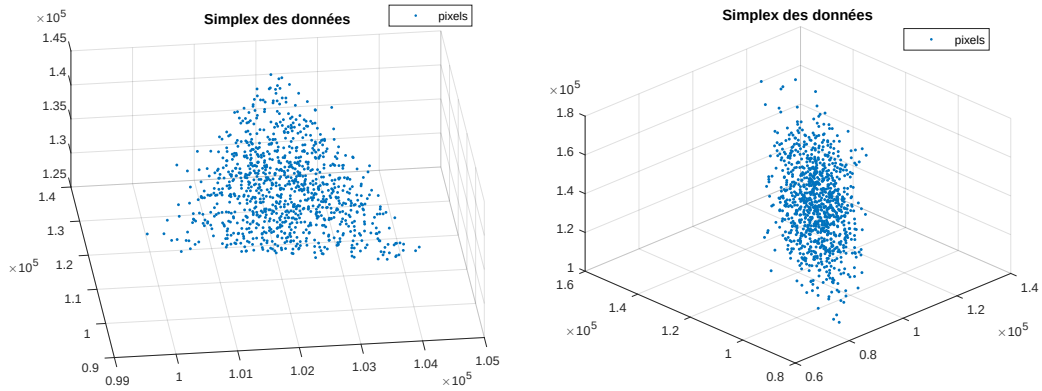

FIGURE 1.8 – Représentation des données dans un simplex, $N = 1000$, $m = 900$, données corrélées.

La figure 1.9 suivante montre l'évolution de $J(\mathbf{M}, \mathbf{S})$ en fonction du rapport signal sur bruit pour les différentes méthodes présentées précédemment. Il est à noter que l'augmentation de la dimension dégrade les performances des méthodes NMF. Sur la figure 1.9, le bruit simulé est un bruit CES, tel que les τ_i suivent une loi inverse gamma de paramètre 1.1 et la matrice de corrélation est telle que $\mathbf{W} = \mathcal{L}(\rho_{\mathbf{0}}, \ldots, \rho_{\mathbf{m-1}})$, pour $\rho_i = 0.99^i$, $i \in [\![0, m-1]\!]$. Le bruit est donc fortement corrélé. Les dimensions choisies sont $N = 1000$ et $m = 960$. Les sources sont également issues de la base de donnée en ligne [60], et les RSB, définis comme pour la figure 1.5, varient également entre 1 et $\frac{1}{0.001}$. Les performances des estimateurs sont dégradées, sauf pour la méthode VCA, qui renvoie des résultats assez similaires. Pour les autres méthodes, l'erreur calculée est assez importante, d'autant plus que le nombre d'*endmembers* était supposé a priori connu. Les différentes contraintes rajoutées sur le volume notamment, n'apportent rien de plus par rapport à la méthode VCA, puisque ces contraintes n'ont plus le même sens au vu de la figure 1.8.

Enfin, les deux graphes de la figure 1.10 montrent les performances d'estimation de la matrice $\mathbf{M}$, c'est à dire $\left\|\hat{\mathbf{M}} - \mathbf{M}\right\|_F$. Pour la première figure, le bruit est gaussien, $N = 1000$ et $m = 96$, tandis que sur la seconde, le bruit est un bruit CES, de mêmes paramètres (RSB, forme du bruit) que pour la figure 1.9. Les résultats montrés sur ces figures sont les moyennes des résultats obtenus pour 10 itérations. Au vu de la définition du RSB, les résultats doivent rester plus ou moins constants selon le RSB. Sur la figure de droite, la méthode NMF2 a été retirée car donnait de trop mauvais résultats.

Sur les figures 1.10, les ordres des méthodes donnant les meilleures performances sont changés par rapport à la figure 1.9. Ainsi, la méthode VCA, donnant les meilleurs résultats pour la minimisation du critère, ne donne plus les meilleurs résultats en ce qui concerne l'estimation de $\mathbf{M}$.

Ainsi, plusieurs éléments sont à considérer : la bonne minimisation du critère mais aussi l'estimation correcte des matrice $\mathbf{M}$ et $\mathbf{S}$. La meilleure méthode serait celle qui vérifie ces éléments.

1.5.4 Motivations pour la théorie des matrices aléatoires

Les images hyperspectrales, comme présentées précédemment, sont des signaux de grande dimension qui peuvent être corrélées spatialement. Certaines méthodes, pourtant reconnues

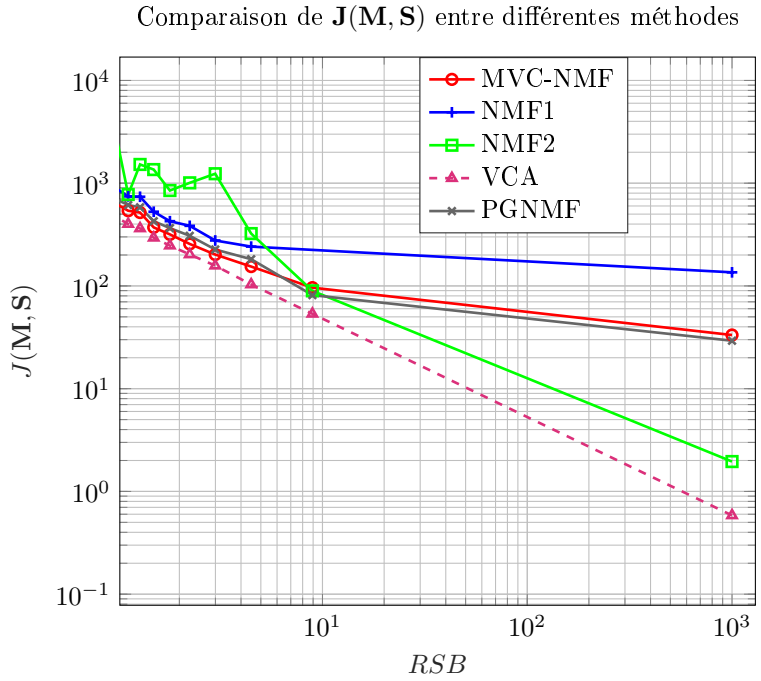

FIGURE 1.9 – $J(\mathbf{M},\mathbf{S})$ pour différentes méthodes, avec $N = 1000$ et $m = 960$, sur images simulées avec des spectres hyperspectraux de [60] et un bruit additif CES

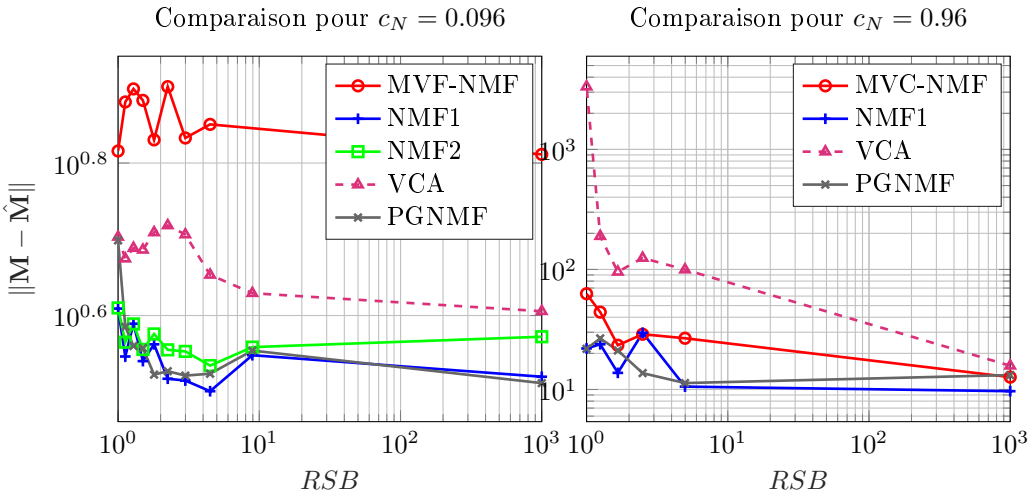

FIGURE 1.10 – Comparaison entre les différentes méthodes de $\left\|\hat{\mathbf{M}} - \mathbf{M}\right\|_F$.

pour leur efficacité, ne le sont plus pour résoudre les problèmes fréquemment rencontrés présentés précédemment si l'on considère une image hyperspectrale plus complexe, sans faire d'hypothèses parfois restrictives. Une autre solution serait de considérer un très grand nombre de pixels pour ne plus être sous l'hypothèse 3 (1.4.2) mais les temps de traitements seraient beaucoup plus longs et le nombre d'*endmembers* risque d'être autant plus important que le nombre de pixels, lié à la surface de terrain prise en compte dans l'image, est grand. Une dernière solution souvent adoptée est la réduction des données, par projection par exemple, comme le fait E.R. Manolakis dans son article [71], où les données sont réduites pour faciliter leur traitement.

La théorie des matrices aléatoires permet de caractériser des variables aléatoires de grande dimension beaucoup plus précisément que par ces méthodes classiques. Récemment introduite vers 2010 pour l'imagerie hyperspectrale, par exemple avec l'article [18], cette théorie présente l'avantage par rapport aux autres méthodes de prendre en compte la grande dimension des images hyperspectrales sans chercher à la réduire. Elle apporte de nouveaux outils statistiques adaptés, et reste encore à développer pour certains domaines. En effet, celle-ci ne s'applique pas seulement au domaine de l'imagerie hyperspectrale, puisque initialement développée pour la physique nucléaire [121], elle se démocratise pour tout type de domaine nécessitant l'analyse d'un grand nombre de données, telle que la biologie [5] ou les télécommunications sans fil [30].

Pour l'application en imagerie hyperspectrale, l'intérêt est l'étude des valeurs propres de matrices de covariance pour estimer correctement le sous espace des *endmembers*. La théorie des matrices aléatoires permet ce genre d'étude, ce qui a motivé le sujet de cette thèse, cherchant à développer des méthodes faisant appel à des outils statistiques de la théorie des matrices aléatoires pour l'imagerie hyperspectrale.

1.6 La théorie des matrices aléatoires

1.6.1 Introduction à la théorie des matrices aléatoires

La théorie des matrices aléatoires propose des outils théoriques permettant de développer des méthodes d'estimation adaptées au cas de l'imagerie hyperspectrale. Cette théorie traite des matrices, ici de grande dimension, dont les entrées sont des variables aléatoires, définies rigoureusement par :

Definition 1.6.1. *Matrice aléatoire*
Une matrice aléatoire $\mathbf{X}$ *est une matrice considérée comme une variable aléatoire d'un espace probabilisé* (Ω, C, Pr)*, Pr une mesure de probabilité, dont les entrées sont des variables aléatoires dans un espace mesurable.*
Une réalisation de $\mathbf{X}$ *pour un point* $\omega \in \Omega$ *notée rigoureusement* $\mathbf{X}(\omega)$*, sera par la suite notée* $\mathbf{X}$ *en omettant de parler du point de réalisation, et ce, pour simplifier les écritures.*

Cette théorie ne fut développée que récemment, d'abord par les travaux de J. Wishart en 1928 [124], qui étudia les matrices de covariances empiriques de matrices aléatoires, dont notamment, largement étudiées, les matrices de Wishart. Ces matrices peuvent s'écrire comme le produit $\mathbf{W}\mathbf{W}^H$ d'une matrice $\mathbf{W} \in \mathbb{C}^{m\times m}$ dont les colonnes sont des vecteurs indépendants et de moyenne nulle. En ce qui concerne les matrices de grande dimension, elles furent introduites pour la physique nucléaire et la mécanique quantique dans l'étude de particules par E. Wigner en 1955 [121]. Ce fut alors l'occasion d'introduire les matrices de Wigner, elles aussi largement étudiées par la suite. Depuis, l'émergence de nouveaux domaines et méthodes en traitement du signal a souvent conduit à l'acquisition d'un grand nombre de données à traiter. De nouveaux outils statistiques ont ainsi été développés à partir de ces travaux de recherche afin d'analyser correctement ces observations dont le nombre N

et la dimension m sont grands. Ces nouveaux outils sont utiles dans de nombreux domaines tels que les communications et le traitement d'antennes [27], le traitement radar [89], la finance [98], et pour des problèmes plus généraux tels que l'estimation du sous-espace signal [45] et la détection de sources [62]. En ce qui concerne l'imagerie hyperspectrale, elle n'y est encore que très peu utilisée, et ceci fera l'objet d'une section suivante.

L'intérêt de la théorie des matrices aléatoires appliquée à l'imagerie hyperspectrale est qu'elle permet l'étude des valeurs propres de matrices de grande dimension, permettant, notamment, l'estimation du sous espace signal. Les images hyperspectrales pouvant être modélisées par des matrices de grande dimension, cette théorie semble ainsi particulièrement adaptée. Elle nécessite ainsi l'estimation de la matrice de covariance du signal, comme pour les méthodes classiques présentées en 1.2, pour pouvoir ensuite estimer les valeurs propres et vecteurs propres associés au sous-espace signal. Or, comme montré dans la section précédente, en régime classique, c'est à dire lorsque N tend vers l'infini, pour m fixé, les estimations de la dimension du sous espace signal ou les estimations de $\mathbf{M}$ et $\mathbf{S}$ sont consistantes. Mais lorsque N et m tendent vers l'infini avec un rapport constant, ces méthodes ne sont plus consistantes. Le régime des grandes dimensions étudié par la théorie des matrices aléatoires donne des résultats différents pour ce type de données, et propose une interprétation des résultats en lien avec la structure des données. En pratique, le régime des grandes dimensions signifie que les grandeurs N et m sont toutes deux proches en ordre de grandeur, alors que les méthodes présentées précédemment s'utilisent plutôt pour des signaux tels que la dimension N est grande, et surtout importante par rapport à la dimension des données m.

1.6.2 Quelques éléments de base en théorie des matrices aléatoires

La théorie des matrices aléatoires a de nombreuses applications, mais ici la partie utile de cette théorie est l'étude de la matrice de covariance estimée $\hat{\mathbf{C}}$, sous l'hypothèse 3 (1.4.2). Il est en particulier intéressant de regarder la distribution des valeurs propres de cette matrice, pour distinguer, lorsque cela est possible, les valeurs propres issues du bruit et celles des *endmembers*. Ainsi, cette partie et les parties suivantes s'appliquent à présenter quelques résultats sur la distribution des valeurs propres, en sachant que l'intérêt est de pouvoir ensuite séparer les valeurs propres du bruit et celle du signal. La distribution des valeurs propres des matrices de covariance pour un signal constitué uniquement de bruit doit donc être étudiée, pour ensuite pouvoir les distinguer de celles du signal.

Cette étude a déjà été menée, pour différents modèles de matrices de covariance, comme par exemple dans les articles [30], [75], [76] ou encore [31]. Ces études reposent sur la loi de Marchenko-Pastur [72], développée en 1967 par Marchenko et Pastur qui peut être introduite de la manière suivante : soient N vecteurs d'observations $[\mathbf{y}_1, \ldots, \mathbf{y}_N] \in \mathbb{R}$ de taille m avec, pour tout i de l'intervalle $[\![1, N]\!]$, $\mathbf{y}_i$ indépendants et identiquement distribuées, suivant une loi normale de moyenne nulle et de matrice de covariance $\mathbf{C} \in \mathbb{C}^{m \times m}$. Alors la loi des grands nombres stipule que :

$$\left\| \frac{1}{N} \sum_{i=1}^{N} \mathbf{y}_i \, \mathbf{y}_i^T - \mathbf{C} \right\| \overset{p.s.}{\to} 0 \text{ lorsque } N \to \infty \, m \text{ fixé.} \tag{1.35}$$

Maintenant, dans le cadre de l'hypothèse 3, alors l'équation (1.35) n'est plus valable :

$$\left\| \frac{1}{N} \sum_{i=1}^{N} \mathbf{y}_i \, \mathbf{y}_i^T - \mathbf{C} \right\| \overset{p.s.}{\not\to} 0 \text{ lorsque } N \to \infty \text{ et } m \to \infty \text{ avec } \frac{m}{N} \to c, \, c \in \mathbb{R}^+, \tag{1.36}$$

même si, terme à terme, chaque élément de la SCM, soit, de la matrice $\sum_{i=1}^{N} \mathbf{y}_i \mathbf{y}_i^T$ tend $p.s.$ vers l'élément correspondant de la matrice $\mathbf{C}$. Par exemple, pour $\mathbf{C} = \mathbf{I}_m$, c'est à dire dans le cas

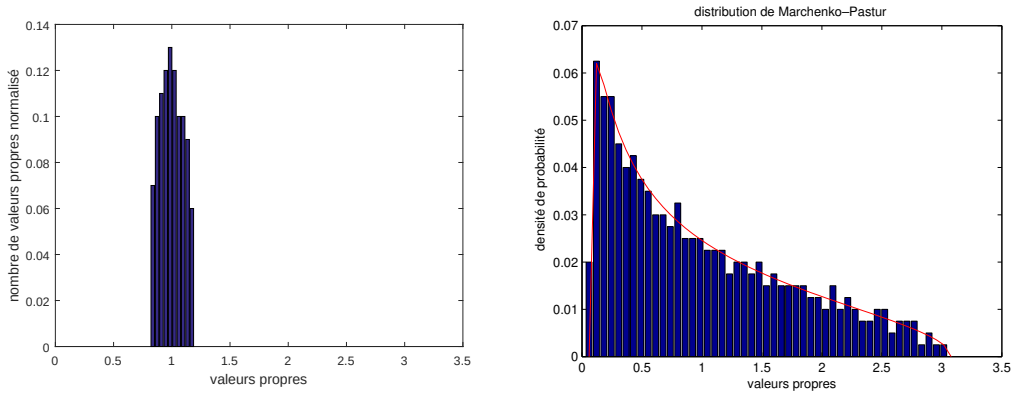

FIGURE 1.11 – Distribution des valeurs propres de la SCM d'un bruit blanc, à gauche $c_N = 0.01$ et à droite $c_N = 0.57$.

d'un bruit blanc gaussien, les valeurs propres de $\frac{1}{N}\sum_{i=1}^{N}\mathbf{y}_i\mathbf{y}_i^T$, correspondant à l'estimation empirique de la matrice de covariance, tendent vers 1, dans le cas où seul N tend vers l'infini et m est fixé. Elles seront notées $\hat{\lambda}_k$ pour $k \in [\![0, m-1]\!]$ par la suite. La figure de gauche de 1.11 illustre la loi des grands nombres, pour $N = 10000$ et $m = 100$. Cependant, sur la figure de droite de 1.11, où $N = 700$ et $m = 400$, les valeurs propres ne sont plus réparties autour de 1 d'une façon symétrique mais suivent la loi de Marchenko-Pastur. Ainsi, beaucoup de méthodes qui reposent sur cette loi des grands nombres ne peuvent s'appliquer au cas où N et m sont du même ordre de grandeur. Cependant, comme la distribution des valeurs propres reste connue (c'est la loi de Marchenko-Pastur), la théorie des matrices aléatoires développe de nouveaux outils théoriques pour appréhender correctement la distribution des valeurs propres du bruit pour ces signaux de grande dimension. Afin de définir la loi de Marchenko-Pastur, soit $\mathbf{A} \in \mathbb{C}^{m\times m}$ une matrice aléatoire de valeurs propres λ_k, $k \in [\![0, m-1]\!]$, alors $\mu_{\mathbf{A}}$, l'e.s.d. (Distribution Spectrale Empirique) de cette matrice est définie de la manière suivante : $\mu_{\mathbf{A}} = \frac{1}{m}\sum_{k=0}^{m-1}\delta_{\lambda_k}$.

La loi de Marchenko-Pastur est définie comme suit :

Théorème 1.6.1. *Loi de Marchenko-Pastur [30]*
Soit $\mathbf{Y} \in \mathbb{C}^{m\times N}$ composée d'éléments indépendants et identiquement distribués de moyennes nulles et de variances unités. Alors, sous l'hypothèse 3 (1.4.2), l'e.s.d. de la SCM $\frac{1}{N}\mathbf{Y}\mathbf{Y}^H$ converge faiblement vers une fonction F_c non aléatoire dont la densité de probabilité est donnée par :

$$f_c(x) = (1-c^{-1})^+\,\delta(x) + \frac{1}{2\,\pi\,c\,x}\sqrt{(x-a)^+\,(x-b)^+}\,\mathbb{1}_{]a,b[}(x)\,, \tag{1.37}$$

où $a = (1-\sqrt{c})^2$, $b = (1+\sqrt{c})^2$ et $\delta(.)$ est la distribution de Dirac en zéro. La fonction F_c est la loi de Marchenko-Pastur.

Un cas un peu plus général peut être montré, lorsque les éléments de la matrice $\mathbf{Y}$ ont une variance égale à σ^2. Dans ce cas, le support de la loi devient $]a,b[=]\sigma^2(1-\sqrt{c})^2, \sigma^2(1+\sqrt{c})^2[$. Cette loi montre que les estimateurs qui supposent que l'e.s.d. des valeurs propres est indépendante de $c_N = \frac{m}{N}$ ne peuvent donner de bons résultats puisque ce postulat est faux.

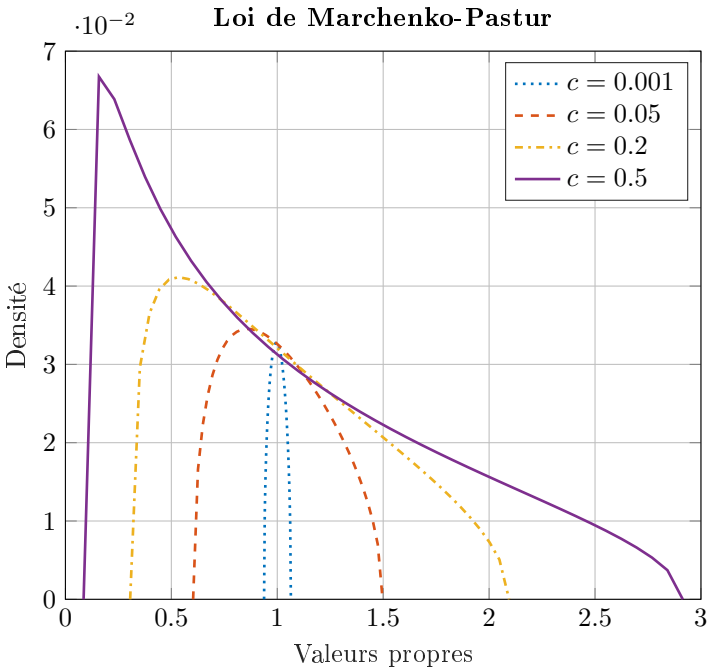

FIGURE 1.12 – Distribution des valeurs propres de $\frac{1}{N}\mathbf{Y}\mathbf{Y}^H$ pour différentes valeurs de c

Sur la figure 1.12 est représenté, pour différentes valeurs de c, la loi de Marchenko-Pastur, afin de bien illustrer ce phénomène. Plus la grandeur c est proche de 0, plus la distribution se rapproche de 1. Ainsi, en partant de matrices aléatoires, il est possible d'obtenir une distribution de probabilité à support déterministe borné et connu pour leurs valeurs propres. Ce support ne dépend que de la grandeur c et de la variance des données.

La figure 1.13 montre, pour une même valeur de c mais pour des valeurs de m et N différentes, les différents histogrammes obtenus pour les valeurs propres de la SCM pour un signal constitué d'un bruit blanc gaussien de moyenne nulle et de variance unité. Sur cette figure, en rouge, est aussi représenté la loi de Marchenko-Pastur. Il est bien visible que cette loi dépend du paramètre c et non directement des grandeurs m et N. Par ailleurs, pour des valeurs relativement faibles de N, (comme par exemple $N = 100$ sur la figure 1.13), il est possible d'être dans le cas où la loi des grands nombres ne s'applique déjà plus.

La preuve de cette loi peut se trouver dans [30] et peut se faire selon différentes approches. L'une d'elle fait appel à un outil primordial pour la suite qui est la transformée de Stieltjes, présentée ci-après.

Mais avant d'introduire cette notion, comme un certain nombre de méthodes reposent sur l'étude des plus grandes valeurs propres ou de la plus grande valeur propre de matrice de covariance et non sur la distribution de toutes les valeurs propres, la littérature propose de nombreuses études de la distribution de la plus grande valeur propre pour des matrices particulières. En effet, dans l'optique de distinguer les valeurs propres du bruit et celles dues aux *endmembers*, il semble judicieux de connaître la valeur de la plus grande valeur propre du bruit, et ce, pour poser par la suite un seuil qui déterminera quelles sont les valeurs propres qui ne peuvent pas être des valeurs propres dues au bruit. Par exemple, les fluctuations de la plus grande valeur propre d'une matrice de Wishart centrée ont été étudiées. Ces matrices, qui ont une définition assez simple, sont rencontrées dans le cas de signaux indépendants et de moyenne nulle, pour lesquels la SCM est étudiée. Leur étude donne le théorème suivant :

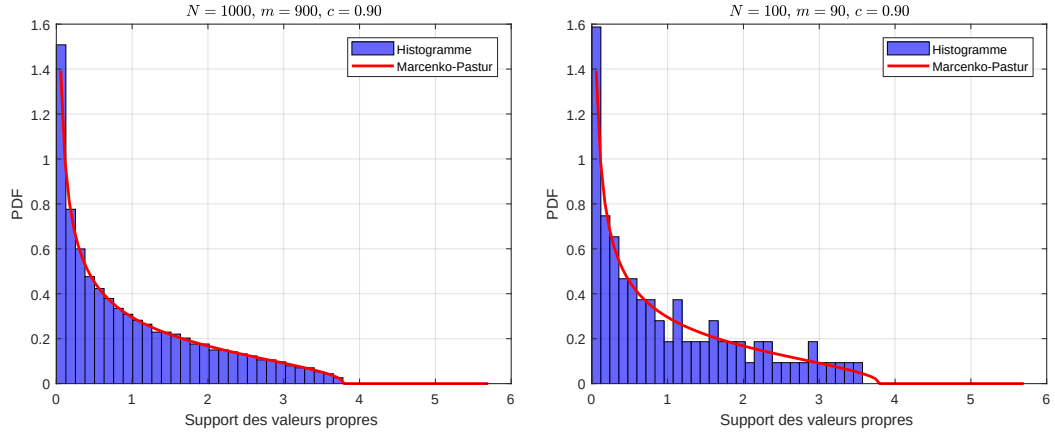

FIGURE 1.13 – Loi de Marchenko-Pastur et histogramme des valeurs propres pour un bruit blanc gaussien, de variance unité et de moyenne nulle, $c = 0,90$, avec, à droite $N = 100$ et $m = 90$, et, à gauche, $m = 900$ et $N = 1000$.

Théorème 1.6.2. *Loi de Tracy-Widom : Soit $\mathbf{Y} \in \mathbb{C}^{m \times N}$ composée d'éléments indépendants et identiquement distribués de moyennes nulles et de variances unités. Soit λ_0 la plus grande valeur propre de $\frac{1}{N}\mathbf{Y}\mathbf{Y}^H$, $c_N = \frac{m}{N}$, $b_N = (1+\sqrt{c_N})^2$ et $\sigma_N = (1+\sqrt{c_N})c_N^{4/3}$. Alors, pour tout x appartenant à un ensemble compact :*

$$\mathbb{P}\left(m^{2/3}\,\frac{\lambda_{0,N} - b_N}{\sigma_N} \geq x\right) \underset{N\to\infty}{\to} F_{TW}(x), \tag{1.38}$$

où $F_{TW}(.)$ représente la fonction de répartition de la loi de Tracy-Widom, définie comme suit :

$$F_{TW}(t) = \exp\left(-\int_t^{+\infty} (x-t)^2\, q(x)\, dx\right), \tag{1.39}$$

avec $q(.)$ la fonction de Painlevé II solution de l'équation différentielle ci-dessous :

$$\begin{aligned} q''(x) &= x\, q(x) + 2\, q^3(x), \\ q(x) &\underset{x\to\infty}{\sim} Ai(x), \\ \text{avec } Ai(x) &= \frac{1}{2\,\pi}\int \exp\left(i\,x\,t + \frac{i\,t^3}{3}\right) dt. \end{aligned}$$

Cette loi est tracée sur la figure 1.14 ainsi qu'un histogramme des plus grandes valeurs propres de $\frac{1}{N}\mathbf{Y}\mathbf{Y}^H$. Sur cette figure, il est visible que la plupart des plus grandes valeurs propres se répartissent autour d'une valeur moyenne, et que la distribution de cette loi n'est pas à queue lourde. Cette loi n'est cependant pas une loi centrée. Cette loi permet ainsi de prédire, pour une valeur de Probabilité de Fausse Alarme (PFA) donnée, quelle sera la valeur de la plus grande valeur propre du bruit d'un signal.

L'expression de le loi de Tracy-Widom n'étant pas explicite, des tableaux de valeurs donnant la fonction de répartition de cette loi sont utilisés. Cette loi caractérise donc la distribution de la plus grande valeur propre de la loi de Marchenko-Pastur. Un cas plus général a été étudié dans l'article [78], pour le cas des matrices $\mathbf{V} = \mathbf{C}^{1/2}\mathbf{Y}$, $\mathbf{C} \in \mathbb{C}^{m\times m}$ une matrice déterministe non négative. Alors dans ce cas le théorème s'applique mais pour

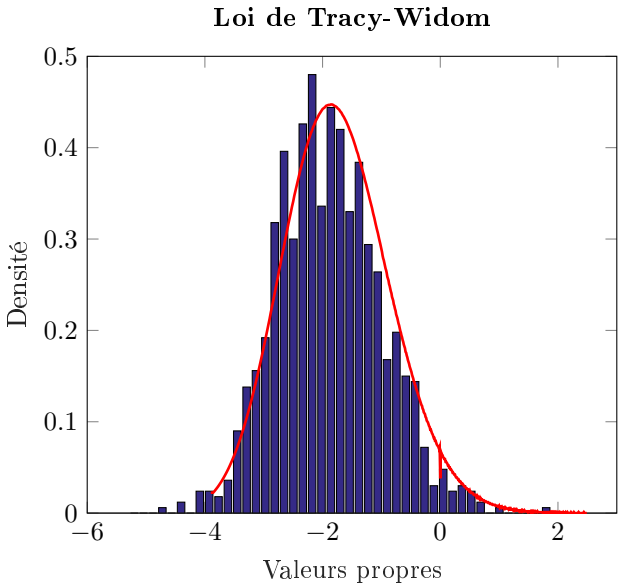

FIGURE 1.14 – Loi de Tracy-Widom et histogramme des valeurs propres, $c_N = 0.57$.

des valeurs différentes de b_N et σ_N qui vont dépendre de la distribution spectrale de $\mathbf{C}$. Ce théorème est ainsi utile lorsqu'il s'agit de donner, pour une PFA donnée, le nombre d'*endmembers* présents dans l'image. Ce point sera développé un peu plus tard.

En ce qui concerne la transformée de Stieltjes, elle est définie de la manière suivante, *Supp* désignant le support :

Definition 1.6.2. Transformée de Stieltjes
Soit μ une mesure finie réelle, $z \in \mathbb{C} \setminus Supp(\mu)$. Alors la transformée de Stieltjes de la mesure μ est définie comme suit :

$$m_\mu(z) = \int \frac{1}{t-z}\,\mu(dt). \tag{1.40}$$

Definition 1.6.3. Transformée inverse de Stieltjes
Soit μ une mesure finie réelle. Alors, pour tout a et pour tout b points du support de μ,

$$\mu([a\,b]) = \lim_{\epsilon\to 0} \frac{1}{\pi} \int_a^b Im\left[m_\mu(x + i\,\epsilon)\right]\,dx. \tag{1.41}$$

Si μ possède une densité de probabilité f en x, alors :

$$f(x) = \lim_{\epsilon\to 0} \frac{1}{\pi}\, Im\left[m_\mu(x + i\,\epsilon)\right]. \tag{1.42}$$

La transformée de Stieltjes est utilisée en pratique pour la caractérisation d'e.s.d. des valeurs propres de la matrice de covariance. En effet, la fonction de répartition définie comme :

$$\hat{Q}_m(\lambda) = \frac{1}{m}\sharp k, \{\hat{\lambda}_k \leq \lambda\},$$

n'est, en pratique, pas facile à étudier [75] (Rappel : les $\hat{\lambda}_k$ correspondent aux valeurs propres de $\frac{1}{N}\sum_{i=1}^{N} \mathbf{y}_i\mathbf{y}_i^H$ pour des vecteurs $\mathbf{y}_i$ complexes, $i \in [\![1, N]\!]$). Il est alors plus aisé

de considérer la transformée de Stieltjes de $\hat{Q}_m(\lambda)$, pour en déduire le comportement des valeurs propres dans des cas plus ou moins complexes. De plus, l'expression de la transformée de Stieltjes peut avoir une forme assez simple dans le cas d'une matrice $\frac{1}{N}\sum_{i=1}^{N}\mathbf{y}_i\,\mathbf{y}_i^H$ puisque $m_{\hat{Q}_m}(z) = \frac{1}{m}\sum_{i=0}^{m-1}\frac{1}{\hat{\lambda}_i - z} = \frac{1}{m}\,\mathrm{Tr}\left[\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{y}_i\,\mathbf{y}_i^H - z\,\mathbf{I}_m\right)^{-1}\right]$ pour tout $z \in \mathbb{C}^+$, $\mathbb{C}^+ = \{z \in \mathbb{C} | Im(z) > 0\}$, d'où son utilisation notamment dans la preuve de la loi de Marchenko-Pastur.

1.6.3 Distribution des valeurs propres pour un modèle plus général

Maintenant qu'il a été montré que les valeurs propres d'une matrice de covariance n'ont pas le même comportement en grande dimension, il apparaît plus clairement l'utilité d'étudier de nouveau la distribution de ces valeurs propres. En particulier, dans le cas où le signal n'est constitué que d'un bruit, l'étude des valeurs propres de la matrice de covariance permet de décider ensuite quelles valeurs propres appartiennent au bruit, et quelles valeurs propres, avec une certaine probabilité ou avec certitude, sont dues à la présence d'*endmembers*. Et ceci permet la mise en place d'outils adaptés pour le problème de l'imagerie hyperspectrale, puisque la plupart des méthodes reposent sur l'étude de ces valeurs propres.

Le théorème suivant donne toute son importance à la loi de Marchenko-Pastur puisqu'il assure qu'aucune valeur propre du bruit ne peut se retrouver en dehors de la distribution. Il a été introduit par Z. Bai et J.W. Silverstein [7], et s'occupe d'un cas un peu plus général.

Théorème 1.6.3. *Bai & Silverstein 1*

- *Soit* $\mathbf{X} \in \mathbb{C}^{m\times N}$ *une matrice de variables aléatoires indépendantes, identiquement distribuées, de moyennes nulles, de variances unités et de moment d'ordre 4 finis ;*
- *soit* $\mathbf{C}$ *de taille* $m \times m$ *une matrice hermitienne définie positive , de valeurs propres* $\lambda_i(\mathbf{C})$, $i \in [\![1, m]\!]$, *et telle que l'e.s.d.* $F^{\mathbf{C}} = \frac{1}{m}\sum_{i=0}^{m}\delta_{\lambda_i(\mathbf{C})}$ *de* $\mathbf{C}$ *converge en distribution vers* ν *;*
- *soit* $\mathbf{W} = \frac{1}{N}\mathbf{C}^{1/2}\mathbf{X}\mathbf{X}^H\,\mathbf{C}^{1/2} \in \mathbb{C}^{m\times m}$.

Alors, sous l'hypothèse 3 (1.4.2), si $\|\mathbf{C}\|$, *la norme spectrale, est bornée, pour tout intervalle* $[a, b]$, $a > 0$ *en dehors du support de* $F^{\mathbf{C}}$ *et de* ν, *la probabilité qu'aucune valeur propre de* $\mathbf{W}$ *n'apparaisse dans* $[a, b]$ *sous l'hypothèse 3 est égale à 1.*

Ainsi, grâce à ce théorème, la distribution des valeurs propres de la matrice de covariance du bruit a un support connu. Ce théorème ne donne cependant pas la distribution spectrale des valeurs propres, alors qu'il est possible de la déterminer d'une façon non explicite, en fonction de μ, par l'intermédiaire de la transformée de Stieltjes. C'est ce que montre le théorème suivant, toujours issu de l'article [7].

Théorème 1.6.4. *Bai & Silverstein 2*

Soit $\mathbf{C} \in \mathbb{C}^{m\times m}$ *telle que l'e.s.d. de* $\mathbf{C}$ *converge en distribution vers* ν. *Et soit* $\underline{\mathbf{W}} = \frac{1}{N}\mathbf{X}^H\mathbf{C}\mathbf{X} \in \mathbb{C}^{N\times N}$.

Alors $\underline{\mu}_N = \frac{1}{N}\sum_{i=}^{N}\delta_{\lambda_i(\underline{\mathbf{W}})}$ *converge presque sûrement et en distribution vers* $\underline{\mu}$ *dont la transformée de Stieltjes* $m_{\underline{\mu}}$ *est l'unique solution dans* $\mathbb{C}^+ = \{z \in \mathbb{C}\ Im(z) > 0\}$ *pour tout* $z \in \mathbb{C}^+$ *de l'équation :*

$$m_{\underline{\mu}}(z) = \left(-z + c\int \frac{1}{1 + t\,m_{\underline{\mu}}(z)}\,\nu(dt)\right)^{-1}. \tag{1.43}$$

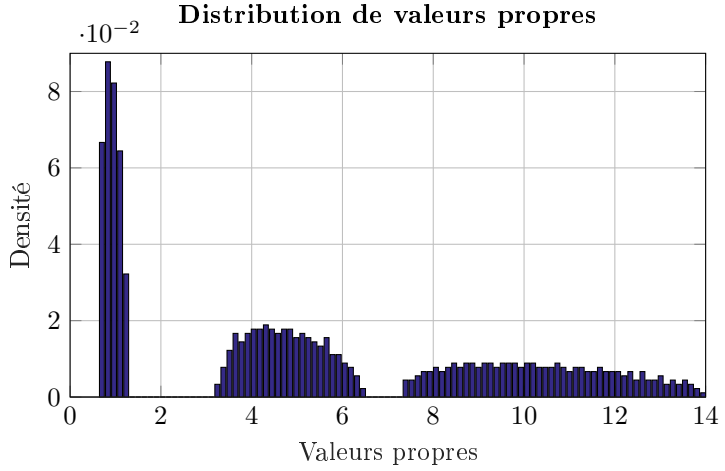

FIGURE 1.15 – Distribution des valeurs propres pour $c = 0.09$, $\nu = \frac{1}{3}\delta_1 + \frac{1}{3}\delta_5 + \frac{1}{3}\delta_{10}$.

L'équation :

$$\underline{\mu}(x) = (1-c)\,\mathbb{1}_{[0,\infty]}(x) + \mu(x)\,,$$

permet de relier $\underline{\mu}(x)$ et $\mu(x)$, l'e.s.d. limite pour $\mathbf{W}$ définie comme $\underline{\mu}$ l'est pour $\underline{\mathbf{W}}$. Il est ainsi possible d'étudier aussi bien la distribution des valeurs propres de $\mathbf{W}$ que de $\underline{\mathbf{W}}$, qui ne vont différer que par un certain nombre de valeurs propres nulles. Or, il se trouve que $\underline{\mu}$ est plus facile à étudier que μ. La relation suivante permet de lier leur deux transformées de Stieltjes, pour pouvoir passer facilement de l'une à l'autre :

$$m_{\underline{\mu}}(z) = c\, m_{\mu}(z) + (c-1)\frac{1}{z}\,. \tag{1.44}$$

La figure 1.15 montre une distribution possible des valeurs propres lorsque par exemple $\nu = \frac{1}{3}\delta_1 + \frac{1}{3}\delta_5 + \frac{1}{3}\delta_{10}$. Sur cette figure, plusieurs paquets de valeurs propres sont visibles, autant de paquets que ν contient de "zones" contenant des valeurs propres, à condition que la valeur choisie pour c le permette [75]. Ce choix de c impose une condition de séparabilité sur les paquets de valeurs propres. En effet, si le paramètre c est trop grand, certains paquets fusionnent. Ainsi sur la figure 1.15 sont visibles trois paquets de valeurs propres correspondant aux trois valeurs propres de la matrice $\mathbf{C}$ choisie. Alors, le théorème précédent permet de retrouver ces paquets de valeurs propres.

Le problème reste que ces expressions ne donnent pas le moyen de calculer explicitement la transformée de Stieltjes voulue.

Pour obtenir le support de distribution des valeurs propres de la SCM, il est nécessaire de considérer les manipulations mathématiques suivantes. Il est tout d'abord possible d'inverser la relation, pour trouver, en fonction de cette transformée de Stieltjes, les z correspondants :

$$z(m_{\underline{\mu}}) = -\frac{1}{m_{\underline{\mu}}} + c\int \frac{t}{1+t\,m_{\underline{\mu}}}\,d(\nu)\,.$$

Puis, les théorèmes suivant permettent de déterminer la distribution des valeurs propres de la matrice $\mathbf{W}$ ou $\underline{\mathbf{W}}$. Ils sont repris dans [78], [101] et [30] notamment.

Tout d'abord, d'après la formule de la transformée de Stieltjes inverse, pour repasser à la densité de probabilité de $\underline{\mathbf{W}}$, il suffit de calculer, pour tout point a, b de $\underline{\mu}$:

$$\underline{\mu}(b) - \underline{\mu}(a) = \lim_{\epsilon\to 0}\frac{1}{\pi}\int_a^b Im\left[m_{\underline{\mu}}(x+i\epsilon)\right]\,dx\,.$$

Il s'agit donc de calculer la valeur limite de $m_{\underline{\mu}}(z)$ lorsque $z \in \mathbb{C}^+ \to x \in \mathbb{R}$.

Théorème 1.6.5. *[30]*
Soient $\underline{\mathbf{W}} \in \mathbb{C}^{m\times m}$ *et* $\underline{\mu}$ *définis comme précédemment.*
Alors, $\forall x \in \mathbb{R}^+$, $\lim\limits_{\substack{z\to x \\ z\in\mathbb{C}^+}} m_{\underline{\mu}}(z)$ *existe et est notée* $m^\circ(x)$. m° *est de plus continue sur* $\mathbb{R}^\star$.

De plus, pour tout x du support de $\underline{\mu}$, la densité de probabilité $f(x) = \underline{\mu}(x) = \frac{1}{\pi} Im(m^\circ(x))$. Si f est une fonction analytique alors, $\forall x \in \mathbb{R}^\star$, $f(x) > 0$ [30]. Ainsi, le théorème suivant permet de déduire $m^\circ(x)$ à partir de m.

Théorème 1.6.6. *[30]*
Soit $Supp^c$ *le support complémentaire d'une distribution, et soient :*

- $G = \left\{ m_{\underline{\mu}} \ / \ m_{\underline{\mu}} \neq 0 \ , \ -\frac{1}{m_{\underline{\mu}}} \in Supp(\nu) \right\}$
- $x_{\underline{\mu}}$ *une fonction définie sur* G *par* $x_{\underline{\mu}} : m_{\underline{\mu}} \to -\frac{1}{m_{\underline{\mu}}} + c \int \frac{t}{1 + t\, m_{\underline{\mu}}} d(\nu)$.

Alors pour tout $x \in \mathbb{R}^\star$, *il est possible de calculer* $m^\circ(x)$ *de la manière suivante :*

- *Si* $x \in Supp(\mu)$, *alors* $m^\circ(x)$ *est l'unique solution sur* G *telle que* $Im(m^\circ(x)) > 0$ *de l'équation* $x = x_{\underline{\mu}}(m_{\underline{\mu}})$.
- *Si* $x \in Supp(\mu)^c$, *alors* $m^\circ(x)$ *est l'unique solution réelle sur* G *de l'équation* $x = x_{\underline{\mu}}(m_{\underline{\mu}})$ *telle que* $x'_{\underline{\mu}}(m^\circ) > 0$.

De plus, pour tout $m_{\underline{\mu}} \in G$, *si* $x'_{\underline{\mu}}(m_{\underline{\mu}}) > 0$ *alors* $x_{\underline{\mu}}(m_{\underline{\mu}}) \in Supp^c(\underline{\mu})$.

Ainsi, pour aller plus loin, le support de $\underline{\mu}$ correspond aux zones où $x'_{\underline{\mu}}(m_{\underline{\mu}}) > 0$. Les bords du support de $\underline{\mu}$ sont ainsi les solutions de :

$$\frac{1}{m^2} - c \int \frac{t^2}{(1 + t\, m)^2)} \nu(dt) = 0 \,,$$

soit les solutions de

$$\int \frac{t^2 m^2}{(1+tm)^2} \nu(dt) = \frac{1}{c} \,. \tag{1.45}$$

De plus, résoudre $x''_{\underline{\mu}}(m_{\underline{\mu}}) = 0$ donnera le nombre de paquets présents, comme résumé dans l'article de P. Vallet [113]. Et de cette façon, il est possible de retrouver une expression un peu plus facile à calculer du support de la loi de la distribution de valeurs propres de la matrice de covariance du bruit. Ceci peut être illustré par la figure suivante 1.16, où est représenté la fonction x en fonction de $m_{\underline{\mu}}$. Le support complémentaire correspond aux zones mises en valeur en rouge. Sur cette figure, $c = 0.009$ pour plus de visibilité. Les autres paramètres sont les mêmes que pour la figure 1.15. Les bornes du support sont ainsi définis par les ordonnées des points en rouge sur la figure 1.16. Des asymptotes sont également visibles à chaque valeur propre c'est à dire pour $m_{\underline{\mu}} = -1$, $\frac{-1}{5}$ et $\frac{-1}{10}$. Il est donc possible de retrouver le support de la distribution des valeurs propres de la SCM.

Le théorème suivant permet d'estimer correctement les valeurs propres de $\mathbf{C}$, dans le cas où les valeurs propres de la matrice $\mathbf{W}$ sont connues. Il sera repris et utilisé par la suite en tant que théorème à l'origine de l'algorithme G-MUSIC.

Théorème 1.6.7. *[75]*

- *Soient* $c_1 < \ldots < c_K$ *les* K, $K < m$, *différentes valeurs propres de la matrice* $\mathbf{C}$, *chacune de multiplicité* $m_1, \ldots, m_k$ *et soit* $\mathcal{M}_k = \left\{ \sum\limits_{j=1-1}^{k} m_j + 1, \ldots, \sum\limits_{j=1}^{k} m_j \right\}$, *pour* $k \in [\![1, \ldots, m]\!]$, *représentant en pratique les indices des valeurs propres correspondant à la même multiplicité.*

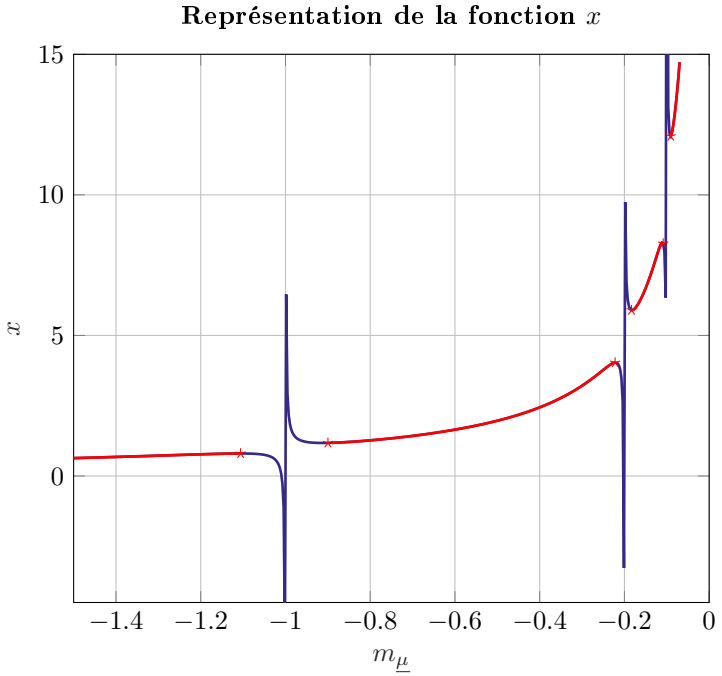

FIGURE 1.16 – Zones correspondants au support complémentaire pour la distribution des valeurs propres de la matrice $\mathbf{W}$.

- *Soient* $\lambda_1 \leq \ldots \leq \lambda_m$ *les valeurs propres de* $\mathbf{W}$ *et soit le vecteur* $\boldsymbol{\lambda} = (\lambda_1 \leq \ldots \leq \lambda_m)^T$.
- *Soient* $\mu_1 \leq \ldots \leq \mu_m$ *les valeurs propres de la matrice* $\left(\mathrm{diag}(\boldsymbol{\lambda}) - \frac{1}{N}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^T\right)$, $\mathrm{diag}(\boldsymbol{\lambda})$ *une matrice contenant sur sa diagonale* $\boldsymbol{\lambda}$ *et des zéro ailleurs.*
- *Soient aussi* $\hat{c}_k$, $k \in [\![1, \ldots, m]\!]$ *tels que*

$$\hat{c}_k = \frac{N}{m_k} \sum_{i \in \mathcal{M}} (\lambda_i - \mu_i)$$

Alors, sous l'hypothèse 3 (1.4.2), et si les paquets des valeurs propres de $\mathbf{C}$ *vérifient les conditions assurant leur séparabilité (voir plus haut et voir [30]),*

$$\forall k\,,\ \hat{c}_k - c_k \overset{p.s.}{\to} 0 \tag{1.46}$$

Ce théorème donne ainsi une estimation des valeurs propres de la matrice $\mathbf{C}$ lorsque celles de $\mathbf{W}$ sont connues. Dans un cas assez général, il est ainsi toujours possible d'estimer correctement la distribution des valeurs propres de la matrice d'intérêt, même si des lois valables dans le cas $m \ll N$ ne s'appliquent plus.

1.6.4 Distribution des valeurs propres pour le modèle *Spiked*

La distribution des valeurs propres de la matrice de covariance d'un signal constitué uniquement d'un bruit venant d'être étudiée, il est maintenant important d'étudier l'influence d'un signal qui s'ajouterait aux données sur cette distribution.

À présent, il sera supposé que la matrice $\mathbf{C}$ est égale à l'identité et donc que la densité de probabilité de la distribution des valeurs propres est la loi de Marchenko-Pastur. Lorsque les valeurs propres du bruit et les valeurs propres des *endmembers* sont clairement différentiables,

il devient alors possible d'estimer le nombre d'*endmembers* en comptant le nombre de valeurs propres différentes de celles du bruit, évidemment à condition que le RSB des *endmembers* soit suffisamment élevé pour leur permettre de sortir du support de la loi de Marchenko-Pastur. Mais ce point est développé dans le théorème 1.6.8 suivant. Le modèle *Spiked* en théorie des matrices aléatoires correspond à un modèle ou une matrice de rang faible par rapport aux dimensions m et N est ajoutée au bruit $\mathbf{W}$ de la partie précédente. Il fut introduit par [54]. Il existe différents types de modèles *spiked*, additifs ou multiplicatifs, et le modèle choisi ici pour les images hyperspectrales correspond parfaitement au modèle multiplicatif.

Le modèle choisi est de la forme : $\mathbf{Y} = \mathbf{MS} + \mathbf{B}$. Dans un premier temps, le modèle du bruit est un modèle plus simple que celui choisi pour modéliser les images hyperspectrales, c'est à dire $\mathbf{B} = \mathbf{X}$, $\mathbf{X}$ définie comme précédemment c'est à dire une matrice d'élément indépendants, gaussiens, (identiquement distribués) de moyenne nulle et de variance σ^2. Alors la matrice de covariance du signal peut s'écrire comme $\mathbf{\Gamma} = \frac{1}{N}\mathbf{MSS}^H\mathbf{M}^H + \sigma^2\mathbf{I}_m$. La partie $\frac{1}{N}\mathbf{MSS}^H\mathbf{M}^H$ peut être assimilée à une perturbation de rang faible. Le modèle pour la matrice $\mathbf{Y}$ est alors équivalent au modèle suivant :

$$\mathbf{Y} = \mathbf{\Gamma}^{1/2}\,\mathbf{Z}\,, \tag{1.47}$$

ou $\mathbf{Z}$ est une matrice de taille $m \times N$ d'éléments gaussiens, indépendants, de moyenne nulle et de variance unité. Alors le théorème suivant s'applique, pour un cas un peu plus général :

Théorème 1.6.8. *[9]*

Soit $\mathbf{Y} = (\mathbf{P} + \mathbf{I}_m)^{1/2}\,\mathbf{Z}$ *où* $\mathbf{Z} \in \mathbb{C}^{m\times N}$ *est une matrice d'entrées indépendantes, identiquement distribuées, de moyenne nulle, de variance unité et de moment d'ordre 4 fini,* $\mathbf{P} \in \mathbb{C}^{m\times m}$ *une matrice hermitienne non-négative de rang p fixe.*

Soient $\omega_0 \leq \ldots \leq \omega_{p-1}$ *et* $\hat{\lambda}_0 \leq \ldots \leq \hat{\lambda}_{m-1}$ *les valeurs propres de* $\mathbf{P}$ *et de* $\frac{1}{N}\mathbf{YY}^H$ *respectivement.*

Alors :

Si $\omega_k > \sqrt{c}$,

$$\hat{\lambda}_k \overset{a.s.}{\to} (1+\omega_k)\left(1+\frac{c}{\omega_k}\right).$$

Sinon,

$$\hat{\lambda}_k \overset{a.s.}{\to} (1+\sqrt{c})^2\,.$$

Le théorème 1.6.8 donne ainsi la position des valeurs propres de la matrice de covariance empirique engendrées par la partie des *endmembers*. La condition $\omega_k > \sqrt{c}$ peut être vue comme une condition de séparabilité des valeurs propres dues aux *endmembers* et celles dues au bruit. La figure 1.17 montre la distribution des valeurs propres de la SCM lorsque deux *endmembers* indépendants sont ajoutés au signal avec un RSB suffisamment fort. Deux valeurs propres ressortent alors à droite de la distribution des valeurs propres du bruit.

La figure 1.18 montre un autre cas, où 4 *endmembers* cette fois-ci sont présents sur l'image, et où leur position a été calculée théoriquement grâce au théorème 1.6.8. Leur positions théoriques sont en rouge. Cette nouvelle illustration montre à la fois la correspondance entre la théorie des matrices aléatoires et la distribution des valeurs propres de la SCM d'une image hyperspectrale constituée d'*endmembers* auxquels s'ajoute un bruit gaussien.

La distribution de la plus grande valeur propre peut également être connue en fonction de ω_{p-1} [8].

Si le bruit est un bruit gaussien corrélé, des résultats plus ou moins similaires existent, et le nombre de valeurs propres dépassant de la loi de Marchenko-Pastur est connu, en

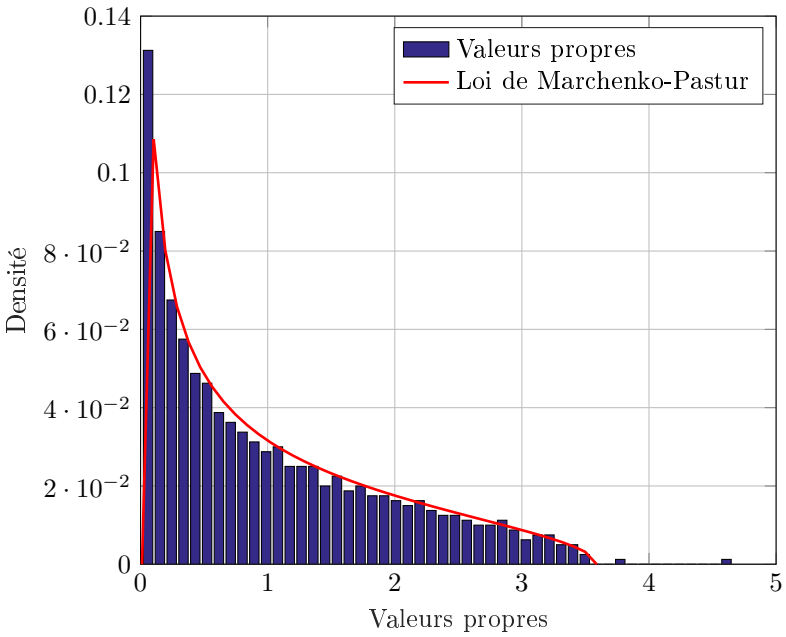

FIGURE 1.17 – Distribution des valeurs propres de l'estimation empirique de la matrice de covariance pour un bruit gaussien et deux sources indépendantes.

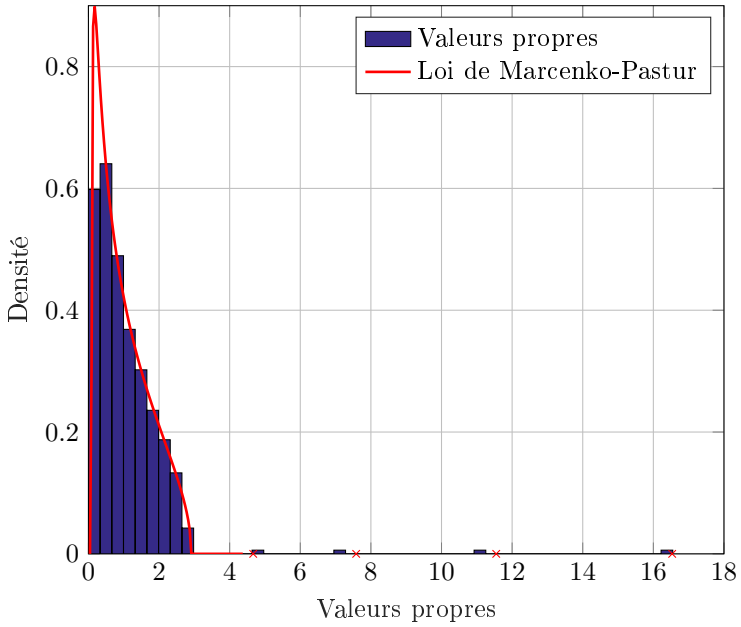

FIGURE 1.18 – Distribution des valeurs propres de l'estimation empirique de la matrice de covariance pour un bruit gaussien et quatre sources indépendantes.

fonction des valeurs propres de la matrice de corrélation du bruit. Ces résultats, résumés dans [23] mais qui ne seront pas exploités ici, montrent qu'il est possible, en fonction de la fonction de corrélation et de la transformée de Stieltjes de sa distribution spectrale, de connaître la distribution des valeurs propres de la matrice $\mathbf{Y}$. Ainsi, même dans le cadre des grandes dimensions, l'étude de la distribution des valeurs propres de matrices particulières reste possible et a déjà donné lieu à la publication de nombreux résultats, qui seront utiles pour la suite de cette thèse.

1.7 Quelques méthodes RMT pour l'imagerie hyperspectrale

Des méthodes utilisant des théorèmes et principes présentés dans la partie précédente ont déjà été développées pour l'imagerie hyperspectrale, pour différents modèles que celui choisi ici. Par exemple, les articles [46], [17], proposent d'estimer le nombre d'*endmembers* pour ensuite pouvoir estimer leur spectre ainsi que leurs abondances en utilisant les résultats sur les modèles *Spiked*.

Le bruit est assimilé à un bruit gaussien, de moyenne nulle et de matrice de covariance $\mathbf{\Sigma} = \sigma^2 \mathbf{I}_m$. La matrice de covariance empirique du signal $\mathbf{Y}$ est notée $\hat{\mathbf{\Lambda}}$ et a pour valeurs propres $\hat{\lambda}_1 \leq \ldots \leq \hat{\lambda}_m$. Les résultats sont donnés toujours sous l'hypothèse 3 (1.4.2). Les hypothèses propres aux images hyperspectrales (somme des abondances égales à un, ...) sont également prises en compte.

L'estimation du nombre d'*endmembers* $\hat{p}$ peut se faire alors de la manière suivante ([46], [19] par exemple) : d'après les résultats déjà présentés sur la loi de Tracy-Widom, la valeur limite ω pour une PFA donnée, notée α, à partir de laquelle une valeur propre est une valeur propre issue des *endmembers*, est

$$\omega = \sigma^2 \left(1+\sqrt{c}\right)^2 + \frac{\sigma^2 \left(1+\sqrt{c}\right) c^{4/3}}{m^{2/3}} F_{TW}^{-1}(1-\alpha)\,. \tag{1.48}$$

Ainsi, pour estimer le nombre d'*endmembers*, il suffit de compter combien de valeurs propres sont au delà de ce seuil, comme présenté dans [19] :

$$\hat{p}_{Caw} = \min_{k=1,\ldots,m} \left(\hat{\lambda}_k < \sigma^2 \left(\frac{\beta_c}{N^{2/3}} s(\alpha) + \left(1+\sqrt{c}\right)^2 \right) \right) - 1\,, \tag{1.49}$$

avec $s(\alpha)$ une fonction de la fonction de distribution inverse de la loi de Tracy-Widom, et $\beta_c = \left(1+\sqrt{c}\right)\left(1+\sqrt{c^{-1}}\right)^{1/3}$. Cependant, cet estimateur n'est pas consistant, et d'autres auteurs se sont appliqués à trouver un meilleur estimateur, comme présenté ci-après.

En notant $\hat{\gamma}_1 \leq \ldots \leq \hat{\gamma}_r$ les valeurs propres de $\hat{\mathbf{\Lambda}}$ dues aux *endmembers* et en supposant que $\forall i\,,\ \hat{\gamma}_i > 1+\sqrt{c}$, en notant $\phi : x \to (x+1)(1+\frac{c}{x})$ pour tout réel x, alors d'après la partie précédente :

$$\hat{\lambda}_j \underset{N\to\infty}{\to} \sigma^2\, \phi(\hat{\gamma}_j) \text{ a.s.}\,, \text{ pour tout } 1 \leq j \leq r\,, \tag{1.50}$$

$$\hat{\lambda}_j \underset{N\to\infty}{\to} \sigma^2 \left(1+\sqrt{c}\right)^2 \text{ a.s.}\,, \text{ pour tout } j > r\,, \tag{1.51}$$

ceci traduisant le fait que les valeurs propres des *endmembers* "sortent" de la distribution de Marchenko-Pastur. L'estimateur alors proposé dans [46] est de considérer $\delta = \hat{\lambda}_j - \hat{\lambda}_{j+1}$ pour $j \in [\![1, m-1]\!]$ pour ensuite calculer :

$$\hat{p}_{Hal} = \min\{ j \in [\![1, m-1]\!] \,/\, \delta_{j+1} < d_N \}\,, \tag{1.52}$$

où d_N est un seuil à déterminer, sachant que l'estimateur $\hat{p}_{Hal}$ est consistant si $d_N \underset{N\to\infty}{\to} 0$ et $N^{2/3} d_N \underset{N\to\infty}{\to} \infty$ (voir l'article [90] pour le détail). Les auteurs de [90] proposent par exemple $d_N = \frac{4\sqrt{2log(log(N))}}{N^{2/3}}\beta_c$.

Une autre façon d'estimer le nombre d'*endmembers*, mais qui n'a pas été spécifiquement appliqué au cas de l'imagerie hyperspectrale, est d'utiliser l'espacement entre les valeurs propres comme proposé dans [118], appliqué ici à un cas plus simple que celui de l'article :

$$\hat{p}_{Vin} = \underset{j\in[\![1,m]\!]}{\text{argmax}} \frac{\hat{\lambda}_{j-1}}{\hat{\lambda}_j} > \epsilon\,, \tag{1.53}$$

où ϵ est un seuil à fixer.

Le problème principal de ces méthodes est l'estimation de la matrice de covariance du signal, qui, si non consistante, peut mener à une surestimation du nombre d'*endmembers*. En effet, chacune de ces méthodes nécéssite l'utilisation d'un seuil, qui est obtenu soit en prenant en compte les valeurs propres de cette matrice, soit fixé arbitrairement (ce qui peut aussi poser problème). Des solutions sont ainsi proposées, comme dans l'article [17], qui consiste à ne plus considérer seulement les valeurs propres de la matrice de covariance estimée mais aussi les estimations des valeurs propres de la matrice de covariance du bruit. Ceci présuppose d'avoir préalablement estimé le sous espace bruit.

1.8 Introduction à l'estimation robuste

Même si la théorie des matrices aléatoires propose des outils qui semblent adaptés à l'imagerie hyperspectrale, il est aussi important de développer des méthodes d'estimation du nombre d'*endmembers* qui soient robustes par exemple à la forme du bruit et notamment à la présence de valeurs aberrantes. Comme les méthodes proposées ici s'appuient sur l'estimation des valeurs propres de la matrice de covariance, cette section présente quelques estimateurs robustes qui seront repris par la suite.

La famille d'estimateurs considérés est la famille des M-estimateurs. Ils sont issus de la généralisation de l'Estimateur du Maximum de Vraisemblance (EMV), qui se présente comme suit : soient $[\mathbf{y}_1, \ldots, \mathbf{y}_N]$ des signaux indépendants suivant une distribution CES $\mathcal{GCE}(\boldsymbol{\mu}, \boldsymbol{\Gamma}, g)$ telle que pour tout $i \in [\![1, N]\!]$, $f(\mathbf{y}) = k\,|\boldsymbol{\Sigma}|^{-1}\, g\left(\left(\mathbf{y} - \boldsymbol{\mu}^H\right)\, \boldsymbol{\Sigma}^{-1}\, \left(\mathbf{y} - \boldsymbol{\mu}^H\right)\right)$, avec k un paramètre d'échelle. g est supposée réelle et dérivable. Alors la log-vraisemblance s'écrit :

$$L\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right) = k - N\, \log\left(|\boldsymbol{\Sigma}|\right) + \sum_{i=1}^{N} \log\left[g\left(\left(\mathbf{y}_i - \boldsymbol{\mu}\right)^H\, \boldsymbol{\Sigma}^{-1}\, \left(\mathbf{y}_i - \boldsymbol{\mu}\right)\right)\right]\,. \tag{1.54}$$

Par dérivation en fonction de $\boldsymbol{\mu}$ et de $\boldsymbol{\Sigma}$, la solution qui correspond au EMV est alors donnée par les deux équations suivantes en $\hat{\boldsymbol{\mu}}$ et $\hat{\boldsymbol{\Sigma}}$:

$$\begin{aligned}
\hat{\boldsymbol{\mu}} &= \frac{\sum_{i=1}^{N} \frac{-g'(u_i)}{g(u_i)} \mathbf{y}_i}{\sum_{i=1}^{N} \frac{-g'(u_i)}{g(u_i)}} \\
\hat{\boldsymbol{\Sigma}} &= \frac{1}{N} \sum_{i=1}^{N} \frac{-g'(u_i)}{g(u_i)} (\mathbf{y}_i - \hat{\boldsymbol{\mu}})(\mathbf{z}_i - \hat{\boldsymbol{\mu}})^H\,,
\end{aligned} \tag{1.55}$$

où $u_i = (\mathbf{y}_i - \hat{\boldsymbol{\mu}})^H \, \hat{\boldsymbol{\Sigma}}^{-1} \, (\mathbf{y}_i - \hat{\boldsymbol{\mu}})$ pour tout $i \in [\![1, N]\!]$. En pratique, il s'agit de résoudre itérativement ces deux équations. Voici deux exemples, où, la fonction g étant connue, il est possible de retrouver l'EMV correspondant :

— pour la distribution gaussienne :

$$\forall i \in [\![1, N]\!] \ , \ g(u_i) = \exp(-u_i) \text{ et } \frac{-g'(u_i)}{g(u_i)} = 1 \, , \tag{1.56}$$

ce qui donne comme solution pour $\hat{\boldsymbol{\Sigma}}$ la SCM.

— pour la K-distribution,

$$\forall i \in [\![1, N]\!] \ , \ \frac{-g'(u_i)}{g(u_i)} = \sqrt{\frac{\mu}{u_i}} \, \frac{K_{\mu-m-1}\left(2\sqrt{\mu\, t}\right)}{K_{\mu-m}\left(2\sqrt{\mu\, t}\right)} \, . \tag{1.57}$$

Cependant, la fonction g n'est pas toujours connue. Une solution consiste à remplacer les équations (1.55) par :

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \frac{\displaystyle\sum_{i=1}^{N} v_1(u_i)\mathbf{y}_i}{\displaystyle\sum_{i=1}^{N} v_1(u_i)} \\ \hat{\boldsymbol{\Sigma}} &= \frac{1}{N}\sum_{i=1}^{N} v_2\left(u_i^2\right) \, (\mathbf{y}_i - \hat{\boldsymbol{\mu}}) \, (\mathbf{z}_i - \hat{\boldsymbol{\mu}})^H \ \ , \end{aligned} \tag{1.58}$$

avec, cette fois ci : u_i la distance de Mahalanobis c'est à dire $u_i = \left((\mathbf{y}_i - \hat{\boldsymbol{\mu}})^H \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{y}_i - \hat{\boldsymbol{\mu}})\right)^{1/2}$ pour tout $i \in [\![1, N]\!]$, et v_1, v_2 sont deux fonctions poids, qui ne dépendent pas de la distribution CES étudiée et qui vérifient certaines hypothèses citées ci-après. L'existence et l'unicité ont été montré par R.A. Maronna dans [74] sous réserve que les hypothèses suivantes soient vérifiées :

Hypothèses de Maronna

- v_1 et v_2 sont deux fonction positives ou nulles, décroissantes et continues sur $[0, \infty[$.
- Les fonctions définies par $\Phi_1 : x \mapsto x\, v_1(x)$ et $\Phi_2 : x \mapsto x\, v_2(x)$ sont bornées.
- La fonction Φ_2 est croissante et elle l'est strictement sur les intervalles où $\Phi_2(x) < sup_x \geq 0 \Phi_2(x)$.
- Soit P_y la distribution empirique des signaux $[\mathbf{y}_1, \ldots, \mathbf{y}_N]$. Il existe un x_0 tel que, pour tout hyperplan H, $dim(H) < 2m - 1$ et $P_y(H) \leq 1 - \dfrac{m}{\sup_x \geq 0 \Phi_2(x)} - x_0$.

La dernière hypothèse peut ne pas être prise en compte comme montré dans [112] et [56].

Pour différents types de distribution du bruit CES, il est possible de trouver l'EMV correspondant en fonction de ces fonctions v_1 et v_2, à condition de connaître cette distribution. Par exemple :

— dans le cas d'une t-distribution présentée précédemment, les fonctions de poids correspondant à l'EMV sont de la forme :

$$v_2(u^2) = \frac{\mu + 2m}{\mu + 2u^2} \ \text{ et } \ v_1(u) = v_2(u^2) \ . \tag{1.59}$$

Même si choisir ces fonctions arbitrairement pour un bruit dont la distribution est inconnue est possible et donne bien un M-estimateur, certaines de ces fonctions sont plus adaptées et plus robustes pour des distributions de bruit inconnues dans le sens où les grandes valeurs sont mieux atténuées. Voici deux exemples souvent utilisés, dont un sera repris par la suite :

— estimateur de Huber [50] et [84] pour le cas complexe :

$$v_2\left(u^2\right)\begin{cases} 1/\beta \text{ si } u^2 < k^2\,, \\ k^2/\left(u^2\,\beta\right) \text{ sinon.} \end{cases} \tag{1.60}$$

k^2 et β deux paramètres ajustables, qui permettent de choisir la proportion de données à atténuer. k est défini grâce à la distribution du χ^2 à $2m$ degrés de libertés, avec F_{2m} la fonction de répartition d'une loi du χ^2, $q = F_{2\,m}(2\,k^2)$. Quant au paramètre β, il permet de rendre l'estimation de la matrice de covariance non biaisée dans le cas gaussien, et est également déterminée à l'aide de la fonction de répartition d'une loi du χ^2 : $\beta = F_{2m+2}(2k^2) + k^2\dfrac{1-q}{m}$. Par exemple, $q = 1$ renvoie à l'estimateur de la SCM.

— estimateur du point fixe, aussi appelé estimateur de Tyler :

$$v_1(u) = \frac{1}{u} \text{ et } v_2(u^2) = \frac{m}{u^2}\,. \tag{1.61}$$

Ce dernier estimateur a été largement étudié [112], [111], [86] et [88]. En notant $\hat{\mathbf{\Sigma}}_{FP}$ l'estimateur du point fixe pour la matrice de covariance d'un signal CES, c'est à dire l'unique solution de :

$$\hat{\mathbf{\Sigma}}_{FP} = \frac{m}{N}\sum_{i=1}^{N}\frac{\mathbf{y}_i\mathbf{y}_i^H}{\mathbf{y}_i^H\,\hat{\mathbf{\Sigma}}_{FP}^{-1}\,\mathbf{y}_i} \tag{1.62}$$

Alors les résultats suivant sont vrais, pour $\boldsymbol{\mu}$ connue :

- $\hat{\mathbf{\Sigma}}_{FP}$ est un estimateur consistant de $\mathbf{\Sigma}$ [87],
- Pour N grand, $\hat{\mathbf{\Sigma}}_{FP}$ est une matrice de Wishart à $\frac{mN}{m+1}$ degrés de liberté.

La figure 1.19 suivante montre les différentes fonctions de pondération v_2 pour trois estimateurs : la SCM, plusieurs estimateurs de Huber et le point fixe. Sur cette figure, plus le coefficient q se rapproche de 1, plus l'estimateur considéré est proche de la SCM. A l'inverse, une faible valeur de q rapproche l'estimateur du point fixe. Selon l'application, il peut être avantageux de considérer certaines valeurs pour q.

Dans cette thèse, c'est sur l'estimation de la matrice de covariance que sera porté un intérêt particulier. Les estimateurs pris en considération seront donc ceux de la forme simplifiée suivante, avec les nouvelles notations qui suivent :

$$\hat{\mathbf{\Sigma}} = \frac{1}{N}\sum_{i=1}^{N} u\left(\left(\mathbf{y}_i - \hat{\boldsymbol{\mu}}\right)^H\,\hat{\mathbf{\Sigma}}^{-1}\,\left(\mathbf{y}_i - \hat{\boldsymbol{\mu}}\right)\right)\,\left(\mathbf{y}_i - \hat{\boldsymbol{\mu}}\right)\,\left(\mathbf{y}_i - \hat{\boldsymbol{\mu}}\right)^H\,, \tag{1.63}$$

ou, pour des données de moyenne nulle :

$$\hat{\mathbf{\Sigma}} = \frac{1}{N}\sum_{i=1}^{N} u\left(\mathbf{y}_i^H\hat{\mathbf{\Sigma}}^{-1}\mathbf{y}_i\right)\mathbf{y}_i\,\mathbf{y}_i^H\,, \tag{1.64}$$

la fonction u vérifiant les hypothèses de Maronna 1.8. L'estimateur du point fixe a l'avantage de n'être sensible à aucun facteur d'échelle sur les $\mathbf{y}_i$. En effet, pour toute constante a et pour des signaux $[\mathbf{y}_1, \ldots, \mathbf{y}_N]$ de moyenne nulle,

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\left(a\,\mathbf{y}_i\right)\,\left(a\,\mathbf{y}_i\right)^H}{\left(a\,\mathbf{y}_i\right)^H\,\hat{\mathbf{\Sigma}}^{-1}\,\left(a\,\mathbf{y}_i\right)} = \frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{y}_i\,\mathbf{y}_i^H}{\mathbf{y}_i^H\,\hat{\mathbf{\Sigma}}^{-1}\,\mathbf{y}_i}\,. \tag{1.65}$$

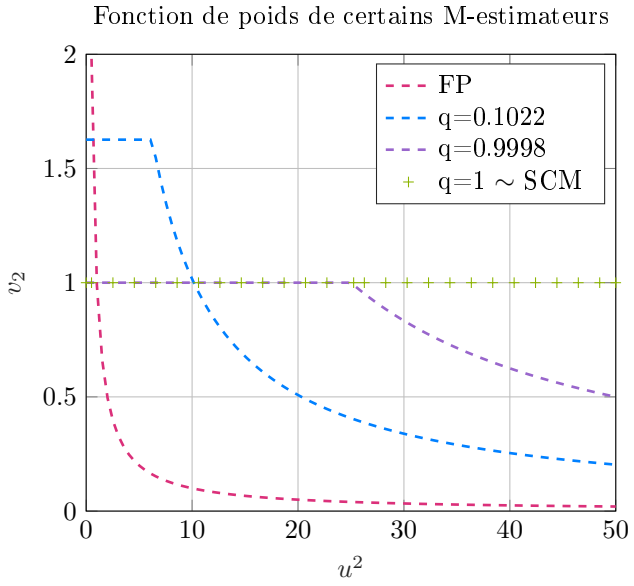

FIGURE 1.19 – Fonctions de poids de certains M-estimateurs pour $m = 10$ et différentes valeurs de q pour l'estimateur de Huber.

Pour finir, la distribution asymptotique des ces M-estimateurs a été étudiée, et la consistance démontrée [87].

Par la suite, la fonction Φ_2 sera notée Φ, et la limite $lim_{N\to\infty}\Phi$ sera notée Φ_∞.

Enfin, dans le cas ou le nombre de données serait insuffisant, c'est à dire si la taille m des spectres recueillis est trop important par rapport au nombre de pixels N, il est toujours possible d'utiliser un estimateur de type "régularisation" (ou "*shrinkage*") qui se présente sous la forme suivante :

$$\hat{\mathbf{\Sigma}}_{shr}(\beta) = (1-\beta)\,\hat{\mathbf{\Sigma}} + \beta\,\mathbf{I}_m\,, \tag{1.66}$$

où cette fois-ci, $\beta \in [0,1]$ est un paramètre à fixer en fonction des résultats souhaités. Un tel estimateur, par exemple dans le cas du point fixe, n'existe pas forcément puisque l'équation (1.65) n'admet pas obligatoirement de solution selon le β choisi. De récents travaux [88] ont permis de montrer que cette equation par exemple admet une solution si $\beta \in [\alpha, 1]$ avec $\alpha = \max\left(0, 1 - \frac{N}{m}\right)$. Cet estimateur de shrinkage ne sera pas développé par la suite mais peut être utilisé sur des images hyperspectrales réelles où la zone spatiale prise en compte est trop petite par rapport au nombre de bandes spectrales.

1.9 Synthèse

Ce chapitre de présentation de l'imagerie hyperspectrale, des problèmes et des solutions envisageables a permis d'introduire le contexte de cette thèse.

Cette thèse se place dans le contexte du traitement des données de grandes dimensions. Il s'agit également de considérer des images hyperspectrales ayant un bruit additif modélisé comme CES, corrélé, afin de prendre en compte un cas plus général que le cas du bruit gaussien. Les hypothèses qui seront considérées par la suite ont été énoncées dans ce premier chapitre d'introduction : le modèle général qui permettrait de représenter un grand nombre d'images hyperspectrales a été introduit. Ce modèle sera repris par la suite. Le problème qui

se pose alors est celui de la forme du bruit et de sa corrélation, ainsi que le problème de la taille des données. En effet, certaines images se rapprochent du cas applicatif de la RMT, c'est à dire un grand nombre de pixels et un grand nombre de bandes spectrales. Sous ces deux conditions, certaines méthodes ne s'appliquent plus tandis que d'autres renvoient des résultats moins bons que ceux sur des images qui ne vérifient pas ces hypothèses ou qui ont été réduites par des opérations de projection.

Dans ce premier chapitre d'introduction, des notions sur l'imagerie hyperspectrale, sur les problèmes associés et considérés dans cette thèse et sur les bruits CES ont été rappelées. Les limites de quelques méthodes souvent utilisées ont été démontrées pour le type de modèle choisi. Les fondements et quelques théorèmes de la RMT ont été exposés, ainsi que quelques éléments de la robustesse, utiles pour traiter les images hyperspectrales, comme cela sera démontré par la suite.

L'objet de cette thèse est de retrouver les informations qui se trouvent sur une image hyperspectrale. La difficulté est qu'elle n'utilise aucun a priori sur les images hyperspectrales utilisées : le bruit est supposé inconnu, sa matrice de covariance est aussi supposée inconnue, les *endmembers* sont inconnus, leur nombre est inconnu, leur distribution sur l'image est inconnue. La seule hypothèse faite sur les images traitées est celle de la forme du bruit. Celui-ci est supposé CES corrélé, c'est à dire un cas assez général. La seule méthode qui semble envisageable alors pour retrouver les *endmembers* de l'image hyperspectrale est d'abord estimer le nombre d'*endmembers* pour en déduire la dimension des matrices de mélange et d'abondances pour réussir ensuite à les séparer.

Chapitre 2

Sélection d'ordre pour le démélange d'images hyperspectrales

Ce chapitre propose différentes méthodes d'estimation du nombre d'endmembers, pour différents cas de figure. Chaque cas de figure correspond à un modèle de bruit différent, ce chapitre considérant d'abord un modèle de bruit simple pour aboutir au modèle plus complexe présenté en introduction. Ainsi, en revenant sur le modèle choisi pour le bruit, ce chapitre présente des méthodes variées dont la complexification permet d'aboutir aux dernières méthodes proposées dans ce chapitre. Dans ce chapitre sont aussi données en illustration quelques simulations sur des images hyperspectrales simulées.

2.1 Rappel des principales hypothèses

Hypothèse 1

- $\mathbf{T} \in \mathbb{R}^{N \times N}$ est une matrice diagonale, contenant sur sa diagonale, pour tout $i \in [\![1, N]\!]$, les τ_i correspondant aux valeurs de τ de la distribution gaussienne composée pour chaque pixel.
- $\mathbf{X} \in \mathbb{R}^{m \times N}$ est un bruit d'entrées indépendantes et identiquement distribuées, de moyennes zéro et de variances unité.

Hypothèse 2

- $\mathbf{C}$ est une matrice de Toeplitz de taille $m \times m$.

Hypothèse 3 :

- m et N sont supposés grands, avec $c_N = \dfrac{m}{N} \to c > 0$ lorsque $N \to \infty$ et $m \to \infty$.

Les deux premières parties de ce chapitre présentent des méthodes déjà existantes mais adaptées ici au cas de l'imagerie hyperspectrale. Ces méthodes s'appliquent sur un modèle qui ne vérifie pas forcément toutes les hypothèses du modèle général choisi pour cette thèse. Il est cependant intéressant de les considérer puisqu'elles s'appliquent à un cas particulier de ce modèle général. Elles permettent également de présenter d'abord des méthodes qui utilisent un modèle plus simple pour arriver à des méthodes utilisant le modèle plus complexe énoncé. Ces premières méthodes proposées, qui ont pour certaines déjà été appliquées à d'autres domaines, permettent de démontrer l'adaptabilité de la théorie des matrices aléatoires au cas

de l'imagerie hyperspectrale. Les autres parties correspondent à la présentation de méthodes développées entièrement durant cette thèse, qui ne s'appuient pas sur des méthodes déjà existantes dans d'autres domaines.

2.2 Estimation du nombre d'*endmembers* pour un bruit non corrélé, d'entrées indépendantes et identiquement distribuées.

Comme montré précédemment, certaines méthodes déjà développées dans le cadre des matrices aléatoires peuvent s'adapter au cas de l'imagerie hyperspectrale. Pour un modèle de bruit blanc, c'est à dire pour le modèle suivant :

$$\mathbf{Y} = \mathbf{M}\,\mathbf{S} + \sigma\,\mathbf{X}\,, \tag{2.1}$$

il est toujours possible de regarder le nombre de valeurs propres de l'estimation empirique de la matrice de covariance de $\mathbf{Y}$, c'est à dire $\frac{1}{N}\mathbf{Y}\mathbf{Y}'$, qui dépassent un seuil fixé sur la loi de Tracy-Widom. Ce nombre de valeurs propres correspond ainsi au nombre d'*endmembers* estimé $\hat{p}$ pour une PFA donnée. La méthode proposée consiste ainsi à considérer successivement les deux hypothèses suivantes :

$$\begin{cases} H_0 : \text{ au plus } k \textit{ endmembers} \\ H_1 : \text{au moins } k+1 \textit{ endmembers} \end{cases}$$

avec $k \in [\![0, \min(m, N) - 1]\!]$. Les valeurs propres de la SCM (l'estimation empirique de la matrice de covariance) de $\mathbf{Y}$ sont classées dans l'ordre décroissant : $\hat{\lambda}_0 \geqslant ... \geqslant \hat{\lambda}_{m-1}$. À chaque itération en k, la k-ième valeur propre est comparée à un seuil, permettant de rejeter ou de valider l'hypothèse en question, comme suit :

$$\hat{\lambda}_k \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \zeta_N\,, \tag{2.2}$$

avec, pour une PFA fixée et notée α :

$$\zeta_N = \hat{\sigma}^2(k)\left(b_N + \frac{\sigma_N}{m^{2/3}}\left(F_{TW}^{-1}(1-\alpha)\right)\right), \tag{2.3}$$

où $\hat{\sigma}^2(k)$ est défini ci après. Les autres paramètres sont les mêmes que ceux définis dans le chapitre précédent. Le test d'hypothèses prend fin pour le plus petit k tel que $\hat{\lambda}_k < \zeta_N$. L'estimation du nombre d'*endmembers* est alors le même que celui proposé par J. Vinogradova dans sa thèse [116] :

$$\hat{p} = \arg\min_k \left(\hat{\lambda}_k < \zeta_N\right) - 1\,. \tag{2.4}$$

Le paramètre $\hat{\sigma}^2(k)$ est défini dans [62] et correspond à un estimateur non biaisé et consistant de σ^2, analysant les $k+1$ dernières valeurs propres de la SCM, qui sont donc, si l'hypothèse H_0 est vérifiée, égales à σ^2. L'estimation de σ^2 est obtenue en résolvant itérativement le système d'équations non linéaires suivant, en prenant pour $\hat{\rho}_j$ la plus grande racine de l'équation quadratique ci-après :

$$\hat{\sigma}^2(k) - \frac{1}{m-k-1}\left(\sum_{j=k+1}^{m-1}\hat{\lambda}_j + \sum_{j=0}^{k}(\hat{\lambda}_j - \hat{\rho}_j)\right) = 0\,,$$

$$\hat{\rho}_j^2 - \hat{\rho}_j\left(\hat{\lambda}_j + \hat{\sigma}^2(k) - \hat{\sigma}^2(k)\frac{m-k-1}{N}\right) + \hat{\lambda}_j\hat{\sigma}^2(k) = 0\,.$$

TABLE 2.1 – Résultat de la méthode pour différents RSB, avec $\alpha = 0.01$, $m = 200$, $N = 400$.

RSB	32	28	27	26	24	18
$\hat{p}$	4.00	4.00	3.34	2.12	0.47	0
$\hat{p}_{AIC}$	4.0	3.64	1.8	0.27	0	0
$v_{\hat{p}}$	0	0	0.27	0.25	0.27	0
v_{AIC}	0	0.23	0.40	0.20	0	0

TABLE 2.2 – Résultats de la méthode pour différentes valeurs de α, RSB = 27 dB, $m = 200$, $N = 400$.

PFA	1	0.005	0.001
$\hat{p}$	4.56	3.20	3.08
$\hat{p}_{AIC}$	1.7	1.7	1.7
$v_{\hat{p}}$	0.25	0.30	0.25

En prenant pour valeur initiale, une valeur de $\hat{\sigma}^2(k)$ qui correspond à un estimateur biaisé de σ^2 :

$$\hat{\sigma}^2(0) = \frac{1}{m-k-1} \sum_{j=k+1}^{m-1} \hat{\lambda}_j \, .$$

Ces équations donnent une solution pour $\hat{\sigma}$ sous les conditions données dans [62].

Les tableaux 2.1 et 2.2 montrent les performances de cet estimateur. Pour ces simulations, des spectres hyperspectraux ont été simulés de la manière suivante : chaque pixel $\mathbf{y}_i$, $i \in [\![1, N]\!]$ s'écrit $\mathbf{y}_i = \sum_{j=1}^{p} s_{i,j}\mathbf{m}_j + \sigma\,\mathbf{x}_i$ avec $\mathbf{m}_j = \exp\left(2i\pi j \left(0, \ldots, \frac{m-1}{m}\right)^T\right)$. Le RSB est alors, pour chaque *endmember* j : RSB $= 20\log_{10}(s_{i,j})$. Les simulations sont les moyennes sur 100 tirages de type Monte-Carlo. $p = 4$, $m = 200$ et $N = 400$. Le seuil ζ_N est calculé pour $k = 10$, c'est à dire avec $\hat{\sigma}^2(10)$. Le tableau 2.1 montre, pour différents RSB, les résultats de la méthode, notés $\hat{p}$ par rapport à la méthode AIC, notée $\hat{p}_{AIC}$. Les variances, notées $v_{\hat{p}}$ et v_{AIC}, sont également calculées pour ces 100 itérations. La PFA est fixée à 0.01.

Le tableau 2.2 montre, pour différentes valeurs de PFA, les résultats obtenus pour les simulations faites avec les mêmes paramètres et un RSB de 27 dB.

Les résultats de ces tableaux sont en faveur de la méthode développée dans [116] et proposée ici pour le cas de l'imagerie hyperspectrale. La méthode AIC, quant à elle, ne donne pas de résultats suffisamment bons pour pouvoir être considérée comme une méthode fiable pour les images hyperspectrales.

2.3 Estimation du nombre d'*endmembers* pour un bruit corrélé, d'entrées indépendantes et identiquement distribuées

Cette section s'intéresse au cas où les données sont du type :

$$\mathbf{Y} = \mathbf{M}\mathbf{S} + \mathbf{C}^{1/2}\,\mathbf{X}\,, \tag{2.5}$$

ou $\mathbf{C}$ est une matrice hermitienne non-négative.

Dans cette section sont également reprises des méthodes déjà présentées ou qui ne sont pas encore adaptées au cas de l'imagerie hyperspectrale. Cette section expose donc des méthodes déjà existantes, pour permettre au lecteur de se rendre compte de ce qu'il est possible de faire, même si les deux premières méthodes n'ont pas de lien avec la suite.

L'hypothèse suivante sera nécessaire :

Hypothèse 4 réduite

- $\{c_k\}_{k\in[\![0,m-1]\!]}$ sont des coefficients absolument sommables, tels que $c_0 \neq 0$, et tels que la limite$\frac{1}{N}\sum \delta_{\lambda_i(\mathbf{C})}$ lorsque $N \to \infty$ existe et est notée ν.
- $\mathrm{dist}(\lambda_i(\mathbf{C}), \mathrm{Supp}(\nu)) \to 0$ avec ν la limite de $\frac{1}{N}\sum \delta_{\lambda_i(\mathbf{C})}$ lorsque $N \to \infty$,

Dans ce cas, il est possible de blanchir le signal avec une estimée de la matrice $\mathbf{C}$, puis de regarder la distribution des valeurs propres de la matrice de covariance estimée des données blanchies.

C'est ce qui est par exemple proposé dans [78], qui suppose cependant l'existence de données secondaires pour estimer correctement la matrice $\mathbf{C}$, c'est à dire qu'il est supposé qu'il existe des pixels de bruits sans *endmembers* dont la position est connue, et qui permettent, en ne considérant qu'eux, d'estimer la matrice de covariance du bruit d'une façon plus précise. Pour ce faire, les données sont d'abord blanchies à l'aide de la SCM des données secondaires. Une fois les données blanchies, la SCM est calculée ainsi que ses valeurs propres, notées $\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_K$, K étant le nombre de valeurs propres distinctes de la SCM des données blanchies. Pour une PFA choisie, notée α, il est ensuite possible de calculer le seuil suivant :

$$\zeta = TW^{-1}(1-\alpha)\,.$$

Alors le théorème suivant s'applique :

Théorème 2.3.1. *[78]*

Le nombre d'endmembers dans le signal est égal au plus grand $\hat{p} < \min(N,m)$ tel que :

$$\frac{\log m\,\dfrac{\hat{\lambda}_{\hat{p}+1}}{N} - \mu}{\sigma} < \zeta\,, \tag{2.6}$$

avec

$$\mu = \mu_{m,N,N'} = \frac{\dfrac{u_{\tilde{N}'}}{\tau_{\tilde{N}'}} + \dfrac{u_{\tilde{N}'-1}}{\tau_{\tilde{N}'-1}}}{\dfrac{1}{\tau_{\tilde{N}'}} + \dfrac{1}{\tau_{\tilde{N}'-1}}}\,,$$

$$\sigma = \sigma_{m,N,N'} = \frac{2}{\dfrac{1}{\tau_{\tilde{N}'}} + \dfrac{1}{\tau_{\tilde{N}'-1}}}\,,$$

et $\tilde{N}' = \min(m,N)$
N' *le nombre de données secondaires indépendantes*

$$\gamma_{\tilde{N}'} = 2\,\sin^{-1}\sqrt{\frac{\tilde{N}'+0.5}{2\,\tilde{N}'+N'-m+|N-m|+1}}$$

$$\phi_{\tilde{N}'} = 2\,\sin^{-1}\sqrt{\frac{\tilde{N}'+|N-m|+0.5}{2\,\tilde{N}'+N'-m+|N-m|+1}}$$

$$\tau_{\tilde{N}'} = \left(\frac{16}{\left(2\,\tilde{N}' + N' - m + |N-m| + 1\right)^2} \frac{1}{\sin^2(\phi_{\tilde{N}'} + \gamma_{\tilde{N}'}) \sin\phi_{\tilde{N}'} \sin\gamma_{\tilde{N}'}} \right)^{1/3}$$

$$u_{\tilde{N}'} = 2 \log\tan\left(\frac{\phi_{\tilde{N}'} + \gamma_{\tilde{N}'}}{2}\right)$$

Or ici, il n'est pas supposé qu'il y ait un accès à des données secondaires sans *endmembers*. Ainsi, cet algorithme, donné à titre informatif, ne peut être utilisé sur une image hyperspectrale où le bruit ne serait pas connu.

Estimer les valeurs propres du bruit est tout de même possible sans données secondaires, notamment dans le cas où la matrice $\mathbf{C}$ a des valeurs propres d'ordre de multiplicité égaux ou supérieurs à 1, comme proposé par l'algorithme G-MUSIC, de l'article [75]. Cet algorithme est une amélioration de l'algorithme MUSIC lorsque m et N sont grands, et de même ordre, c'est à dire lorsque le nombre d'échantillons disponibles n'est pas très grand par rapport à leur taille. Cet algorithme se construit à partir du théorème suivant, résumé ici dans le cas le plus général, et déjà introduit dans le chapitre 1 (1.6.7) :

Théorème 2.3.2. *[75]*

- *Soient les données* $\mathbf{Y} = \mathbf{C}^{1/2}\,\mathbf{X}$, *de moyennes nulles et de matrice de covariance* $\mathbf{C}$. $\mathbf{C}$ *a* K *valeurs propres distinctes* $\lambda_1 < \ldots < \lambda_K$ *de multiplicités respectives* $m_1, \ldots, m_K$.
- *Soit la matrice* $\check{\mathbf{C}} = \mathbf{C}^{1/2}\,\mathbf{X}\mathbf{X}^H\,\mathbf{C}^{1/2} \in \mathbb{C}^{m\times m}$.
- *Soient* $\check{\lambda}_1 \leq \ldots \leq \check{\lambda}_m$ *les valeurs propres de* $\check{\mathbf{C}}$ *et le vecteur* $\check{\boldsymbol{\lambda}} = \left(\check{\lambda}_1, \ldots, \check{\lambda}_m\right)^T$.
- *Soit* $k \in [\![1, K]\!]$, $\mathcal{N} = \left\{ \sum_{j=1}^{k-1} m_j + 1, \ldots, \sum_{j=1}^{k} m_j \right\}$, *qui correspondent aux ordres de multiplicités d'une même valeur propre.*
- *Soient* $\hat{\lambda}_k = \dfrac{N}{m_k} \sum_{i\in\mathcal{N}_k} (\check{\lambda}_i - \mu_i)$, *avec* $\mu_1 \leq \ldots \leq \mu_m$ *les valeurs propres de la matrice* $\mathrm{diag}(\check{\boldsymbol{\lambda}}) - \dfrac{1}{N}\sqrt{\check{\boldsymbol{\lambda}}}\sqrt{\check{\boldsymbol{\lambda}}}^T$.

Alors, sous l'hypothèse 3 (2.1) et sous des conditions de séparabilité sur les paquets des valeurs propres de $\mathbf{C}$ *[75],*

$$\forall k, \hat{\lambda}_k - \lambda \overset{p.s.}{\to} 0\,. \tag{2.7}$$

L'algorithme G-MUSIC consiste alors à calculer les valeurs propres estimées $\hat{\lambda}_k$ pour tout k pour en déduire celles n'appartenant pas au bruit. Un algorithme G-MUSIC robuste a, par ailleurs, été développé dans l'article [33], en remplaçant $\check{\mathbf{C}}$ par une estimation robuste de la matrice $\mathbf{C}$. Cet article montre en effet qu'à un facteur près, et dans le cas d'un bruit blanc corrélé, les valeurs propres de la SCM et d'un M-estimateur convergent les unes vers les autres presque sûrement. Le facteur prés dépend de la fonction u choisie pour le M-estimateur. Ainsi, dans le cas d'un bruit blanc corrélé, il est possible de lier la distribution des valeurs propres de la SCM est celle des M-estimateurs. Étudier les valeurs propres de l'un ou de l'autre de ces estimateurs devient alors équivalent.

Les autres approches possibles ont déjà été introduites dans le chapitre précédent, telles que l'étude des rapports des valeurs propres ou des différences entre les valeurs propres par exemple, et il est possible, sur une image hyperspectrale, de mettre en place le test suivant :

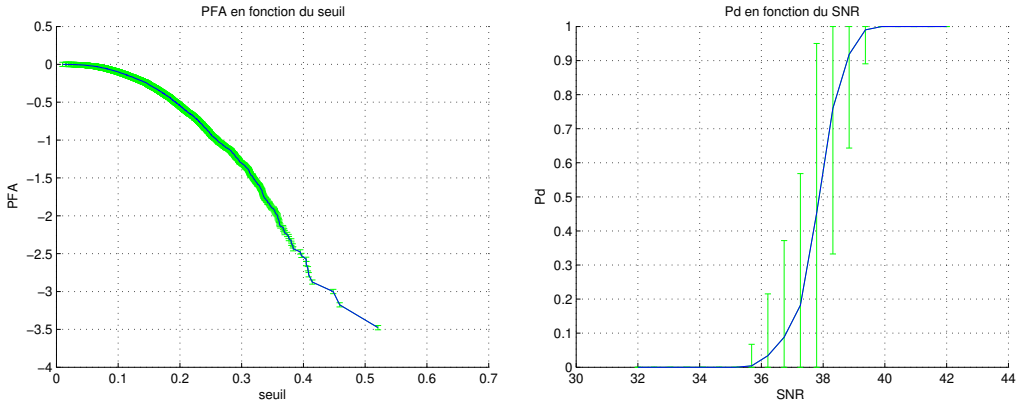

FIGURE 2.1 – à gauche : PFA (en $\log_{10}$) en fonction du seuil ζ pour 3000 itérations et à droite : probabilité de détection en fonction du RSB, pour 500 itérations, $m = 200$, $N = 1000$ et une PFA de 0.01, en vert les barres d'erreurs.

Théorème 2.3.3. *[116]*

Sous l'hypothèse 3 (2.1) et 4 réduite (2.3), en supposant de plus que le signal et le bruit ne sont pas corrélés à la fois dans dans l'espace et dans la dimension spectrale, et que $\frac{\|\mathbf{MS}\|}{N} < \infty$*, en notant* $\hat{\lambda}_0 \geq \ldots \geq \hat{\lambda}_m$ *les valeurs propres de la SCM des données* $\mathbf{Y}$*, alors*

$$\hat{p} = \arg\max_k \left(\frac{\hat{\lambda}_{k-1,N}}{\hat{\lambda}_{k,N}} > 1 + \zeta \right) , \tag{2.8}$$

pour $k \in [\![0, \ldots, L[\![$*,* L *un entier tel que* $L \geq p$ *et* $\hat{\lambda}_{-1,N} = \infty$*,* ζ *un seuil à fixer.*

Démonstration. La preuve se trouve dans [116]. □

L'inconvénient de cette méthode est le seuil ζ qui reste à déterminer. Cependant, il est toujours possible de déterminer ce seuil de façon heuristique, comme montré par les figures suivantes. Pour ces figures sont utilisées des images simulées, et, pour plusieurs tirages aléatoires de bruit, il est possible de tracer la PFA en fonction du seuil, sur la figure de gauche de 2.1. C'est à dire que pour chaque seuil est calculé le nombre de fausse alarme pour 3000 bruits tirés aléatoirement. Grâce à cette courbe, le seuil peut donc être choisi en fonction de la PFA voulue, pour l'image considérée. De même, il est possible de tracer la Probabilité de Détection (Pd) en fonction du RSB, toujours pour une image simulée, représenté à droite de la figure 2.1. Les performances obtenues permettent d'obtenir des résultats satisfaisants mais à condition de pouvoir simuler le bruit de l'image. Ces résultats ont été publiés dans [108].

Ces méthodes ne semblent pas suffisantes, soit parce qu'elles ne sont pas adaptées au modèle choisi, soit parce qu'il reste des indéterminations dans la méthode. Ces méthodes, développées pour les modèles corrélés, ne seront donc pas utilisées par la suite. Il semble préférable, dans le cas d'un modèle gaussien non corrélé et dans le cas d'une image hyperspectrale, d'utiliser une des méthodes développées pour un bruit CES et de considérer les τ comme des scalaires tous égaux à 1. En effet, le cas gaussien corrélé est un cas particulier du cas CES corrélé. Les méthodes développées dans le cas CES proposées par la suite sont donc également adaptées à tous les cas particuliers que le modèle CES regroupe, et ne présentent pas les inconvénients cités au dessus.

2.4 Estimation du nombre d'*endmembers* pour un bruit CES

Pour un bruit CES, et sous les hypothèses 1, 2 (voir 2.1) et 3 (2.1), le modèle considéré est le modèle (1.33), c'est à dire :

$$\mathbf{Y} = \mathbf{M}\,\mathbf{S} + \mathbf{C}^{1/2}\,\mathbf{X}\,\mathbf{T}^{1/2}\,,$$

ou

$$\mathbf{y}_i = \sum_{j=1}^{p} s_{i,j}\mathbf{m}_j + \sqrt{\tau_i}\,\mathbf{C}^{1/2}\,\mathbf{x}_i\,, \quad i \in [\![1, N]\!]\,. \tag{2.9}$$

Pour cette partie, quelques hypothèses supplémentaires sont nécessaires. Il est supposé que la matrice $\mathbf{S} = (s_{i,j})_{i,j} \in \mathbb{R}^{p\times N}$ peut se décomposer comme suit : $\mathbf{S} = \boldsymbol{\delta}^H\boldsymbol{\Gamma}^{1/2}$, avec, $\boldsymbol{\delta}$ une matrice aléatoire de taille $N \times p$, indépendante des éléments de $\mathbf{X}$ et $\boldsymbol{\Gamma}$ une matrice hermitienne de taille $N \times N$. L'hypothèse suivante est aussi prise en compte :

Hypothèse 4

- $\{c_k\}_{k\in[\![0,m-1]\!]}$ sont des coefficients absolument sommables, tels que $c_0 \neq 0$, et tels que la limite$\dfrac{1}{N}\sum \delta_{\lambda_i(\mathbf{C})}$ lorsque $N \to \infty$ existe et est notée ν.
- $\mathbf{T}$ a ses entrées indépendantes et identiquement distribuées, avec des moments d'ordre 1 finis.
- $\mathrm{dist}(\lambda_i(\mathbf{C}), \mathrm{Supp}(\nu)) \to 0$ avec ν la limite de $\dfrac{1}{N}\sum \delta_{\lambda_i(\mathbf{C})}$ lorsque $N \to \infty$,
- La mesure aléatoire $\mu_N = \dfrac{1}{N}\sum_{i=1}^{N}\delta_{\tau_i}$ vérifie $\displaystyle\int \tau\,\mu_N(d\tau) \xrightarrow{a.s.} 1$
- Les variables aléatoires τ_i suivent des distributions à queue non lourde telles qu'une distribution gaussienne, exponentielle, gamma, Student, de Pareto ou de Weibull.
- $\|\boldsymbol{\Gamma}\| < \infty$ et $\|\mathbf{M}\| < \infty$.

Pour estimer le nombre d'*endmembers*, la méthode proposée consiste à d'abord blanchir le signal par une estimation de la matrice de covariance du bruit, c'est à dire par un estimateur $\hat{\mathbf{C}}$, pour obtenir $\check{\mathbf{Y}}_w = \hat{\mathbf{C}}^{-1/2}\,\mathbf{M}\,\mathbf{S} + \hat{\mathbf{C}}^{-1/2}\,\mathbf{C}^{1/2}\,\mathbf{X}\,\mathbf{T}^{1/2}$. La manière d'estimer la matrice de covariance peut se faire avec des méthodes robustes ou non, ce qui sera présenté par la suite. Puis une estimation robuste de la matrice de covariance du signal blanchi est proposée. Cette étape de blanchiment, comme cela sera montré dans les résultats, est primordiale pour le bon fonctionnement de la méthode. Les valeurs propres de cette matrice sont ensuite analysées, seuillées, pour en déduire le nombre d'*endmembers*. Ces résultats, qui sont présentés dans les parties suivantes, ont fait l'objet de différentes publications [109], [110] et [107]. L'originalité de ces méthodes proposées vient du modèle considéré qui demande une étape de blanchiment supplémentaire. Cette étape de blanchiment doit ensuite être prise en considération pour la suite des algorithmes proposés.

2.4.1 Bruit CES "blanc"

Avant de présenter les solutions proposées, un cas déjà étudié dans les articles [34] et [29], correspondant à un modèle légèrement simplifié est décrit. Ce cas correspond au modèle suivant : $\mathbf{Y} = \mathbf{X}\,\mathbf{T}^{1/2}$.

Sous l'hypothèse 1 (2.1) et si le bruit est un bruit blanc gaussien, alors, comme présenté dans le premier chapitre, aucune valeur propre de la matrice de covariance du bruit estimée ne peut se trouver en dehors du support de la loi de Marchenko-Pastur. Cependant, la

présence de la matrice de texture $\mathbf{T}$ modifie ce résultat, et sans même tenir compte d'une hypothétique corrélation du bruit $\mathbf{C}$, des valeurs propres peuvent apparaître à droite du support de la loi. Il est donc nécessaire de déterminer un autre seuil qui assure que les valeurs propres qui lui sont supérieures sont dues aux *endmembers*. Cela dit, il n'est pas assuré que tous les *endmembers* donneront des valeurs propres dont la valeur dépassera ce seuil, puisque cela dépend de leur RSB ; mais cela assure que si des valeurs propres sont situées au delà, elles seront forcément dues à la présence d'*endmembers*.

Soient les données, notées $\mathbf{Y} = \mathbf{X}\,\mathbf{T}^{1/2}$, ou $\mathbf{y}_i = \sqrt{\tau_i}\,\mathbf{x}_i$ pour tout $i \in [\![1, N]\!]$, ce qui correspond, par rapport au modèle présenté, au bruit CES sans la matrice de covariance $\mathbf{C}$. Il n'est donc pas nécessaire de blanchir le bruit. Les hypothèses 1 (2.1), 3 (2.1) et 4 (2.4) sont considérées, en plus de l'hypothèse suivante :

Hypothèse 5

- Les données $\mathbf{x}_1, \ldots, \mathbf{x}_N$ sont supposées indépendantes des variables aléatoires $\tau_1, \ldots, \tau_N$,
- Il existe un $\epsilon < 1 - \Phi_\infty^{-1} < 1 - c$ et $n > 0$, tel que, pour N assez grand, presque sûrement, $\mu_N([0, n)) < \epsilon$.

Il est de plus supposé que $\|\mathbf{x}_i\|_F^2 = m$ pour tout $i \in [\![1, N]\!]$, ou que les x_i sont égaux à $\sqrt{m}\,\dfrac{\tilde{\mathbf{x}}_i}{\|\tilde{\mathbf{x}}_i\|_F}$ avec $\tilde{\mathbf{x}}_i \in \mathbb{C}^m$ gaussien.

Pour estimer la matrice de covariance du bruit et limiter les effets de la texture sur la distribution des valeurs propres (la finalité étant de se rapprocher d'une distribution connue ou exploitable comme la loi de Marchenko-Pastur), un estimateur robuste de $\mathbf{\Sigma}$ est utilisé. Il correspond à l'unique solution de l'équation suivante et est noté $\hat{\mathbf{\Sigma}}$:

$$\mathbf{\Sigma} = \frac{1}{N}\sum_{i=1}^{N} u\left(\frac{1}{m}\,\mathbf{y}_i^H\,\mathbf{\Sigma}^{-1}\,\mathbf{y}_i\right)\,\mathbf{y}_i\mathbf{y}_i^H\,, \tag{2.10}$$

où la fonction u vérifie les hypothèses de Maronna (1.8), légèrement modifiées :

Hypothèses de Maronna modifiées

(i) $u : [0, +\infty) \mapsto (0, +\infty)$ positive, continue et décroissante,

(ii) $\phi : x \mapsto x\,u(x)$ est strictement croissante, bornée, et vérifie $\lim\limits_{x\to\infty} \phi(x) = \Phi_\infty > 1$,

(iii) $\lim\limits_{N\longrightarrow\infty} c_N < \Phi_\infty^{-1}$.

L'hypothèse (ii) est légèrement plus restrictive que les hypothèses classiques de Maronna (1.8) puisque la fonction ϕ doit être strictement croissante, et ceci est nécessaire dans certaines des preuves des théorèmes suivants. Enfin, une dernière hypothèse nécessaire ici est la suivante :

pour tout $a > b > 0$, alors $\limsup\limits_{t\to\infty} \dfrac{\limsup \nu((t,\infty))}{\Phi(a\,t) - \Phi(b\,t)} = 0$ p.s. .

Cette hypothèse permet de contrôler la vitesse relative de la queue de distribution des variables aléatoires $\tau_1, \ldots, \tau_N$ par rapport à la vitesse d'atténuation de la fonction Φ, et permet d'assurer p.s. que les valeurs propres extrêmes dues à la matrice $\mathbf{T}$ seront correctement atténuées.

L'idée est donc d'étudier la distribution des valeurs propres de l'unique solution de l'équation du M-estimateur de Maronna choisi, c'est à dire de l'équation (2.10). Cependant, la

dépendance du terme $u\left(\frac{1}{m}\,\mathbf{y}_i^H\,\mathbf{\Sigma}^{-1}\,\mathbf{y}_i\right)$ aux vecteurs $\mathbf{y}_i$ ne permet pas d'étudier cette distribution spectrale [34], [91]. L'article [34] s'applique donc à reformuler le problème pour aboutir à une forme équivalente sans cette dépendance. Ceci passe par la réécriture de $\hat{\mathbf{\Sigma}}$ en enlevant un terme dépendant des $\mathbf{y}_i$:

$$\hat{\mathbf{\Sigma}}_i = \hat{\mathbf{\Sigma}} - \frac{1}{N} u\left(\frac{1}{m}\,\mathbf{y}_i^H\,\hat{\mathbf{\Sigma}}^{-1}\,\mathbf{y}_i\right)\,\mathbf{y}_i\mathbf{y}_i^H\,.$$

$\hat{\mathbf{\Sigma}}_i$ ne dépend de $\mathbf{y}_i$ que par $\hat{\mathbf{\Sigma}}$ qui est défini comme l'unique solution de (2.10), équation qui dépend des $\mathbf{y}_i$. Mais cette dépendance des données est assez faible et des approximations peuvent être faites pour s'en affranchir. La réécriture de ce terme montre, en fin de compte, que trouver l'unique solution de (2.10), ainsi que montrer son existence et son unicité, revient à faire de même pour la nouvelle écriture de $\hat{\mathbf{\Sigma}}$:

$$\hat{\mathbf{\Sigma}} = \frac{1}{N}\sum_{i=1}^{N} \tau_i\, v\left(\tau_i\, d_i\right)\,\mathbf{x}_i\,\mathbf{x}_i^H\,, \tag{2.11}$$

avec $d_i = \frac{1}{m}\,\mathbf{x}_i^H\,\hat{\mathbf{\Sigma}}_i\,\mathbf{x}_i$ et $v : x \mapsto (u \circ h^{-1})(x)$, $h : x \mapsto \frac{x}{1 - c\,\Phi(x)}$. Il ne reste ensuite qu'à trouver un équivalent pour d_i, en faisant les approximations nécessaires, afin de s'affranchir de la dépendance des termes $\mathbf{x}_i$ de l'équation (2.10) et permettre l'étude des valeurs propres de la solution de cette équation. L'article [34] montre alors qu'il existe une matrice $\hat{\mathbf{S}}$ qui ne dépend pas de d_i, dont la distribution spectrale est connue, telle que :

$$\left\|\hat{\mathbf{\Sigma}} - \hat{\mathbf{S}}\right\| \overset{p.s.}{\to} 0\,. \tag{2.12}$$

La matrice $\hat{\mathbf{S}}$ est définie comme suit :

$$\hat{\mathbf{S}} = \frac{1}{N}\sum_{i=1}^{N} v\left(\tau_i\,\xi\right)\,\tau_i\,\mathbf{x}_i\,\mathbf{x}_i^H\,, \tag{2.13}$$

avec ψ la fonction $\psi : x \mapsto x\,v(x)$, ξ donnée par la solution de :

$$\frac{1}{N}\sum_{i=1}^{N} \frac{\psi(\tau_i\,\xi)}{1 + c_N\,\psi(\tau_i\,\xi)} = 1\,. \tag{2.14}$$

Alors, la distribution spectrale de la matrice $\hat{\mathbf{\Sigma}}$ est connue : sa transformée de Stieltjes se définit comme l'unique solution de l'équation suivante, sur $\mathbb{C}^+$:

$$m(z) = \left(-z + \xi^{-1}\,\frac{1}{N}\sum_{i=1}^{N} \frac{\psi(\tau_i\,\xi)}{1 + c\,\xi^{-1}\,\psi(\tau_i\,\xi)\,m(z)}\right)^{-1}\,. \tag{2.15}$$

Le support de cette distribution est borné, et a déjà été présenté dans la partie introduction (1.43) pour un cas plus simple. Il est ainsi possible de trouver un seuil t tel qu'aucune valeur propre de $\hat{\mathbf{S}}$ ne dépasse se seuil. En effet, comme $\tau_i\,v(\tau_i\,\xi) = \xi^{-1}\,\psi(\tau_i\,\xi) < \xi^{-1}\,\psi_\infty$ avec $\psi_\infty = \frac{\phi_\infty}{(1 - c\,\phi_\infty)}$, alors :

$$\hat{\mathbf{S}} \preceq \frac{\phi_\infty}{(1 - c\,\phi_\infty)}\,\frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i\,\mathbf{x}_i^H\,. \tag{2.16}$$

Puis, d'après [7], la borne supérieure de la loi de Marchenko-Pastur étant $(1 + \sqrt{c})^2$:

$$\left\|\hat{\mathbf{S}}\right\| \leq \frac{\phi_\infty\,(1 + \sqrt{c})^2}{\xi\,(1 - c\,\phi_\infty)}\,. \tag{2.17}$$

Il est donc possible de choisir le seuil t comme suit : [29]

$$t = \frac{\phi_\infty \, (1+\sqrt{c})^2}{\xi \, (1 - c \, \phi_\infty)} \, . \tag{2.18}$$

Ce seuil prend bien en compte la distribution des variables aléatoires $\tau_1, \ldots, \tau_N$ à travers ξ.

Dans le cas où p *endmembers* sont ajoutés au bruit, c'est à dire pour $\mathbf{Y} = \mathbf{M}\,\mathbf{S} + \mathbf{X}\,\mathbf{T}^{1/2}$, si $p \ll N$, alors les résultats précédents ne sont pas modifiés [29], à part qu'un nombre p de valeurs propres peuvent dépasser le seuil t, sous réserve d'avoir un RSB assez élevé. En effet, la matrice de covariance estimée contient une partie signal et une partie bruit. La partie bruit est contrôlée par les résultats précédents, et la distribution de ces valeurs propres est notée μ, tandis que la partie signal peut présenter des valeurs propres au delà du seuil t, les parties bruit et signal étant supposées indépendantes, leurs valeurs propres le sont aussi. La position de ces valeurs propres en fonction de leur RSB est décrite dans [29], et dépend de la distribution ν.

Ces résultats sont résumés ci-après pour information (puisqu'en pratique, la distribution des variables aléatoires $\tau_1, \ldots, \tau_N$ n'est pas connue) : pour des *endmembers* de la forme $\sum_{j=1}^{p} s_{i,j}\mathbf{m}_j$ pour $i \in [\![1, N]\!]$, avec $\mathbf{m}_j = \sqrt{\mathrm{RSB}_j}\mathbf{a}_j$, RSB_j le RSB du $j$ième *endmember*, alors les valeurs propres de $\hat{\mathbf{\Sigma}}$: λ_j, $j \in \mathcal{J}$ tel que $\mathcal{J} = \{j \ / \ \lambda_j > t\}$ vérifient :

$$\lambda_j \overset{p.s.}{\to} \Lambda_j > t \text{ pour tout } j \text{ tel que } \lambda_j > t \, , \tag{2.19}$$

où Λ_j sont les uniques solutions positives de

$$-c \left(\delta(\Lambda_j) \int \frac{v(\tau\,\xi)}{1 + \delta(\Lambda_j)\,\tau\,v(\tau\,\xi)} \, \nu(d\tau) \right)^{-1} = \mathrm{RSB}_j \, ,$$

et, pour tout $x \in \mathbb{R} \ \ [\inf(Supp(\mu)), \sup(Supp(\mu))]$:

$$\delta(x) = c \left(-x + \int \frac{v\,(t\,\xi)}{1 + \delta(\Lambda)\,t\,v\,(t\,\xi)} \, \nu(dt) \right) \, .$$

Cela dit, l'article [29] propose une méthode pour estimer ces positions dans le cas où la distribution des τ est inconnue. Dans ce cas, il s'agit d'utiliser les valeurs observées des τ_i en remplaçant dans les équations précédentes, les intégrales par des sommes empiriques. Mais encore faut il avoir accès aux valeurs de ces τ_i, ce qui n'est pas le cas, en pratique, ici.

L'estimation de p se fait donc de la manière suivante :

$$\hat{p} = \sharp\{\mathcal{J}\} \, . \tag{2.20}$$

2.4.2 Approche non-robuste pour un bruit CES corrélé

Le modèle considéré est de nouveau celui présenté en introduction, sans la présence de sources, c'est à dire : $\mathbf{Y} = \mathbf{C}^{1/2}\,\mathbf{X}\,\mathbf{T}^{1/2}$. La plupart des résultats présentés ci-après peuvent être vu comme une généralisation de l'article [34].

Étape de Blanchiment

L'approche non robuste consiste à calculer d'abord $\hat{\mathbf{C}}_{SCM} = \frac{1}{N}\,\mathbf{Y}\,\mathbf{Y}^H$ qui correspond ainsi à la SCM. Une fois cet estimateur calculé, la matrice $\mathbf{C}$ étant de Toeplitz, l'estimateur biaisé suivant peut être considéré (l'opérateur $\mathcal{T}$ défini dans les notations) :

$$\tilde{\mathbf{C}}_{SCM} = \mathcal{T}\left(\hat{\mathbf{C}}_{SCM}\right) \, , \tag{2.21}$$

ou, autrement écrit,

$$\left[\tilde{\mathbf{C}}_{SCM}\right]_{i,j} = [\mathcal{L}(\tilde{\mathbf{c}}_{SCM})]_{i,j} \quad \text{avec} \quad \tilde{c}_{SCM,k} = \frac{1}{m\,N} \sum_{i=0}^{m-1} \sum_{j=0}^{N-1} y_{i,j}\, y^{\star}_{i+k,j}\, \mathbb{1}_{0\leq i+k<m}\,. \tag{2.22}$$

L'estimateur est biaisé dans le sens où les derniers termes de la matrice se rapprochent de zéro pour m grand. Cela dit, l'avantage de cet estimateur est qu'il donne de meilleurs résultats que l'estimateur non biaisé qui consisterait à diviser chaque somme d'éléments diagonaux de la matrice $\hat{\mathbf{C}}_{SCM}$ par le nombre d'élément sur la diagonale. En effet, les erreurs d'estimations sont minimisées par la moyenne faite sur chaque diagonale. Or, comme moins il y a d'éléments sur la diagonale, moins l'estimation de la diagonale sera consistante, une solution est de faire tendre ces éléments vers zéro, pour leur donner en quelque sorte moins de poids. De plus, la présence des p *endmembers* sera en quelque sorte atténuée par cette étape, de la même manière que les erreurs d'estimations sont minimisées. Ainsi, cette étape a le double avantage d'ajouter de la robustesse à la méthode et de minimiser l'influence des sources dans l'estimation de la matrice de covariance du bruit.

Le théorème suivant établit la consistance de cet estimateur :

Théorème 2.4.1. *Consistence de* $\tilde{\mathbf{C}}_{SCM}$

Sous les hypothèses ***1****,* ***2*** *(2.1),* ***3*** *(2.1) et* ***4*** *(2.4), et pour* $\mathbf{Y} = \mathbf{C}^{1/2}\,\mathbf{X}\,\mathbf{T}^{1/2}$*, le résultat suivant s'applique :*

$$\| \tilde{\mathbf{C}}_{SCM} - \mathbb{E}[\tau]\mathbf{C} \| \overset{p.s.}{\to} 0\,. \tag{2.23}$$

La matrice définie par $\check{\mathbf{C}}_{SCM} = \dfrac{1}{\mathbb{E}[\tau]}\,\tilde{\mathbf{C}}_{SCM}$ *caractérise l'estimation biaisée de Toeplitz de* $\mathbf{C}$*.*

Démonstration.

La preuve de ce théorème est donnée en Annexe B. Il est à noter que la dernière hypothèse prise en compte dans la partie précédente 2.4.1 n'est ici pas nécessaire puisque la preuve s'appuie sur d'autres propriétés de comportement des plus grandes valeurs propres de certaines distribution, et assure p.s. la correcte atténuation des valeurs propres.

□

Dans le cadre plus général où $\mathbf{Y} = \mathbf{M}\,\mathbf{S} + \mathbf{C}^{1/2}\,\mathbf{X}\,\mathbf{T}^{1/2}$, le résultat est biaisé. Le biais dépend ne sera pas étudié dans la preuve mais peut être développé en suivant la preuve de [119] . Afin d'utiliser ce théorème, en pratique, il est donc nécessaire de disposer de données secondaires pour estimer correctement et séparément la partie liée aux *endmembers* afin de pouvoir la soustraire aux données $\mathbf{Y}$.

La figure 2.2 trace $\left\| \tilde{\mathbf{C}}_{SCM} - \mathbb{E}[\tau]\,\mathbf{C} \right\|$ pour des N allant de 0 à 6000. Le signal est constitué d'un bruit blanc gaussien $\mathbf{X}$ corrélé par une matrice de Toeplitz $\mathbf{C} = \mathcal{L}\left(\left(\rho^0, \rho^1, \ldots, \rho^{m-1}\right)^T\right)$ avec $\rho = 0.7$. Les textures $\{\tau_i\}_{i\in[\![0,N-1]\!]}$ sont des variables aléatoires de lois inverse gamma de moyenne unité. La figure 2.2 représente la norme spectrale de la différence entre la vraie matrice de covariance $\mathbf{C}$ et $\check{\mathbf{C}}_{SCM}$ pour $c = 0.7$. Le taux de convergence de 2.4.1 est assez lent. Même s'il semble possible d'améliorer cette vitesse de convergence par exemple avec des méthodes de "banding" ou de "tapering" comme proposé dans l'article [15], ce problème n'est ici pas l'intérêt principal et ne sera pas traité.

La figure 2.3 montre, sur la moyenne de 20 échantillons type Monte Carlo, les mêmes résultats que sur la figure 2.2, pour le même bruit mais pour $c = 0.2$ et sur une échelle logarithmique. Est rajoutée, sur cette figure, la différence en norme spectrale entre la SCM et la vraie matrice $\mathbf{C}$. Ainsi, la SCM ne renvoie pas du tout les bonnes valeurs propres et aucune étude des valeurs propres ne peut être menée sur la SCM.

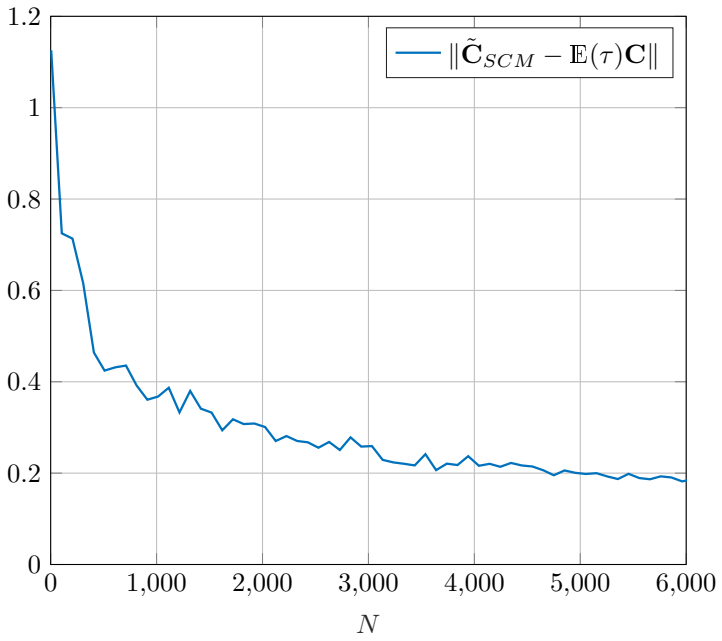

FIGURE 2.2 – Illustration du Théorème 2.4.1 pour $c = 0.7$ et pour des $\{\tau_i\}_{i\in[\![1,N]\!]}$ suivant une loi inverse gamma ($N \in [\![20, 6000]\!]$)

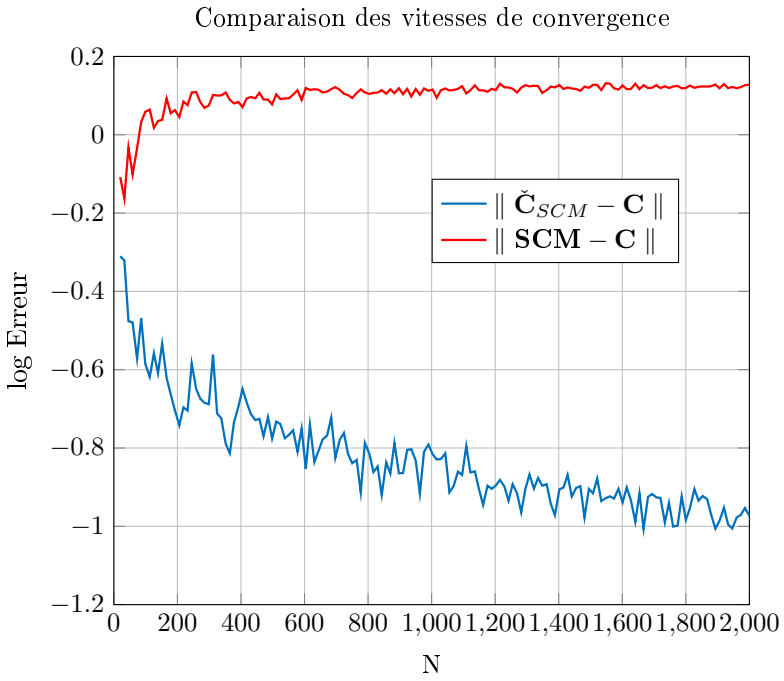

FIGURE 2.3 – Comparaison de la convergence entre $\check{\mathbf{C}}_{SCM}$ et la SCM, pour $c = 0.2$, $\rho = 0.7$, échelle logarithmique.

Une fois la matrice $\check{\mathbf{C}}_{SCM}$ estimée, le signal est blanchi : $\check{\mathbf{Y}}_{wSCM} = \check{\mathbf{C}}_{SCM}^{-1/2}\,\mathbf{M}\,\mathbf{S} + \check{\mathbf{C}}_{SCM}^{-1/2}\,\mathbf{C}^{1/2}\,\mathbf{X}\,\mathbf{T}^{1/2}$, avec $\check{\mathbf{Y}}_{wSCM} = [\check{\mathbf{y}}_{wS0}, ..., \check{\mathbf{y}}_{wS\,N-1}]$.

Estimation de la matrice de covariance

Il est ensuite possible d'utiliser ces données blanchies pour estimer leur matrice de covariance $\mathbb{E}[\mathbf{X}\mathbf{T}\mathbf{X}^H]$ et ensuite étudier la distribution théorique de ses valeurs propres. Cette estimation s'effectue à l'aide d'un estimateur robuste, et ceci, pour éviter que des valeurs isolées mais élevées de la texture $\mathbf{T}$ ne viennent altérer la structure de la matrice de covariance estimée. L'estimateur robuste choisi est du même type que présenté en introduction, parfois appelé estimateur du Point Fixe (FP) bien que cette dénomination corresponde en fait à un cas particulier de ces M-estimateurs où la fonction u est la fonction inverse. Ici, cette dénomination correspondra au cas général des M-estimateur, et l'estimateur dit de Tyler au cas particulier de la fonction inverse. (Cette double dénomination se retrouve dans la littérature sur le sujet et peut parfois porter à confusion.) Il est noté $\check{\mathbf{\Sigma}}_{SCM}$ et défini comme l'unique solution de cette équation :

$$\mathbf{\Sigma} = \frac{1}{N}\sum_{i=1}^{N} u\left(\frac{1}{m}\,\check{\mathbf{y}}_{wSi}^H\,\mathbf{\Sigma}^{-1}\,\check{\mathbf{y}}_{wSi}\right)\,\check{\mathbf{y}}_{wSi}\check{\mathbf{y}}_{wSi}^H\,, \tag{2.24}$$

où la fonction u vérifie les hypothèses de Maronna modifiées (2.4.1), c'est à dire, pour rappel :

(i) $u : [0, +\infty(\mapsto (0, +\infty($ positive, continue et décroissante,

(ii) $\phi : x \mapsto x\,u(x)$ est strictement croissante, bornée, et vérifie $\lim\limits_{x\to\infty} \phi(x) = \Phi_\infty > 1$,

(iii) $\lim\limits_{N\longrightarrow\infty} c_N < \Phi_\infty^{-1}$,

sachant que pour cette partie, les hypothèses de Maronna (1.8) suffisent également.

Il s'agit ensuite d'étudier la distribution des valeurs propres de cette matrice.

Une fois la matrice $\mathbf{Y}$ blanchie, il est possible d'étudier la distribution des valeurs propres de $\check{\mathbf{\Sigma}}_{SCM}$, toujours en exploitant la matrice $\hat{\mathbf{S}}$ pour permettre de s'affranchir de la dépendance du terme $u\left(\frac{1}{m}\,\mathbf{y}_i^H\,\mathbf{\Sigma}^{-1}\,\mathbf{y}_i\right)$ aux vecteurs $\mathbf{y}_i$. La matrice $\hat{\mathbf{S}}$ est la même que celle de la partie précédente, et vérifie l'équation (2.13). Soit $\mathbf{Y}_w$ la matrice telle que : $\mathbf{Y}_w = [\mathbf{y}_{w0}, \ldots, \mathbf{y}_{w\,N-1}] = \mathbf{C}^{-1/2}\,\mathbf{M}\,\mathbf{S} + \mathbf{X}\,\mathbf{T}^{1/2}$. Dans ce cas, il a déjà été prouvé, dans l'article [29], le théorème suivant :

$$\left\|\hat{\mathbf{\Sigma}} - \hat{\mathbf{S}}\right\| \overset{p.s.}{\to} 0\,, \tag{2.25}$$

où $\hat{\mathbf{\Sigma}}$ est l'unique solution de l'équation :

$$\mathbf{\Sigma} = \frac{1}{N}\sum_{i=1}^{N} u\left(\frac{1}{m}\,\mathbf{y}_{wi}^H\,\mathbf{\Sigma}^{-1}\,\mathbf{y}_{wi}\right)\,\mathbf{y}_{wi}\mathbf{y}_{wi}^H\,. \tag{2.26}$$

Alors le théorème 2.4.2 permet d'établir la relation entre les matrices $\hat{\mathbf{S}}$ et $\check{\mathbf{\Sigma}}_{SCM}$:

Théorème 2.4.2. *Convergence de la matrice $\check{\mathbf{\Sigma}}_{SCM}$*
Sous les mêmes hypothèses que précédemment, la convergence suivante est établie :

$$\left\|\check{\mathbf{\Sigma}}_{SCM} - \hat{\mathbf{S}}\right\| \overset{p.s.}{\to} 0\,. \tag{2.27}$$

Démonstration. La preuve de ce théorème, qui est également la preuve du théorème 2.4.4, est donnée en Annexe B. □

Sélection de l'ordre du modèle

Ainsi, la distribution des valeurs propres de la matrice $\check{\mathbf{\Sigma}}_{SCM}$ étant liée à celle de $\hat{\mathbf{S}}$, il devient possible, en utilisant des résultats sur la matrice $\hat{\mathbf{S}}$, d'obtenir le nombre d'*endmembers* ou de sélectionner l'ordre du modèle. Les résultats de la partie précédente s'appliquent alors, et un seuil t vérifiant la même équation que précédemment, c'est à dire :

$$t = \frac{\phi_\infty (1+\sqrt{c})^2}{\xi \left(1 - c\,\phi_\infty\right)} ,$$

peut être appliqué sur la distribution des valeurs propres de la matrice $\check{\mathbf{\Sigma}}_{SCM}$. Soient $\check{\lambda}_{SCM\,i}$, $i \in [\![1, m]\!]$ les valeurs propres de $\check{\mathbf{\Sigma}}_{SCM}$. Soit $\mathcal{J}_{SCM} = \{i \in [\![1, m]\!] \,/\, \check{\lambda}_{SCM\,i} > t\}$. Le nombre d'*endmembers* est alors :

$$\hat{p} = \sharp\{\mathcal{J}_{SCM}\} . \tag{2.28}$$

Une autre façon équivalente de noter ce nombre d'*endmembers*, en triant les valeurs propres dans l'ordre croissant : $\check{\lambda}_{SCM\,1} \geq \ldots \geq \check{\lambda}_{SCM\,m}$, est :

$$\hat{p} = \min_k \left(\check{\lambda}_{SCM\,k} > t\right) . \tag{2.29}$$

La figure 2.4 montre la correspondance entre la distribution des valeurs propres de $\check{\mathbf{\Sigma}}_{SCM}$ et de $\hat{\mathbf{S}}$ lorsqu'il n'y a pas d'*endmembers* dans le signal, et illustre le théorème 2.4.2. Le seuil t, qui dépend de la fonction u par la fonction Φ_∞, est également traçé, en rouge. La fonction u est choisie pour correspondre à l'EMV pour des variables aléatoires τ_i, $i \in [\![1, N]\!]$, distribués selon une loi inverse gamma. Ainsi, $u : x \mapsto \dfrac{1+\alpha}{x+\alpha}$. Le paramètre α est fixé à 0.1. Le seuil t a alors l'expression suivante : $t = \dfrac{(1+\alpha)\,(1+\sqrt{c})^2}{\xi\,\left(1 - c\,(1+\alpha)\right)}$.

Comme la distribution des valeurs propres de $\check{\mathbf{\Sigma}}_{SCM}$ suit celle de $\hat{\mathbf{S}}$, aucune valeur propre ne dépasse le seuil t. L'utilisation d'un estimateur robuste pour l'estimation de la matrice $\mathbf{C}$ permet donc d'annihiler correctement l'influence des τ qui peuvent engendrer de grandes valeurs propres.

La figure 2.5 montre que si l'étape de blanchiment n'est pas suivie, les valeurs propres de l'unique solution de

$$\mathbf{\Sigma} = \frac{1}{N} \sum_{i=1}^{N} u\left(\frac{1}{m}\, \mathbf{y}_i^H\, \mathbf{\Sigma}^{-1}\, \mathbf{y}_i\right) \mathbf{y}_i \mathbf{y}_i^H , \tag{2.30}$$

notée $\check{\mathbf{\Sigma}}$, ne permettent pas de vérifier le théorème 2.4.2. Les valeurs propres de $\check{\mathbf{\Sigma}}$ ne suivent en effet pas la même distribution que la matrice $\hat{\mathbf{S}}$, et le seuil est dépassé par un grand nombre de valeurs propres alors qu'aucun *endmember* n'est présent dans le signal. Enfin, la figure 2.6 montre les mêmes résultats mais pour une fonction u qui ne correspond pas à la distribution des variables τ_i, $i \in [\![1, N[\![$. En effet, pour rappel, la fonction u peut être choisie de telle façon à correspondre à l'EMV pour une distribution des τ_i connue. Dans le cas où la fonction u ne correspond pas à l'EMV, la méthode est toujours valable puisqu'aucune hypothèse ne requiert cette condition pour que le théorème puisse être appliqué. Ceci est ainsi illustré par la figure 2.6, où le cas choisi est assez éloigné de la distribution qui correspondrait à l'EMV pour la fonction choisie. Cependant, la méthode semble moins fiable dans le sens où

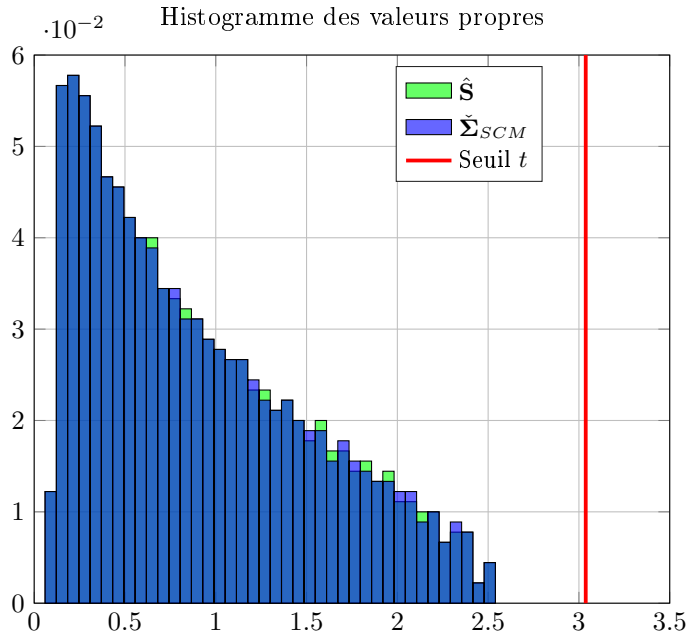

FIGURE 2.4 – Valeurs propres des matrices $\check{\mathbf{\Sigma}}_{SCM}$ et $\hat{\mathbf{S}}$ lorsque le signal $\mathbf{Y}$ a été blanchi par $\check{\mathbf{C}}_{SCM}$ et le seuil correspondant t pour $\rho = 0.7$, $m = 900$, $N = 2000$, $\tau =$ inverse gamma, $\alpha = 0.1$, $p = 0$.

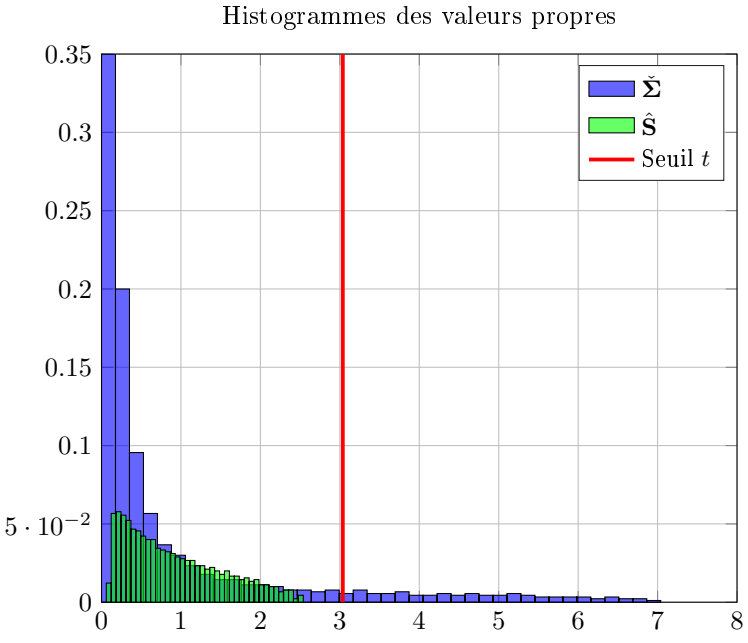

FIGURE 2.5 – Valeurs propres des matrices $\check{\mathbf{\Sigma}}$ et $\hat{\mathbf{S}}$ lorsque le signal $\mathbf{Y}$ n'a pas été blanchi par $\check{\mathbf{C}}_{SCM}$, et le seuil correspondant t pour $\rho = 0.7$, $m = 900$, $N = 2000$, $\tau =$ inverse gamma, $\alpha = 0.1$, $p = 0$.

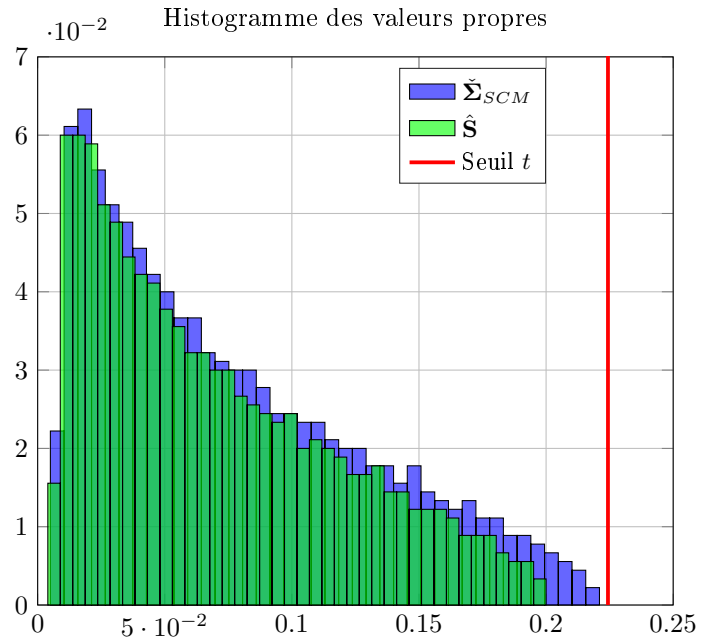

FIGURE 2.6 – Valeurs propres des matrices $\check{\mathbf{\Sigma}}_{SCM}$ et $\hat{\mathbf{S}}$ et le seuil t, pour $\rho = 0.7$, $m = 900$, $N = 2000$, $\tau = t^2 - student$, $\alpha = 0.1$, $p = 0$.

les valeurs propres de $\check{\mathbf{\Sigma}}_{SCM}$ se rapprochent du seuil, car la convergence vers les valeurs propres de la matrice $\hat{\mathbf{S}}$ est moins rapide. Il faut donc un N (et donc un m) plus grand pour assurer la superposition des deux distributions.

De plus, comme, en pratique, la distribution des variables aléatoires $\tau_1, \dots \tau_N$ est inconnue, il semble judicieux d'améliorer la méthode, ce qui est fait par la suite.

2.4.3 Approche robuste pour un bruit CES corrélé

Au lieu d'utiliser la SCM préalablement forcée à être de Toeplitz (équation 2.21) pour estimer la matrice de covariance du bruit, cette partie présente les résultats dans le cas où l'estimateur choisi serait un M-estimateur de Maronna. Cette fois-ci, la matrice $\mathbf{Y}$ contient des sources $\mathbf{MS}$. La première sous-partie présente le cas général des M-estimateurs. La seconde, quant à elle, propose d'utiliser un estimateur de Tyler. Les étapes sont les mêmes que celles de la partie précédente : une étape de blanchiment, une étape d'estimation et une étape de seuillage des valeurs propres. Cette approche a fait l'objet de deux publications, [110] et [107].

Cas Général

Étape de Blanchiment

Soit $\widetilde{\mathbf{C}}_{FP}$ un estimateur biaisé de la matrice de covariance $\mathbf{C}$ tel que $\widetilde{\mathbf{C}}_{FP} = \mathcal{T}(\widehat{\mathbf{C}}_{FP})$ où $\widehat{\mathbf{C}}_{FP}$ est l'unique solution de l'équation suivante (correspondant toujours aux M-estimateurs de Maronna [74]) :

$$\mathbf{Z} = \frac{1}{N} \sum_{i=0}^{N-1} u\left(\frac{1}{m} \mathbf{y}_i^H \mathbf{Z}^{-1} \mathbf{y}_i\right) \mathbf{y}_i \mathbf{y}_i^H .$$

Tout comme dans la partie précédente, la fonction u vérifie les hypothèses de Maronna modifiées (2.4.1). Ceci correspond au même estimateur que $\tilde{\mathbf{C}}_{SCM}$ mais pour un estimateur de type Maronna. Alors le théorème suivant s'applique pour $\widetilde{\mathbf{C}}_{FP}$:

Théorème 2.4.3. *Consistance de $\widetilde{\mathbf{C}}_{FP}$*
*Soit $\widetilde{\mathbf{C}}_{FP}$ l'estimateur défini ci-dessus, alors, sous les hypothèses **1**, **2** (2.1), **3** (2.1) et **4** (2.4), et pour $\mathbf{Y} = \mathbf{C}^{1/2}\,\mathbf{X}\,\mathbf{T}^{1/2}$ le résultat suivant est vrai :*

$$\left\| \widetilde{\mathbf{C}}_{FP} - \mathbb{E}\left[v(\tau\,\xi)\,\tau\right]\,\mathbf{C} \right\| \xrightarrow{a.s.} 0\,, \tag{2.31}$$

où v et ξ sont définis comme précédemment. Pour rappel :

- *u vérifie les hypothèses de Marronna (2.4.1)*
- $h : x \mapsto \dfrac{x}{1 - c_N\,\Phi(x)}$
- $v : x \mapsto (u \circ h^{-1})(x)$
- $\psi : x \mapsto x\,v(x)$
- *ξ est l'unique solution de l'équation*

$$\frac{1}{N}\sum_{i=1}^{N} \frac{\psi(\tau_i\,\xi)}{1 + c_N\,\psi(\tau_i\,\xi)} = 1\,. \tag{2.32}$$

La matrice $\check{\mathbf{C}}_{FP} = \dfrac{\widetilde{\mathbf{C}}_{FP}}{\mathbb{E}\left[v(\tau\,\xi)\,\tau\right]}$ caractérise l'estimation de la matrice de covariance $\mathbf{C}$.

Démonstration. La preuve, inspirée de [119] et de [66], est donnée en Annexe B. □

Première remarque : [32] montre que $\tilde{\mathbf{C}}_{SCM} = \phi^{-1}(1)\,\tilde{\mathbf{C}}_{FP}$, avec ϕ définie comme précédemment, c'est à dire $\phi(x) = x\,u(x)$, et $\tilde{\mathbf{C}}_{SCM}$ définie en 2.21. Lorsque la fonction u est bien choisie, il est possible d'obtenir $\phi^{-1}(1) = 1$ et $\tilde{\mathbf{C}}_{SCM} = \tilde{\mathbf{C}}_{FP}$, comme c'est le cas par exemple pour la fonction $u : x \mapsto \dfrac{1+\alpha}{x+\alpha}$, où α est un paramètre à fixer. Mais même dans ce cas, $\check{\mathbf{C}}_{SCM}$ et $\check{\mathbf{C}}_{FP}$ diffèrent d'un facteur d'échelle puisque $\check{\mathbf{C}}_{SCM} = \dfrac{\mathbb{E}\left[v(\tau\,\xi)\,\tau\right]}{\mathbb{E}(\tau)}\,\check{\mathbf{C}}_{FP}$.

Seconde remarque : le résultat peut être considéré comme vrai dans le cas où $\mathbf{Y} = \mathbf{M}\,\mathbf{S} + \mathbf{C}^{1/2}\,\mathbf{X}\,\mathbf{T}^{1/2}$. En effet, la partie signal sera fortement atténuée par l'estimation robuste de la matrice. Le biais résultant est calculé de manière exacte dans [67]. Ce biais ne dépend pas du SNR du signal, et ne sera donc pas impacté par la puissance des sources.

Afin d'illustrer le théorème 2.4.3, la figure 2.7 montre la consistance de l'estimateur $\check{\mathbf{C}}_{FP}$ pour un c fixé, lorsque $N \to \infty$. Les paramètres de la figure sont les mêmes que ceux de la figure 2.2 et sont rappelés ici : le signal est constitué d'un bruit blanc gaussien $\mathbf{X}$ corrélé par une matrice de Toeplitz $\mathbf{C} = \mathcal{L}\left(\left(\rho^0, \rho^1, \ldots, \rho^{m-1}\right)^T\right)$ avec $\rho = 0.7$. Les texture $\{\tau_i\}_{i\in[\![0,N-1]\!]}$ sont des variables aléatoires de loi inverse gamma de moyenne unité. La figure 2.7 représente la norme spectrale de la différence entre la vraie matrice de covariance $\mathbf{C}$ et $\check{\mathbf{C}}_{FP}$ lorsque $c = 0.7$.

Remarque : La convergence observée sur la figure 2.7 est légèrement plus rapide que celle de la figure 2.2 même si cette vitesse de convergence reste assez faible.

Comme dans la partie précédente, les données $\mathbf{Y}$ peuvent être blanchies avec $\check{\mathbf{C}}_{FP}^{-1/2}$. Soient $\check{\mathbf{Y}}_{wFP} = [\check{\mathbf{y}}_{wF0}, \ldots \check{\mathbf{y}}_{wF\,N-1}]$ les données blanchies, c'est à dire :

$$\begin{aligned} \check{\mathbf{Y}}_{wFP} &= \check{\mathbf{C}}_{FP}^{-1/2}\,\mathbf{Y}\,, \\ \text{soit : } \check{\mathbf{Y}}_{wFP} &= \check{\mathbf{C}}_{FP}^{-1/2}\,\mathbf{M}\,\boldsymbol{\delta}^H\,\boldsymbol{\Gamma}^{1/2} + \check{\mathbf{C}}_{FP}^{-1/2}\,\mathbf{C}^{1/2}\,\mathbf{X}\,\mathbf{T}^{1/2}\,. \end{aligned}$$

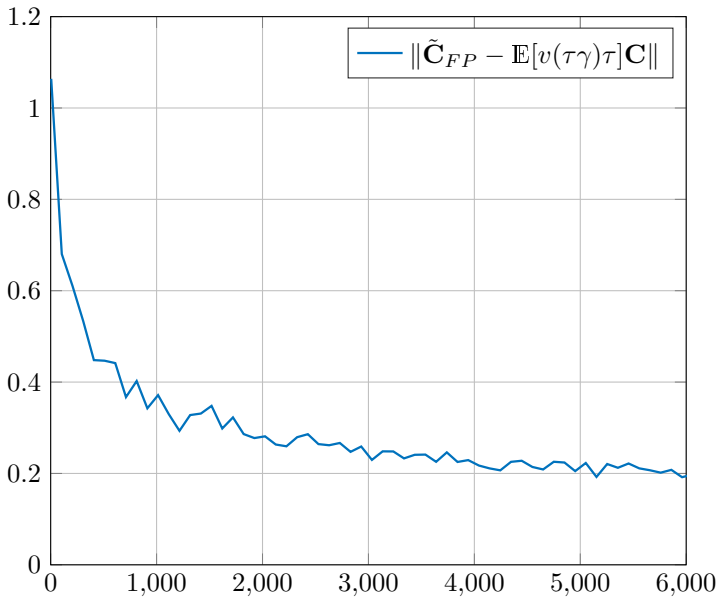

FIGURE 2.7 – Illustration de la consistance du Théorème 2.4.3 pour $c = 0.7$ et pour des $\{\tau_i\}_{i\in[\![1,N]\!]}$ suivant une loi inverse gamma ($N \in [\![20, 6000]\!]$).

Estimation de la matrice de covariance

L'estimation robuste de la matrice de covariance et celle du nombre d'*endmembers* (qui correspond à l'ordre du modèle) sont menées comme précédemment. L'estimateur utilisé est toujours un M-estimateur de Maronna, noté $\check{\mathbf{\Sigma}}_{FP}$ et est défini comme l'unique solution de l'équation suivante :

$$\mathbf{\Sigma} = \frac{1}{N}\sum_{i=0}^{N-1} u\left(\frac{1}{m}\,\check{\mathbf{y}}_{wFi}^{H}\,\mathbf{\Sigma}^{-1}\,\check{\mathbf{y}}_{wFi}\right)\,\check{\mathbf{y}}_{wFi}\,\check{\mathbf{y}}_{wFi}^{H}\,. \tag{2.33}$$

De cette façon, $\check{\mathbf{\Sigma}}_{FP}$ est un estimateur robuste de la matrice de covariance du signal blanchi. Un équivalent au théorème 2.4.2 peut être proposé :

Théorème 2.4.4. *Convergence de la matrice* $\check{\mathbf{\Sigma}}_{FP}$
Sous les mêmes hypothèses que précédemment, la convergence suivante est établie :

$$\left\|\check{\mathbf{\Sigma}}_{FP} - \hat{\mathbf{S}}\right\| \xrightarrow{p.s.} 0\,. \tag{2.34}$$

Démonstration. La preuve, fournie en Annexe B, est la même que pour le théorème 2.4.2. □

Sélection de l'ordre du modèle

Le même seuil t que celui vérifiant l'équation (2.18) peut donc être retenu sur la distribution des valeurs propres de la matrice $\check{\mathbf{\Sigma}}_{FP}$, et ce, grâce au Théorème 2.4.4. Ainsi, le dénombrement du nombre d'*endmembers* ou l'estimation de l'ordre du modèle se fait comme précédemment, c'est à dire, en considérant les valeurs propres de $\check{\mathbf{\Sigma}}_{FP}$, notées $\check{\lambda}_{FP\,i}$, $i \in [\![1, m]\!]$, et avec $\mathcal{J}_{FP} = \{i \in [\![1, m]\!] \;/\; \check{\lambda}_{FP\,i} > t\}$. Le nombre d'*endmembers* est alors estimé selon :

$$\hat{p} = \sharp\{\mathcal{J}_{FP}\}\,, \tag{2.35}$$

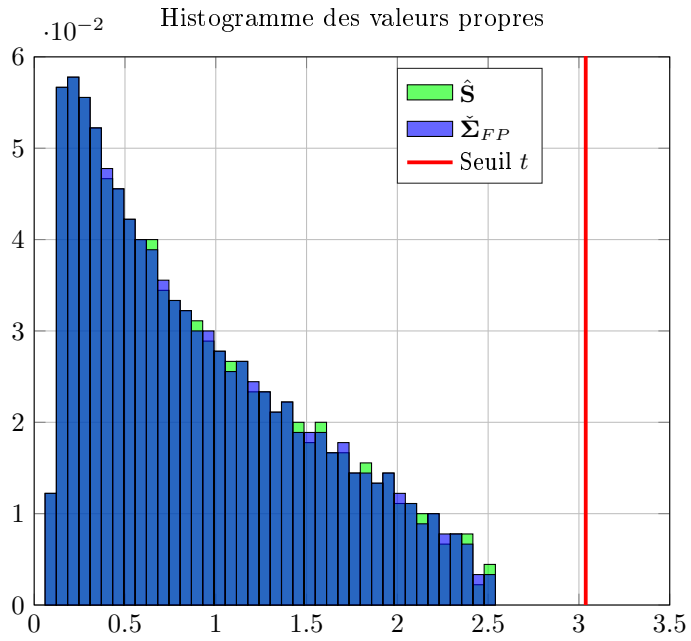

FIGURE 2.8 – Distribution des valeurs propres des matrices $\hat{\mathbf{S}}$ et $\check{\mathbf{\Sigma}}_{FP}$ ainsi que le seuil t pour $\rho = 0.7$, $m = 900$, $N = 2000$, $\tau \sim$ inverse gamma.

ou encore, en classant les valeurs propres par ordre décroissant : $\check{\lambda}_{FP\,1} \geq \ldots \geq \check{\lambda}_{FP\,m}$:

$$\hat{p} = \min_k \left(\check{\lambda}_{FP\,k} > t\right). \tag{2.36}$$

Les figures suivantes représentent différentes distributions de valeurs propres, pour différents cas de figure. Pour chaque figure, les paramètres sont les suivants $\rho = 0.7$, $m = 900$ et $N = 2000$ soit $c = 0.45$. La fonction u est choisie pour correspondre à l'EMV pour des variables aléatoires $\tau_1, \ldots, \tau_N$ distribuées selon une loi inverse gamma. Ainsi, $u : x \mapsto \dfrac{1+\alpha}{x+\alpha}$, et α est encore choisi égal à 0.1. La figure 2.8 représente la distribution des valeurs propres associées aux matrices $\hat{\mathbf{S}}$ et $\check{\mathbf{\Sigma}}_{FP}$. Sur cette figure, les deux distributions sont parfaitement superposées et correspondent, tout comme sur la figure 2.4.

Quant au cas où la fonction u choisie ne correspond pas à l'EMV, la méthode proposée donne des résultats qui semblent plus robustes, comme le montre la figure 2.9 qui peut être comparée à la figure 2.6 ou, sur cette figure, les variables aléatoires τ_i suivent une distribution t^2, où t est une t-distribution de Student. La figure 2.9 montre que les valeurs propres de $\hat{\mathbf{S}}$ et $\check{\mathbf{\Sigma}}_{FP}$ ont des distributions toujours bien superposées.

Soit la matrice $\tilde{\mathbf{\Sigma}}_{FP} = \check{\mathbf{C}}_{FP}^{1/2}\, \check{\mathbf{\Sigma}}_{FP}\, \check{\mathbf{C}}_{FP}^{1/2}$, c'est à dire la matrice correspondant à la solution d'un M-estimateur sans avoir au préalable blanchi les données. Alors dans ce cas, le théorème ne s'applique plus, ce qui est visible sur la figure 2.10. Sur cette figure, le seuil ne dépasse pas la plus grande valeur propre du bruit. En présence d'*endmembers*, la méthode surestimerait le nombre de spectres purs contenus dans l'image.

La figure 2.11, quant à elle, montre la distribution des valeurs propres de la matrice $\hat{\mathbf{S}}$ et celle de la SCM des données blanchies avec l'estimateur $\check{\mathbf{C}}_{FP}$ proposé pour estimer $\mathbf{C}$. Dans ce cas, le théorème 2.4.4 n'est plus valable, et les distributions des valeurs propres des deux matrices ne correspondent pas, comme illustré par la figure 2.11.

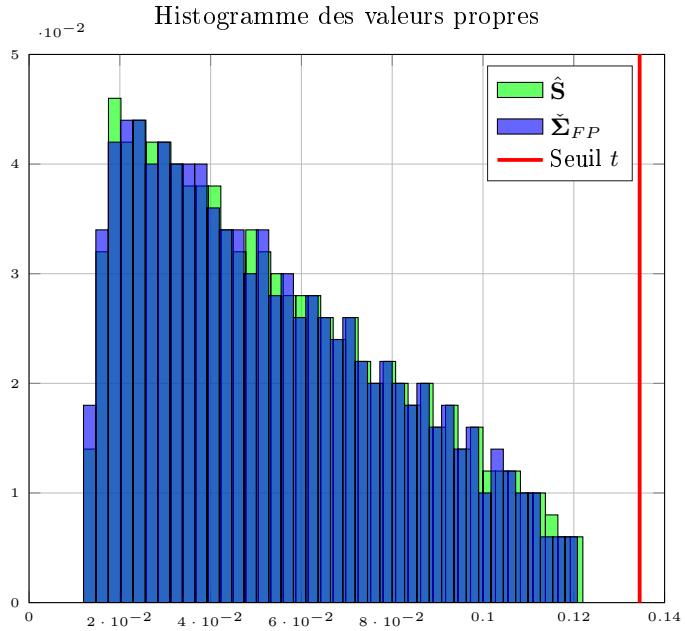

FIGURE 2.9 – Illustration de la robustesse au choix de u. Distribution des valeurs propres des matrices $\hat{\mathbf{S}}$ et $\check{\mathbf{\Sigma}}_{FP}$ et le seuil t pour $\rho = 0.7$, $m = 900$, $N = 2000$, $\tau = t^2$, $t \sim$ Student

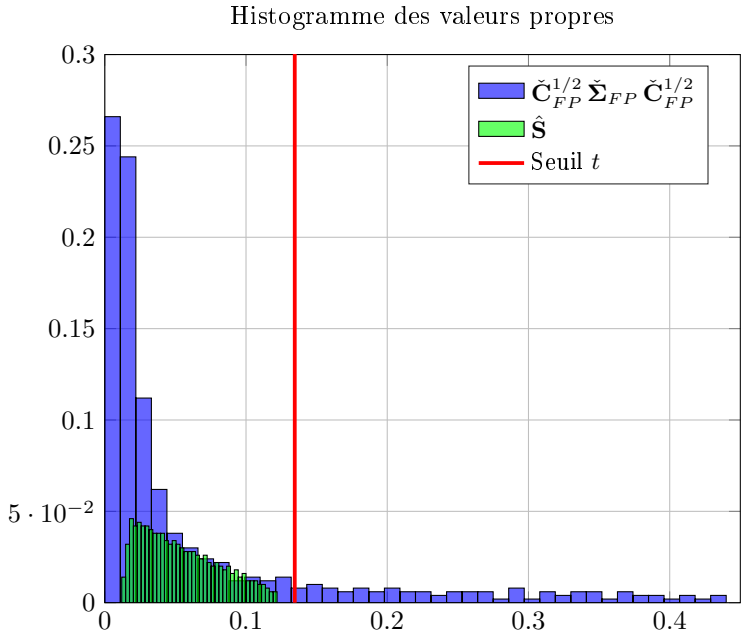

FIGURE 2.10 – Distribution des valeurs propres des matrices $\hat{\mathbf{S}}$ et $\tilde{\mathbf{\Sigma}}_{FP}$ pour un signal non blanchi ainsi que le seuil t pour $\rho = 0.7$, $m = 900$, $N = 2000$, $\tau = t^2$, $t \sim$ Student.

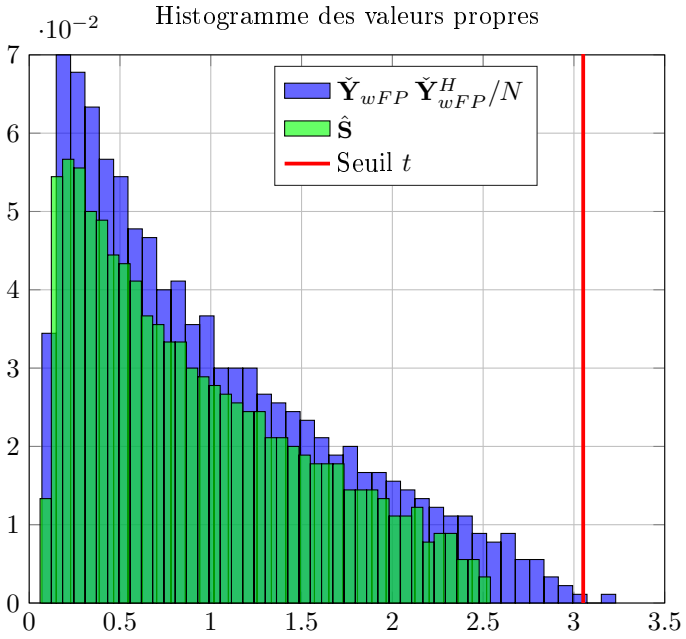

FIGURE 2.11 – Distribution des valeurs propres des matrices $\hat{\mathbf{S}}$ et de la matrice représentant la SCM de $\check{\mathbf{Y}}_{wFP}\,\check{\mathbf{Y}}_{wFP}^H/N$ et le seuil t pour $\rho = 0.7$, $m = 900$, $N = 2000$, $\tau \sim$ inverse gamma.

Enfin, la figure 2.12 montre la distribution des valeurs propres pour les mêmes matrices $\hat{\mathbf{S}}$ et $\check{\mathbf{\Sigma}}_{FP}$ mais cette fois-ci en présence d'un *endmember*, d'un RSB de 10 dB. Le spectre $\mathbf{m}_1$ de cet endmember est simulé de la manière suivante : $[\mathbf{m}]_{j,1} = 10^{\mathrm{RSB}/20} \exp\left(2i\pi \frac{j-1}{m}\right)$, pour $j \in [\![1, m]\!]$. Cet endmember engendre bien la présence d'une valeur propre au delà du seuil, et sera donc détecté. Les autres valeurs propres restent bien en dessous du seuil.

Ainsi, il est possible de connaître et de caractériser un seuil sur la distribution des valeurs propres pour en déduire le nombre d'endmembers. Comme beaucoup de méthodes reposent sur l'étude des valeurs propres, ce nouvel estimateur améliore la consistance de ces méthodes.

Cas de l'estimateur de Tyler

Le cas de l'estimateur de Tyler est un cas particulier de la méthode précédente. En effet, au lieu de considérer n'importe quelle fonction u, il est toujours possible de choisir $u : x \mapsto \frac{1}{x}$. Dans ce cas, les résultats du cas précédent ne s'appliquent pas tous puisque les hypothèses de Maronna modifiées (paragraphe 2.4.1) ne sont pas toutes vérifiées. Cependant, les hypothèses de Maronna du paragraphe 1.8 sont elles, bien vérifiées et il est possible d'obtenir le même type de résultats, plus facilement exploitables. Les étapes des parties précédentes sont donc ici reprises (blanchiment, estimation de la matrice de covariance et seuillage des valeurs propres) mais développées pour le cas particulier de l'estimateur de Tyler, ou aussi appelé estimateur du Point Fixe (FP). Attention toutefois encore une fois à ne pas confondre avec la notation $\check{\mathbf{\Sigma}}_{FP}$ qui représente bien l'estimée dans le cas général d'un M-estimateur (la confusion vient du fait que le Point-Fixe admet dans la communauté du traitement du signal plusieurs définitions). Certains résultats peuvent se simplifier et mènent à d'autres théorèmes présentés ci-dessous. La suite s'appuie sur les récents résultats publiés dans les articles [127] et [33].

Étape de Blanchiment

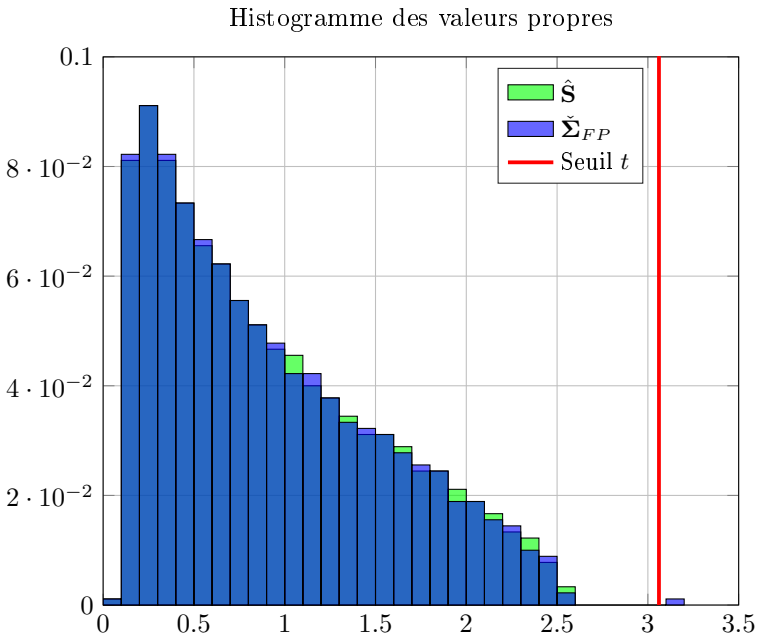

FIGURE 2.12 – Distribution des valeurs propres des matrices $\hat{\mathbf{S}}$ et $\check{\mathbf{\Sigma}}_{FP}$ lorsqu'un endmembers est présent dans la signal, $RSB = 10dB$ ainsi que le seuil t, pour $\rho = 0.7$, $m = 900$, $N = 2000$, $\tau \sim$ inverse gamma.

La première étape est toujours de blanchir le signal, cette fois-ci avec un M-estimateur de Tyler, qui est ensuite rectifié, c'est à dire forcé à être de Toeplitz. Le M-estimateur de Tyler $\hat{\mathbf{C}}_{TYL}$ vérifie l'équation suivante [87] :

$$\mathbf{\Sigma} = \frac{m}{N} \sum_{i=1}^{N} \frac{\mathbf{y}_i \mathbf{y}_i^H}{\mathbf{y}_i^H \mathbf{\Sigma}^{-1} \mathbf{y}_i} \,. \tag{2.37}$$

Cette équation, comme les équations des M-estimateurs précédents, nécessite une approche itérative pour estimer $\hat{\mathbf{C}}_{TYL}$ à partir de la valeur finale de $\mathbf{\Sigma}$. Puis, $\hat{\mathbf{C}}_{TYL}$ est ensuite rectifié en utilisant l'opérateur $\mathcal{T}$ défini précédemment (équation 2.21), afin d'obtenir $\check{\mathbf{C}}_{TYL} = \mathcal{T}\left(\hat{\mathbf{C}}_{TYL}\right)$. Alors le théorème suivant donne :

Théorème 2.4.5. *Consistance de $\check{\mathbf{C}}_{TYL}$*
*Sous les hypothèses **1**, **2** (paragraphe 2.1), **3** (paragraphe 2.1) et **4** (paragraphe 2.4), la convergence suivante est vraie :*

$$\left\| \check{\mathbf{C}}_{TYL} - \mathbf{C} \right\| \overset{p.s.}{\to} 0 \,. \tag{2.38}$$

Démonstration. La démonstration se trouve en Annexe B et dans l'article [107]. □

La figure 2.13 montre la consistance de cet estimateur pour $c = 0.7$ et des τ_i distribués selon une loi inverse gamma. La décroissance est plus rapide que pour les deux théorèmes précédent 2.4.1 et 2.4.3. Cette méthode semble donc a priori justifiée.

Une fois que la matrice $\check{\mathbf{C}}_{TYL}$ a été calculée, il est possible de blanchir le signal de la

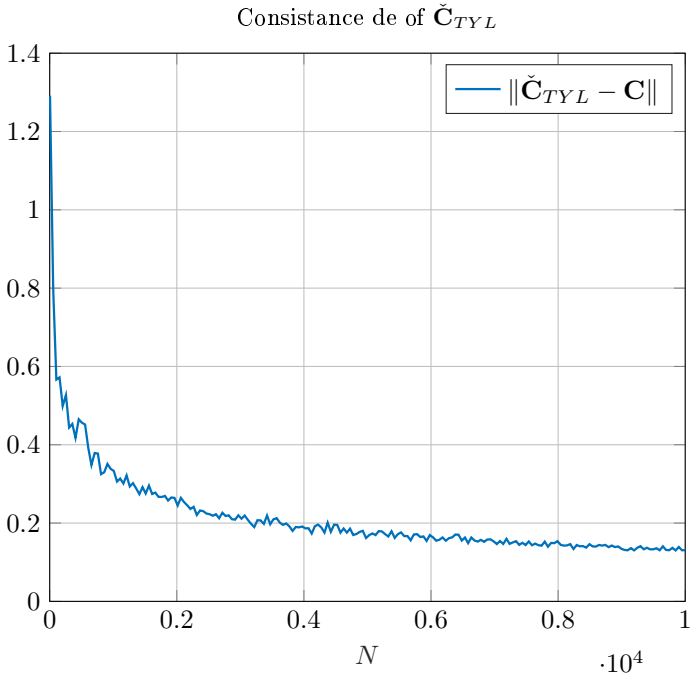

FIGURE 2.13

même manière que précédemment, c'est à dire :

$$\check{\mathbf{Y}}_{wTYL} = \check{\mathbf{C}}_{TYL}^{-1/2}\,\mathbf{Y}\,,$$
$$\text{soit : } \check{\mathbf{Y}}_{wTYL} = \check{\mathbf{C}}_{TYL}^{-1/2}\,\mathbf{M}\,\boldsymbol{\delta}^H\,\boldsymbol{\Gamma}^{1/2} + \check{\mathbf{C}}_{TYL}^{-1/2}\,\mathbf{C}^{1/2}\,\mathbf{X}\,\mathbf{T}^{1/2}\,,$$

et $\check{\mathbf{Y}}_{wTYL} = [\check{\mathbf{y}}_{wT1}, \ldots, \check{\mathbf{y}}_{wTN}]$.

Estimation de la matrice de covariance

Une fois le signal blanchi, il est possible d'utiliser le même estimateur que celui utilisé pour estimer $\mathbf{C}$, mais cette fois-ci, pour estimer la matrice de dispersion du signal blanchi. Soit $\check{\boldsymbol{\Sigma}}_{TYL}$ l'unique solution de l'équation suivante :

$$\boldsymbol{\Sigma} = \frac{m}{N}\sum_{i=1}^{N}\frac{\check{\mathbf{y}}_{wTi}\check{\mathbf{y}}_{wTi}^H}{\check{\mathbf{y}}_{wTi}^H\,\boldsymbol{\Sigma}^{-1}\,\check{\mathbf{y}}_{wTi}}\,. \tag{2.39}$$

Soit également la matrice $\hat{\mathbf{S}}_w$, définie comme :

$$\hat{\mathbf{S}}_w = \frac{1}{N}\,\mathbf{X}\,\mathbf{X}^H\,. \tag{2.40}$$

Alors les valeurs propres de $\hat{\mathbf{S}}_w$ suivent la loi de Marchenko-Pastur. Le théorème suivant permet alors de relier la distribution des valeurs propres de $\check{\boldsymbol{\Sigma}}_{TYL}$ à la distribution de Marchenko-Pastur :

Théorème 2.4.6. *Consistance de la matrice* $\check{\boldsymbol{\Sigma}}_{TYL}$

Sous les mêmes hypothèses que précédemment, l'équation suivante s'applique :

$$\left\|\check{\boldsymbol{\Sigma}}_{TYL} - \hat{\mathbf{S}}_w\right\| \overset{p.s.}{\to} 0\,. \tag{2.41}$$

Démonstration. La démonstration se trouve en Annexe B. Elle s'appuie sur la démonstration du Théorème 2.4.4 et se trouve également dans l'article [107]. □

Sélection de l'ordre du modèle

Ainsi, comme la distribution des valeurs propres de $\hat{\mathbf{S}}_w$ converge en loi vers la loi de Marchenko-Pastur, dont le support est connu et borné, la distribution des valeurs propres de $\check{\mathbf{\Sigma}}_{TYL}$ converge presque sûrement vers cette distribution. Si certains *endmembers* ont un RSB suffisamment élevé, il sera possible de les détecter. Le support de la loi de Marchenko-Pastur étant borné par $(1+\sqrt{c})^2$, il est possible de choisir cette valeur comme seuil sur les valeurs propres de $\check{\mathbf{\Sigma}}_{TYL}$. Le nombre de valeurs propres qui dépassent de ce seuil sera donc lié au nombre d'*endmembers*. L'estimation du nombre d'*endmembers* se fait donc comme précédemment : soient $\check{\lambda}_{TYL\,i}$ les valeurs propres de $\check{\mathbf{\Sigma}}_{TYL}$ pour $i \in [\![1,m]\!]$. Soit $\mathcal{J}_{TYL} = \{i \in [\![1,m]\!] \,/\, \check{\lambda}_{TYL\,i} > t\}$. Alors p peut s'estimer comme suit :

$$\hat{p} = \sharp\{\mathcal{J}_{TYL}\}\,, \tag{2.42}$$

ou encore, si les valeurs propres sont classées par ordre décroissant : $\check{\lambda}_{TYL\,1} \geq \ldots \geq \check{\lambda}_{TYL\,m}$:

$$\hat{p} = \min_k \left(\check{\lambda}_{TYL\,k} > t\right). \tag{2.43}$$

Les figures suivantes permettent d'illustrer les résultats théoriques qui viennent d'être présentés. Sur ces figures, le bruit est un bruit CES, où les variables aléatoires $\tau_1, \ldots, \tau_N$ sont distribuées suivant une loi inverse gamma pour les figures 2.14 et 2.16 et suivant une loi t^2, où t suit une loi de Student pour la figure 2.15. Sur la figure 2.14 sont tracées les valeurs propres de $\check{\mathbf{\Sigma}}_{TYL}$ pour un signal contenant 4 *endmembers* avec un RSB = 10 dB. Ces *endmembers* sont simulés comme précédemment, c'est à dire que leurs spectres, pour $k \in [\![1,4]\!]$ correspondant aux 4 *endmembers*, sont $[\mathbf{m}]_{j,k} = 10^{\mathrm{RSB}/20} \exp\left(2i\pi \frac{(k-1)(j-1)}{m}\right)$, pour $j \in [\![1,m]\!]$. Même si ce RSB peut sembler faible, il permet cependant aux 4 sources d'être détectées, puisque les valeurs propres correspondantes sont au-delà du seuil de la loi de Marchenko-Pastur.

Pour les données permettant de tracer la figure 2.15, le bruit est donc différent mais elles contiennent aussi 4 *endmembers* avec un RSB = 30 dB, et simulés comme pour la figure 2.14. Les valeurs propres sont plus fortes que sur la figure 2.14 puisque les RSB sont plus élevés. De plus, la méthode est beaucoup plus robuste aux différents types de bruits, comme ici, où les variables aléatoires τ_i ne suivent pas la même distribution que sur la figure 2.14. En effet, la fonction u choisie, c'est à dire ici $u : x \to \dfrac{1}{x}$ rend l'estimation insensible aux facteurs multiplicatifs des données puisque $\dfrac{m}{N}\displaystyle\sum_{i=1}^{N} \frac{\tau_i\,\mathbf{y}_i\,\mathbf{y}_i^H}{\sqrt{\tau_i}\,\mathbf{y}_i^H\,\mathbf{\Sigma}^{-1}\,\sqrt{\tau_i}\,\mathbf{y}_i} = \frac{m}{N}\displaystyle\sum_{i=1}^{N} \frac{\mathbf{y}_i\,\mathbf{y}_i^H}{\mathbf{y}_i^H\,\mathbf{\Sigma}^{-1}\,\mathbf{y}_i}$ pour tout scalaire τ_i, $i \in [\![1,N]\!]$. L'estimateur est donc parfaitement robuste aux différents types de τ choisis, ce qui peut se remarquer sur la figure 2.15.

La figure 2.16, quant à elle, superpose les valeurs propres de la loi de Marchenko-Pastur, et les valeurs propres de l'estimateur proposé dans le cas où le bruit est blanchi à l'aide d'un estimateur du Point fixe où la fonction u est choisie égale à $\dfrac{1+t}{t+x}$, $t = 0.1$. Dans ce cas, les deux distributions de valeurs propres ne correspondent plus et il n'est pas possible de considérer la limite droite de la loi de Marchenko-Pastur comme seuil pour déterminer quelles valeurs propres appartiennent au bruit.

En utilisant un estimateur de Tyler, la méthode est rendue plus robuste, puisque les résultats ne dépendent plus de la distribution des $\tau_{i,\,i\in[\![1,N]\!]}$. Elle est aussi moins laborieuse,

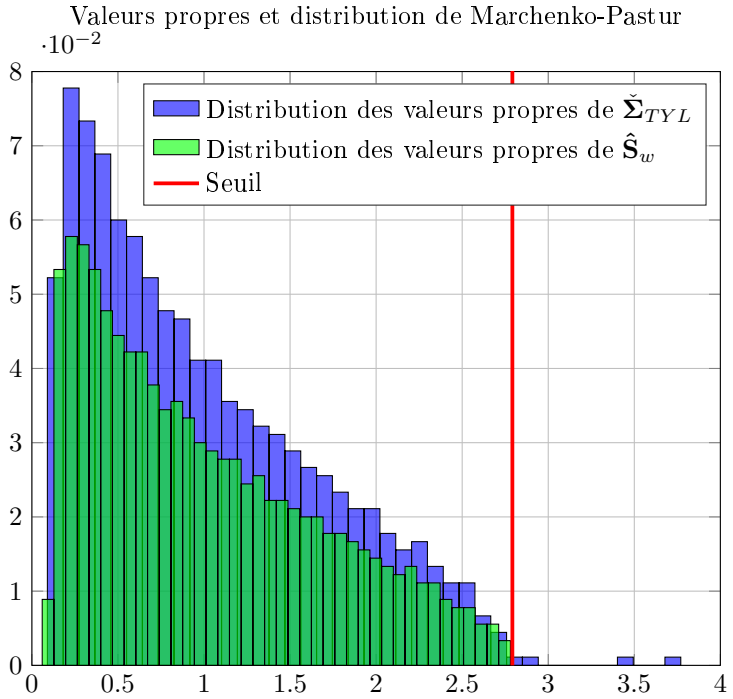

FIGURE 2.14 – Distribution de Marchenko-Pastur et distribution des valeurs propres de $\check{\mathbf{\Sigma}}_{TYL}$ pour 4 *endmembers* de RSB=10 dB, $N = 2000$, $m = 900$, $c = 0.7$, les $\{\tau\}_{i\in[\![1,N]\!]}$ suivant une loi inverse gamma.

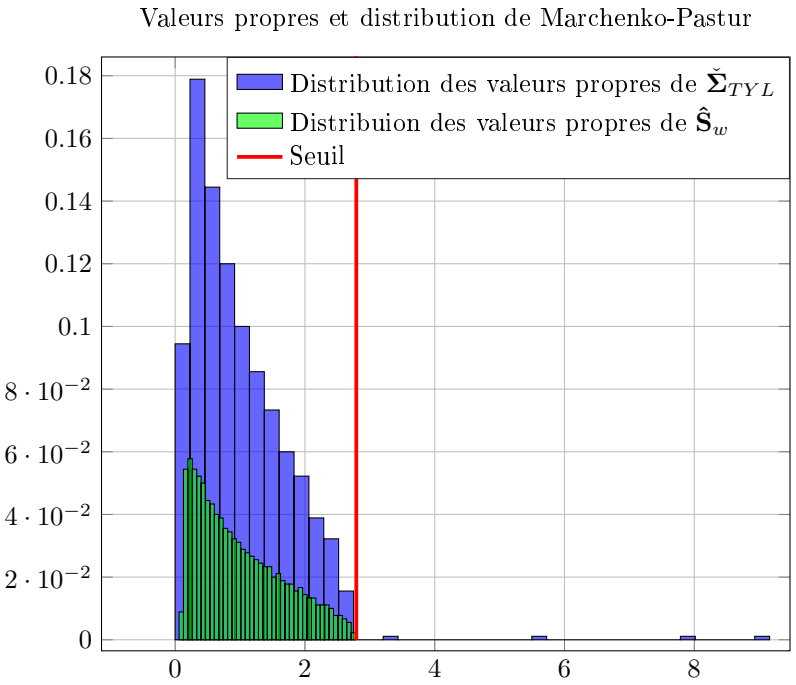

FIGURE 2.15 – Distribution de Marchenko-Pastur et distribution des valeurs propres de $\check{\mathbf{\Sigma}}_{TYL}$ pour 4 *endmembers* de RSB=30 dB, $N = 2000$, $m = 900$, $c = 0.7$, les $\{\tau\}_{i\in[\![1,N]\!]}$ suivant une loi t^2, t distribué selon Student-t.

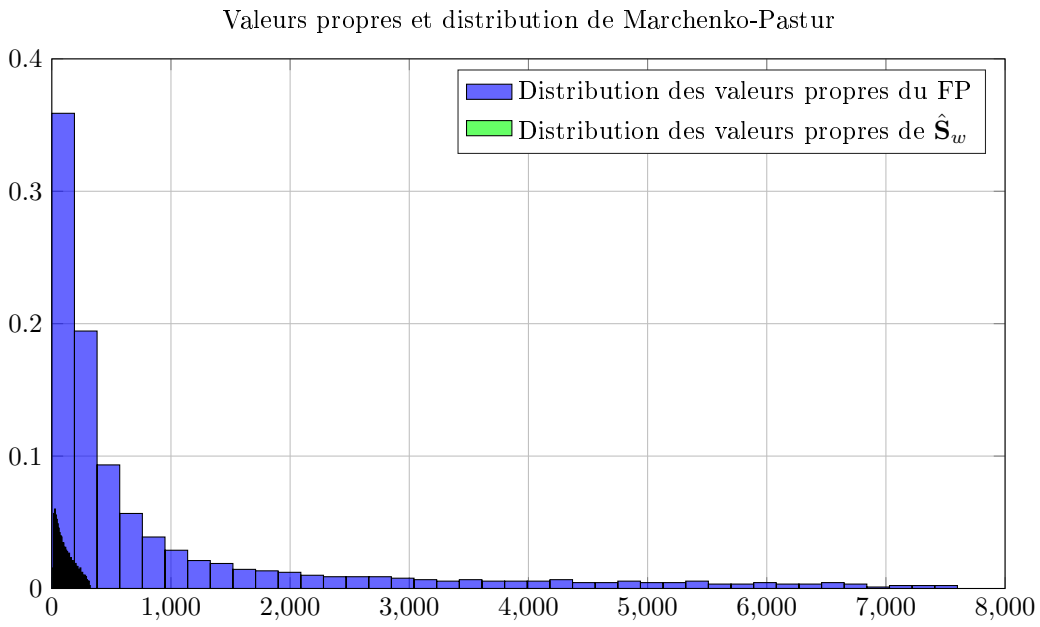

FIGURE 2.16 – Distribution de Marchenko-Pastur et distribution des valeurs propres de l'estimation de la matrice de covariance avec un blanchiment FP $u(x) = \frac{1+0.1}{x+0.1}$, pour 4 *endmembers* de $RSB = 10dB$, $N = 2000$, $m = 900$, $c = 0.7$, les $\{\tau\}_{i\in[\![1,N]\!]}$ suivant une loi inverse gamma.

car le seuil est directement calculable en fonction des dimensions de l'image et vaut $(1+\sqrt{c})^2$. Le seul obstacle à la détection est alors les RSB des sources, qui peut être trop faible et entraîner une non-détection. Quant au cas de fausse alarme, théoriquement impossible d'après [7], il ne peut être dû en pratique qu'aux erreurs de modèle : l'image ne correspondrait pas au modèle supposé ici et des valeurs propres du bruit dépasseraient du seuil.

2.5 Récapitulatif des algorithmes proposés pour l'estimation du nombre d'endmembers

Afin de donner une vue d'ensemble des différentes méthodes proposées, cette partie récapitule les différents algorithmes qu'il est possible de mettre en place dans les différents cas de figure du bruit. Il est rappelé au lecteur que tous les algorithmes sont applicables pour un cas gaussien blanc ou corrélé, le cas CES étant une généralisation de ce cas.

2.5.1 Bruit gaussien blanc

Algorithme 2.5.1 :

Acquisition des données de l'image	$\mathbf{Y}$
Calcul de la SCM	$\frac{1}{N}\mathbf{Y}\mathbf{Y}^T$ ou $\frac{1}{N}\mathbf{Y}\mathbf{Y}^H$ si le bruit simulé est complexe
Calcul des valeurs propres de la SCM	$\hat{\lambda}_0 \geq \ldots \geq \hat{\lambda}_{m-1}$
Calcul du seuil pour une PFA $= \alpha$ choisie	$k = 0$ $\hat{\sigma}^2(0) = \frac{1}{m-k-1}\sum_{k+1}^{m-1}\hat{\lambda}_j$ $\hat{\rho}_j^2 - \hat{\rho}_j\left(\hat{\lambda}_j + \hat{\sigma}^2(k) - \hat{\sigma}^2(k)\frac{m-k-1}{N}\right) + \hat{\lambda}_j\hat{\sigma}^2(k) = 0$ $\underline{\mathbf{A}} : \underline{k = k+1}$ $\hat{\sigma}(k)^2 - \frac{1}{m-k-1}\left(\sum_{j=k+1}^{m-1}\hat{\lambda}_j + \sum_{j=0}^{k}(\hat{\lambda}_j - \hat{\rho}_j)\right) = 0$ $\hat{\rho}_j^2 - \hat{\rho}_j\left(\hat{\lambda}_j + \hat{\sigma}^2(k) - \hat{\sigma}^2(k)\frac{m-k-1}{N}\right) + \hat{\lambda}_j\hat{\sigma}^2(k) = 0$ Tant que $\hat{\sigma}$ ne converge pas : $\Rightarrow \underline{\mathbf{A}}$ $\zeta_N = \hat{\sigma}^2(k)\left(b_N + \frac{\sigma_N}{m^{2/3}}\left(F_{TW}^{-1}(1-\alpha)\right)\right)$
Estimation du nombre d'*endmembers*	$\hat{p} = \arg\min_k\left(\hat{\lambda}_k < \zeta_N\right)$

2.5.2 Bruit corrélé i.i.d. de moyenne nulle

La seule méthode qui présente un intérêt pour l'imagerie hyperspectrale est celle de l'étude de l'espacement entre valeurs propres.

Algorithme 2.5.2 :

Acquisition des données de l'image	$\mathbf{Y}$
Calcul de la SCM	$\frac{1}{N}\mathbf{Y}\mathbf{Y}^T$ ou $\frac{1}{N}\mathbf{Y}\mathbf{Y}^H$ si bruit simulé complexe
Calcul des valeurs propres de la SCM	$\hat{\lambda}_0 \geq \ldots \geq \hat{\lambda}_{m-1}$ Et $\hat{\lambda}_{-1} = \infty$
Choix d'un seuil	ζ
Estimation du nombre d'*endmembers*	$\hat{p} = \arg\max_k\left(\frac{\hat{\lambda}_{k-1,N}}{\hat{\lambda}_{k,N}} > 1 + \zeta\right)$

2.5.3 Bruit CES blanc

Algorithme 2.5.3 :

Acquisition des données de l'image	$\mathbf{Y}$
Calcul de la SCM	$\frac{1}{N}\mathbf{Y}\mathbf{Y}^H$
Calcul des valeurs propres de la SCM	$\hat{\lambda}_0 \geq \ldots \geq \hat{\lambda}_{m-1}$
Calcul du seuil t	Choix de u $\Phi(x) = x\,u(x)$ et $v(x) = u \circ h^{-1}(x)$ $\psi(x) = x\,v(x)$ Calcul de ξ : $\frac{1}{N}\sum_{i=1}^{N} \frac{\psi(\tau_i\,\xi)}{1 + c_N\,\psi(\tau_i\,\xi)} = 1$ $t = \frac{\phi_\infty(1+\sqrt{c})^2}{\xi\,(1 - c\,\phi_\infty)}$
Estimation du nombre d'*endmembers*	$\hat{p} = \arg\min_k \left(\hat{\lambda}_k < t\right)$

2.5.4 Approche RMT non-robuste

Algorithme 2.5.4 :

Acquisition des données de l'image	$\mathbf{Y}$
Estimation de la matrice $\mathbf{C}$	$\hat{\mathbf{C}}_{SCM} \leftarrow \frac{1}{N}\mathbf{Y}\mathbf{Y}^H$ $\tilde{\mathbf{C}}_{SCM} \leftarrow \mathcal{T}\left(\hat{\mathbf{C}}_{SCM}\right)$ $\check{\mathbf{C}}_{SCM} \leftarrow \frac{1}{\mathbb{E}[\tau]}\tilde{\mathbf{C}}_{SCM}$
Blanchiment	$\check{\mathbf{Y}}_{wSCM} \leftarrow \check{\mathbf{C}}_{SCM}^{-1/2}\mathbf{Y}$
Estimation de la matrice de covariance	Choix d'une erreur ϵ (par exemple 10^{-3}) Choix d'une fonction u $\check{\mathbf{\Sigma}}_{SCM} \leftarrow \mathbf{I}_m$, erreur $= 1$ Tant que erreur $> \epsilon$ $\quad \check{\mathbf{\Sigma}}_2 \leftarrow \frac{1}{N}\sum_{i=1}^{N} u\left(\frac{1}{m}\check{\mathbf{y}}_{wSi}^H\,\check{\mathbf{\Sigma}}_{SCM}^{-1}\,\check{\mathbf{y}}_{wSi}\right)\check{\mathbf{y}}_{wSi}\check{\mathbf{y}}_{wSi}^H$ $\quad$ erreur $= \frac{\|\check{\mathbf{\Sigma}}_2 - \check{\mathbf{\Sigma}}_{SCM}\|_F}{\|\check{\mathbf{\Sigma}}_{SCM}\|_F}$ $\quad \check{\mathbf{\Sigma}}_{SCM} \leftarrow \check{\mathbf{\Sigma}}_2$
Calcul des valeurs propres de $\check{\mathbf{\Sigma}}_{SCM}$	$\check{\lambda}_0 \geq \ldots \geq \check{\lambda}_{m-1}$
Calcul du seuil	$t = \frac{\phi_\infty\,(1+\sqrt{c})^2}{\xi\,(1 - c\,\phi_\infty)}$
Estimation du nombre d'*endmembers*	$\hat{p} = \arg\min_k \left(\check{\lambda}_k < t\right)$

2.5.5 Approche RMT robuste cas général

Algorithme 2.5.5 :

Acquisition des données de l'image	$\mathbf{Y}$
Estimation de la matrice $\mathbf{C}$	$\hat{\mathbf{C}}_{FP} \leftarrow \mathbf{I}_m$ Choix d'un ϵ, choix d'une fonction u, calcul de v erreur $= 1$ Tant que erreur $> \epsilon$ $\hat{\mathbf{C}}_2 \leftarrow \frac{1}{N}\sum_{i=1}^{N} u\left(\frac{1}{m}\,\check{\mathbf{y}}_i^H\,\hat{\mathbf{C}}_{FP}^{-1}\,\check{\mathbf{y}}_i\right)\,\check{\mathbf{y}}_i\,\check{\mathbf{y}}_i^H$ erreur $= \frac{\|\hat{\mathbf{C}}_2 - \hat{\mathbf{C}}_{FP}\|_F}{\|\hat{\mathbf{C}}_{FP}\|_F}$ $\hat{\mathbf{C}}_{FP} \leftarrow \hat{\mathbf{C}}_2$ $\tilde{\mathbf{C}}_{FP} \leftarrow \mathcal{T}\left(\hat{\mathbf{C}}_{FP}\right)$ $\check{\mathbf{C}}_{FP} \leftarrow \frac{1}{\mathbb{E}[v(\tau\,\gamma)\,\tau]}\tilde{\mathbf{C}}_{FP}$
Blanchiment	$\check{\mathbf{Y}}_{wFP} \leftarrow \check{\mathbf{C}}_{FP}^{-1/2}\,\mathbf{Y}$
Estimation de la matrice de covariance	$\check{\mathbf{\Sigma}}_{FP} \leftarrow \mathbf{I}_m$ Choix d'un ϵ erreur $= 1$ Tant que erreur $> \epsilon$ $\check{\mathbf{\Sigma}}_2 \leftarrow \frac{1}{N}\sum_{i=1}^{N} u\left(\frac{1}{m}\,\check{\mathbf{y}}_{wFi}^H\,\check{\mathbf{\Sigma}}_{FP}^{-1}\,\check{\mathbf{y}}_{wFi}\right)\,\check{\mathbf{y}}_{wFi}\check{\mathbf{y}}_{wFi}^H$ erreur $= \frac{\|\check{\mathbf{\Sigma}}_2 - \check{\mathbf{\Sigma}}_{FP}\|_F}{\|\check{\mathbf{\Sigma}}_{FP}\|_F}$ $\check{\mathbf{\Sigma}}_2 \leftarrow \check{\mathbf{\Sigma}}_{FP}$
Calcul des valeurs propres de $\check{\mathbf{\Sigma}}_{FP}$	$\check{\lambda}_0 \geq \ldots \geq \check{\lambda}_{m-1}$
Calcul du seuil	$t = \frac{\phi_\infty(1+\sqrt{c})^2}{\xi(1-c\phi_\infty)}$
Estimation du nombre d'*endmembers*	$\hat{p} = \arg\min_k \left(\check{\lambda}_k < t\right)$

2.5.6 Approche robuste RMT : cas de l'estimation de Tyler

Algorithme 2.5.6 :

Acquisition des données de l'image	$\mathbf{Y}$
Estimation de la matrice $\mathbf{C}$	$\hat{\mathbf{C}}_{TYL} \leftarrow \mathbf{I}_m$ Choix d'un ϵ erreur $= 1$ Tant que erreur $> \epsilon$ $\quad \hat{\mathbf{C}}_2 \leftarrow \frac{m}{N} \sum_{i=1}^{N} \frac{\check{\mathbf{y}}_i \check{\mathbf{y}}_i^H}{\check{\mathbf{y}}_i^H \hat{\mathbf{C}}_{TYL}^{-1} \check{\mathbf{y}}_i}$ $\quad \text{erreur} = \frac{\left\| \hat{\mathbf{C}}_2 - \hat{\mathbf{C}}_{TYL} \right\|_F}{\|\hat{\mathbf{C}}_{TYL}\|_F}$ $\quad \hat{\mathbf{C}}_{TYL} \leftarrow \hat{\mathbf{C}}_2$ $\check{\mathbf{C}}_{TYL} \leftarrow \mathcal{T}\left(\hat{\mathbf{C}}_{TYL}\right)$
Blanchiment	$\check{\mathbf{Y}}_{wTYL} \leftarrow \check{\mathbf{C}}_{TYL}^{-1/2} \mathbf{Y}$
Estimation de la matrice de covariance	$\check{\mathbf{\Sigma}}_{TYL} \leftarrow \mathbf{I}_m$ Choix d'un ϵ erreur $= 1$ Tant que erreur $> \epsilon$ $\quad \check{\mathbf{\Sigma}}_2 \leftarrow \frac{m}{N} \sum_{i=1}^{N} \frac{\check{\mathbf{y}}_{wTi} \check{\mathbf{y}}_{wTi}^H}{\check{\mathbf{y}}_{wTi}^H \check{\mathbf{\Sigma}}_{TYL}^{-1} \check{\mathbf{y}}_{wTi}}$ $\quad \text{erreur} = \frac{\|\check{\mathbf{\Sigma}}_2 - \check{\mathbf{\Sigma}}_{FP}\|_F}{\|\check{\mathbf{\Sigma}}_{TYL}\|_F}$ $\quad \check{\mathbf{\Sigma}}_2 \leftarrow \check{\mathbf{\Sigma}}_{TYL}$
Calcul des valeurs propres de $\check{\mathbf{\Sigma}}_{TYL}$	$\check{\lambda}_0 \geq \ldots \geq \check{\lambda}_{m-1}$
Calcul du seuil	$t = (1 + \sqrt{c})^2$
Estimation du nombre d'*endmembers*	$\hat{p} = \arg\min_k \left(\check{\lambda}_k < t\right)$

2.6 Synthèse

Ce chapitre a présenté les résultats théoriques concernant l'estimation du nombre d'*endmembers*. Les premiers résultats présentés sur les modèles simplifiés sont des résultats déjà publiés, à part pour les simulations et figures tracées dans ce chapitre. Ils ont été ici adaptés au cas de l'imagerie hyperspectrale et permettent de caractériser tout un panel d'algorithmes et de méthodes potentiellement intéressantes pour l'imagerie hyperspectrale.

En ce qui concerne les derniers résultats présentés et prouvés en Annexe, ils donnent la possibilité d'utiliser des résultats de la théorie des matrices aléatoires pour permettre l'estimation du nombre d'*endmembers*. Ces résultats ont fait l'objet de publications [108], [109], et [110]. Ils montrent qu'il est possible, malgré le peu d'a priori sur les images, d'estimer correctement le nombre d'*endmembers* en utilisant des estimées de matrices de covariance.

Chapitre 3

Applications

Les applications présentées dans ce chapitre sont des applications des résultats théoriques du chapitre précédent. Une première partie présente les résultats sur des images hyperspectrales réelles, simulées ou sur des images hyperspectrales simulées avec de vrais spectres hyperspectraux. Une seconde partie présente une adaptation des résultats précédents sur des données du domaine de la finance. Cette partie permet de montrer que tous les résultats théoriques développés dans cette thèse et adaptés jusqu'ici au cas de l'imagerie hyperspectrale peuvent être adaptés à d'autres applications.

3.1 Imagerie hyperspectrale

Dans cette partie sont appliqués les résultats théoriques du chapitre précédent au cas d'images hyperspectrales réelles, simulées et au cas d'images hyperspectrales construites à partir de spectres réels obtenus sur des bases de données diverses. Certains résultats classiques ne peuvent être adaptés au cas de l'imagerie hyperspectrale sous les hypothèses réunies précédemment, notamment les méthodes nécessitant l'accès à des données secondaires ou à des données non-corrélées. Ces méthodes sont donc définitivement écartées et ne feront pas l'objet de simulations dans ce chapitre.

3.1.1 Approche RMT non CES

Avant de parler du cas général de cette thèse, qui considère une image hyperspectrale sans aucun a priori, est considéré ici le cas de la détection de cible ou de la détection de changement, un problème fréquent en imagerie hyperspectrale. La particularité de ce problème par rapport au dénombrement du nombre d'*endmembers* est que l'image avant le changement ou avant la présence de cible est connue. Ainsi, une image hyperspectrale constitué d'information non pertinente est supposée connue. Cette image peut être associée à une image de bruit ou de données secondaires. Sur une image connue, c'est en dire en ayant quelques a priori sur les données tels que la connaissance de la nature du bruit, il est possible d'adapter le Théorème 2.3.3 au cas de l'imagerie hyperspectrale. Il est en effet dans ce cas possible de choisir un seuil ζ (2.8), en fonction du bruit de l'image. En ce qui concerne la détection de cible ou de changement, l'image avant changement ou présence de la cible peut alors être considérée comme une image de bruit, car ne contient pas l'information recherchée, c'est à dire ne contient pas les *endmembers* recherchés. Il devient alors intéressant de montrer que des méthodes développées en théorie des matrices aléatoires peuvent s'appliquer sans grand changement au cas de l'imagerie hyperspectrale dans le cas où une image hyperspectrale sans la présence des *endmembers* recherchés est connue.

L'exemple choisi est celui d'une image hyperspectrale contenant de la végétation, une

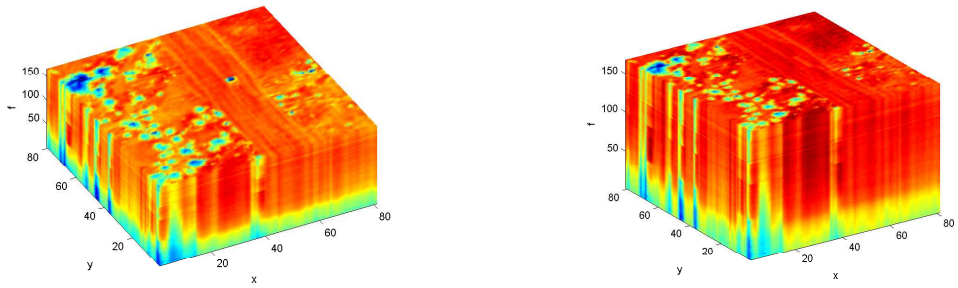

FIGURE 3.1 – Image hyperspectrale utilisée, avec et sans la voiture, de taille $m = 167$, $N = 6561$.

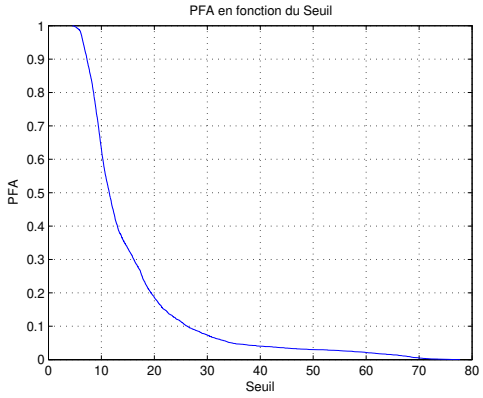

FIGURE 3.2 – Courbe de la PFA en fonction du seuil, $m = 167$, $N = 21 \times 21$.

route et une voiture. Sur cette image, la position de la voiture étant connue, il est possible de la supprimer et de la remplacer par la route afin de disposer de la même image avec et sans la voiture. L'idée est donc non pas de détecter le nombre d'*endmembers* mais le nombre d'*endmembers* différents du fond connu de l'image, soit ici, pour le cas présenté, de détecter la voiture. Ce genre de technique peut être utilisé notamment pour des problèmes de détection de changement entre deux images hyperspectrales, comme déjà mentionné plus haut. Pour appliquer la méthode décrite par le Théorème 2.3.3, il est nécessaire de choisir un seuil $\zeta + 1$. Il faudrait disposer de plusieurs images du bruit ou du fond de la même zone pour estimer statistiquement les seuils associés à différentes PFA. Ceci permettrait de choisir le seuil de détection par rapport au fond de l'image considérée.

Comme l'image est assez homogène, il est possible de la découper en plusieurs images de tailles réduites, ici des images de taille 21×21 pixels, et de considérer ces images comme des réalisations statistiques différentes du même fond. De cette façon, l'ensemble des valeurs propres de la SCM obtenus pour chaque image permet de calculer la PFA pour chaque seuil choisi. La figure 3.1 à gauche, montre l'image analysée, sur laquelle il est possible de distinguer une route sur laquelle se trouve une voiture. A droite, se trouve l'image sans la voiture, de taille 81×81 qui est ensuite divisée en plusieurs images de taille 21×21 à l'aide d'une fenêtre glissante, pour donner la courbe PFA-seuil. La courbe PFA-seuil obtenue est tracée sur la figure 3.2. Grâce à cette figure, il est possible de choisir un seuil adapté à cette image pour une PFA choisie.

TABLE 3.1 – Rapports des valeurs propres successives obtenus pour l'image de taille $81 \times 81 \times 167$

Rapports de valeurs propres						
k	1	2	3	4	5	6
avec voiture	37	3.4	4.3	3.6	1.1	2.0
sans voiture	16	3.1	4.3	3.3	1.3	2.0

L'image originale est ensuite testée, en utilisant l'équation (2.8), pour une PFA choisie de 0.05. Le seuil est alors de 34.7. Le tableau 3.1 présente les différents rapports des valeurs propres de l'algorithme 2.5.2, c'est à dire les valeurs de $\frac{\hat{\lambda}_{k-1,N}}{\hat{\lambda}_{k,N}}$ pour les premières valeurs de $k > 1$.

Sur ce tableau, le premier rapport de la première ligne est le seul à dépasser le seuil, (en plus de celui pour lequel $k = 0$). Le nombre d'*endmembers* estimé est alors $\hat{p} = 1$. De plus, sur la seconde ligne, le seul rapport à dépasser le seuil est pour $k = 0$ ce qui donne bien $\hat{p} = 0$ pour l'image sans la voiture.

L'estimation du nombre d'*endmembers* peut donc se faire correctement ici mais à condition d'avoir certaines connaissances sur les données, afin de pouvoir fixer un seuil.

3.1.2 Comparaison des méthodes adaptés au bruit CES

Dans le cas le plus probable où les images hyperspectrales étudiées ne sont pas connues a priori, les algorithmes 2.5.4, 2.5.5 et 2.5.6 permettent d'estimer le nombre d'*endmembers*. Ces algorithmes ont été testés avec deux types d'images : des images constituées d'un bruit CES simulé sur lesquelles s'ajoutent des spectres simulés, des images constituées d'un bruit CES simulé sur lesquelles s'ajoutent des spectres réels issus de la base de données [60], et des images hyperspectrales réelles issues de la base de données [35] ainsi que des images issues de DSO National Laboratories.

Comparaison sur bruit simulé

Spectres Simulés

Les trois algorithmes proposés ont été appliqués sur des images simulées. La première série de figures compare les méthodes fondées sur une estimation robuste dans le cas général (voir l'algorithme 2.5.5) et celles dans le cas non robuste (voir l'algorithme 2.5.4). À ces courbes est ajoutée, à titre de comparaison, la courbe obtenue pour le même algorithme robuste dans le cas général mais sans l'étape de blanchiment $\check{\mathbf{C}}^{-1/2}\mathbf{Y}$, notée "$\hat{p}$ avec $\check{\mathbf{\Sigma}}_{NB}$". La fonction u choisie est $u : x \mapsto \frac{1+\alpha}{x+\alpha}$, la matrice $\mathbf{C}$ est générée de la même façon que dans la partie précédente, avec $\mathbf{C} = \mathcal{L}\left(\left(\rho^0, \rho^1, \ldots, \rho^{m-1}\right)^T\right)$. Les spectres sont aussi simulés de la même façon que dans la partie précédente, avec, pour $j \in [\![1, m]\!]$: $[\mathbf{m}]_{j,k} = 10^{\mathrm{RSB}/20} \exp\left(2i\pi \frac{(j-1)(k-1)}{m}\right)$, pour $k \in [\![1, p]\!]$. Chaque simulation est le résultat de la moyenne sur 4 tirages de bruit indépendants. Les différents paramètres choisis sont les suivants :

- sur la figure 3.4 : les variables $\{\tau_i\}_{i \in [\![1,N]\!]}$ sont indépendantes, identiquement distribuées selon une loi inverse gamma, de moyenne 1 et de paramètre 10. De plus, le paramètre α a été fixé d'une manière subjective à 0.7. Enfin, $\rho = 0.7$.

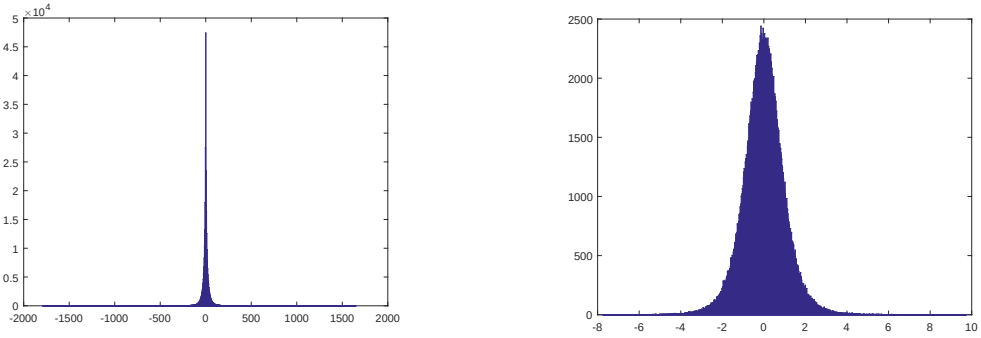

FIGURE 3.3 – Différentes distributions de bruit, à gauche lorsque les $\{\tau_i\}_{i\in[\![1,N]\!]}$ suivent une loi du t^2 (carré d'une loi de Student) et à droite une loi inverse gamma.

- sur la figure 3.5 : les $\{\tau_i\}_{i\in[\![1,N]\!]}$ sont indépendantes, identiquement distribuées et suivent la même loi que t^2 pour $t \sim$ Student-t, de paramètre 100. Le paramètre α a été fixé également d'une manière subjective à 0.1. Et de même, $\rho = 0.7$.

Les données générées avec les variables aléatoires $\{\tau_i\}_{i\in[\![1,N]\!]}$ qui suivent une loi de Student-t au carré sont fortement non gaussiennes, comme illustré par la figure 3.3, où sont comparées les distributions de bruit utilisées pour la figure 3.6 et celle pour la figure 3.7.

Sur la figure 3.4, les résultats sont corrects à partir d'un RSB de 0 dB, c'est à dire qu'ils sont acceptables à partir d'un RSB assez faible, tandis que sur la figure 3.5, les RSB doivent être plus importants avant de pouvoir être détectés. Dans tous les cas, la méthode AIC, comme prévu, donne de mauvais résultats pour ce genre de données. Quant à l'étape de blanchiment, elle est bien indispensable à l'algorithme puisque les résultats théoriques sont faux si les données n'ont pas été blanchies. C'est d'ailleurs ce qui est visible sur les figures 3.4 et 3.5, où les résultats obtenus pour ces méthodes ne sont même plus pertinents.

Sur les figures 3.6 et 3.7 sont comparés les résultats entre l'algorithme robuste général et l'algorithme robuste de Tyler. Les paramètres m et N ne sont pas tout à fait les mêmes, ce qui permet de vérifier que les résultats ne sont pas dépendants de ces paramètres. Les autres paramètres sont les mêmes que pour les deux figures décrites précédemment, c'est à dire :

- sur la figure 3.6 : les variables aléatoires $\{\tau_i\}_{i\in[\![1,N]\!]}$ sont indépendants, identiquement distribués selon une loi inverse gamma, de moyenne 1 et de paramètre 10. Le paramètre α a été également fixé d'une manière subjective à 0.1. Enfin, $\rho = 0.7$.
- sur la figure 3.7 : les variables aléatoires $\{\tau_i\}_{i\in[\![1,N]\!]}$ sont indépendants, identiquement distribués et suivent la même loi que t^2 pour $t \sim$ Student-t, de paramètre 100. Le paramètre α a, encore une fois, été fixé subjectivement à 0.1. Et de même, $\rho = 0.7$.

La méthode reposant sur l'estimateur de Tyler donne des résultats beaucoup plus robustes que la méthode utilisant une fonction u de type $u : x \mapsto \dfrac{1+\alpha}{x+\alpha}$. Ceci s'explique en effet par l'insensibilité de l'estimateur de Tyler aux facteurs multiplicatifs τ mais aussi par le fait que le seuil n'est pas, pour cette dernière méthode, dépendant de la fonction u choisie. Ces simulations sont donc en faveur du dernier algorithme présenté avec l'algorithme 2.5.6.

Spectres Réels

Les algorithmes précédents ont été utilisés avec des spectres réels, issus de la base de données [60]. Deux spectres ont été choisis et sont représentés sur la figure 3.8. Ces spectres sont

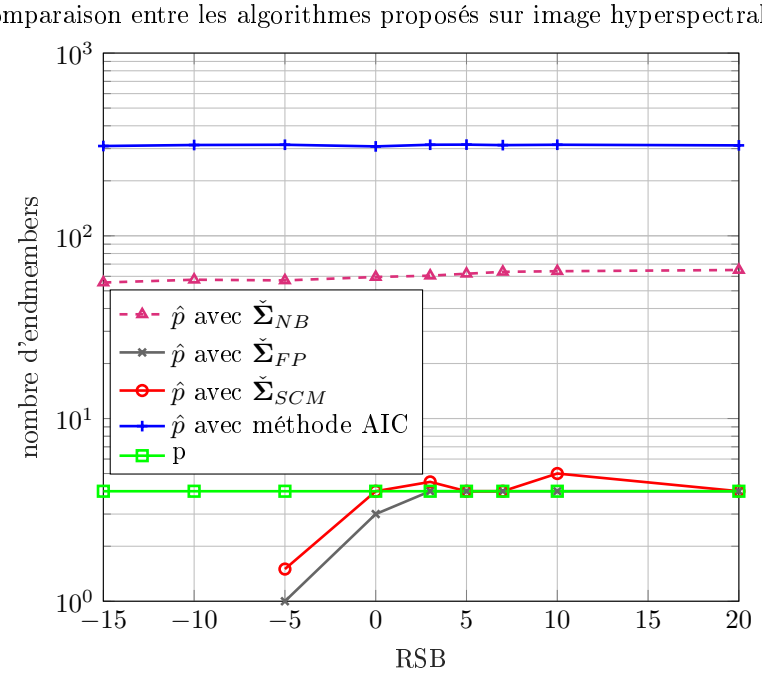

FIGURE 3.4 – Estimation du nombre d'*endmembers* $\hat{p}$ pour un bruit CES corrélé, $m = 400$, $c = 0.2$, $p = 4$, $\rho = 0.7$, $\{\tau\} \sim$ inverse gamma, en fonction du RSB.

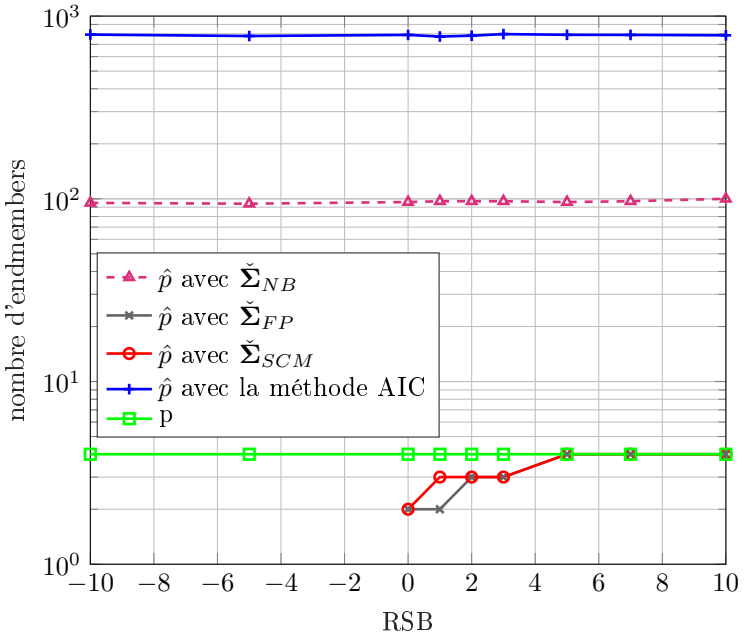

FIGURE 3.5 – Estimation du nombre d'*endmembers* $\hat{p}$ pour un bruit CES corrélé, $m = 400$, $c = 0.2$, $p = 4$, $\rho = 0.7$, $\{\tau\} \sim t^2$ avec $t \sim$ Student-t , en fonction du RSB.

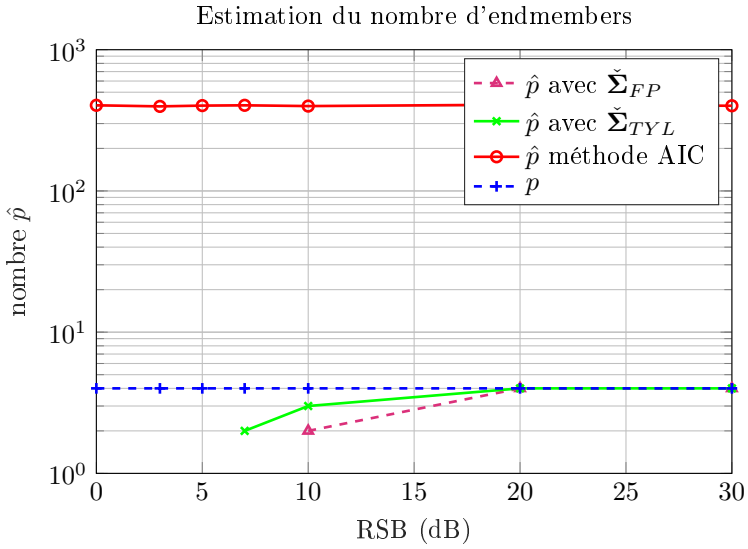

FIGURE 3.6 – Estimation du nombre d'*endmembers* $\hat{p}$ pour un bruit CES corrélé, $m = 900$, $c = 0.45$, $p = 4$,$\rho = 0.7$, $\{\tau\} \sim$ inverse gamma, en fonction du RSB.

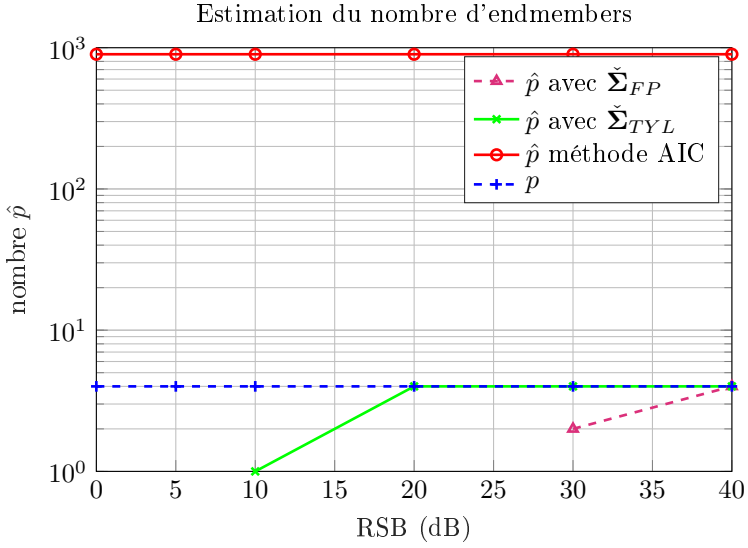

FIGURE 3.7 – Estimation du nombre d'*endmembers* $\hat{p}$ pour un bruit CES corrélé, $m = 900$, $c = 0.45$, $p = 4$,$\rho = 0.7$, $\{\tau\} \sim t^2$ avec $t \sim Student - t$, en fonction du RSB.

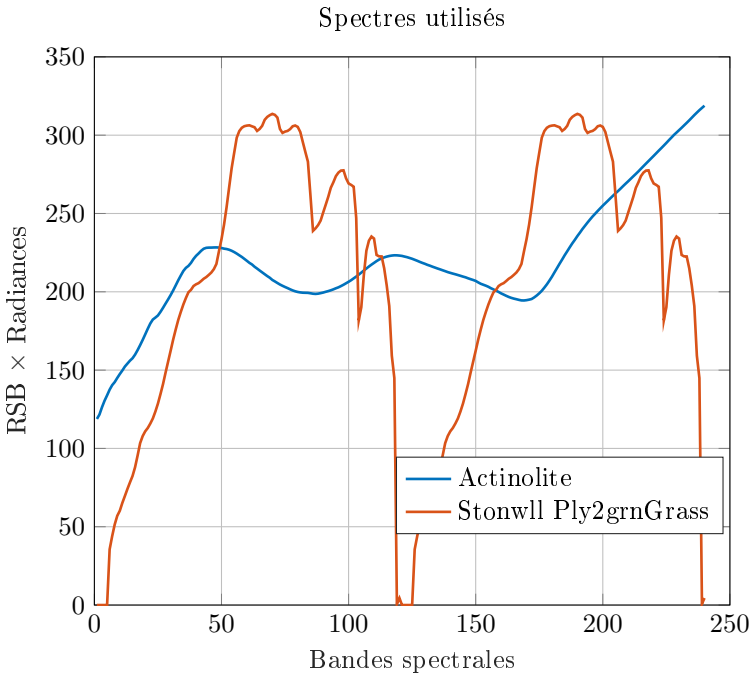

FIGURE 3.8 – Spectres réels utilisés pour la figure 3.9.

ceux de l'`Actinolite` et de `Stonwll_Ply2grnGrass` . Le spectre de `Stonwll_Ply2grnGrass` est répété deux fois pour des raisons de dimension des données, les autres spectres s'étendant sur deux fois plus de bandes spectrales.

À ces spectres est ajouté un bruit CES dont les variables $\{\tau_i\}_{i\in[\![1,N]\!]}$ suivent une loi inverse gamma de paramètre 2. Le paramètre ρ tel que $\mathbf{C} = \mathcal{L}\left(\left(\rho^0, \rho^1, \ldots, \rho^{m-1}\right)^T\right)$ vaut $\rho = 0.7$, et le paramètre α pour la fonction $u : x \mapsto \dfrac{1+\alpha}{x+\alpha}$ est fixé à $\alpha = 0.1$. Les paramètres m et N valent $m = 240$ et $N = 1000$. Sur la figure 3.9 sont tracés différents résultats pour les algorithmes présentés dans le chapitre précédent. La méthode $\check{\mathbf{\Sigma}}_{FP}$ correspond à l'algorithme 2.5.5, la méthode $\check{\mathbf{\Sigma}}_{SCM}$ à l'algorithme 2.5.4, celui $\check{\mathbf{\Sigma}}_{TYL}$ à 2.5.6 et pour finir la méthode $\check{\mathbf{\Sigma}}_{TYL}$ correspond aussi à l'algorithme 2.5.6 mais pour un bruit blanc, ce qui correspond au cas idéal où le bruit CES serait égal à $\mathbf{X}\,\mathbf{T}^{1/2}$. Le RSB correspond comme précédemment à $20\log_{10} P$ avec P la puissance des sources, c'est à dire le facteur multiplicatif du spectre.

Sur la figure 3.9, la méthode qui fonctionne le mieux est encore la méthode robuste utilisant un estimateur de Tyler, comme proposé par l'algorithme 2.5.6. Elle commence à détecter un *endmember* pour un RSB de 0 dB, et le bon nombre d'*endmembers* à partir de 40 dB. Les méthodes robustes 2.5.5 et non-robustes 2.5.4 sous-estiment le nombre d'*endmembers* tandis que la méthode AIC le surestime. L'idéal serait de trouver une méthode qui se rapproche le plus que possible du cas du bruit blanc. la méthode qui s'en approche le plus parmi celles proposées est donc la méthode robuste utilisant un estimateur de Tyler. En ce qui concerne le temps de calcul, pour des images de cette dimension, il est de quelques secondes.

Comparaison sur images réelles

Les images réelles utilisées sont des images hyperspectrales de types assez variés. L'image Indian Pines est une image contenant entre 10 et 16 *endmembers* différents, dépendant du

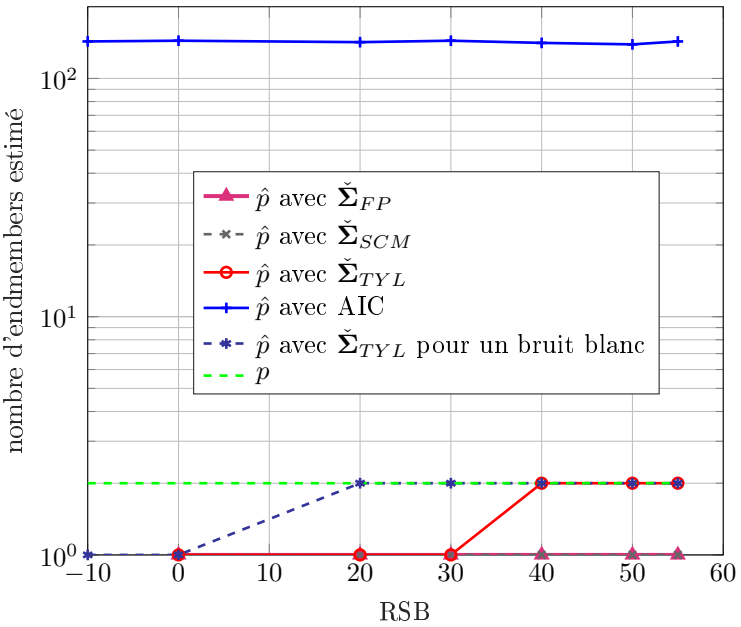

FIGURE 3.9 – Estimation du nombre d'*endmembers* $\hat{p}$ pour un bruit CES corrélé, $m = 240$, $c = 0.24$, $p = 2$,$\rho = 0.7$, $\{\tau\} \sim$ inverse gamma , en fonction du RSB.

TABLE 3.2 – Estimation du nombre d'*endmembers* sur des images hyperspectrales

Images	Indian Pines	SalinasA	PaviaU	Cube
p	16	9	9	6
$\hat{p}_{AIC}$	219	203	102	143
$\hat{p}_{Hysime}$	**19**	14	60	19
$\hat{p}_{FP}$	11	**9**	1	**3**
$\hat{p}_{TYL}$	**13**	2	**10**	**13**

niveau de précision de l'imageur hyperspectral, chaque *endmember* représentant des types de cultures différentes, répartis par zone sur l'image de gauche de 3.10. L'image de droite 3.10 sur montre une partie de l'Université Pavia, contenant théoriquement 9 *endmembers* différents, correspondant à différents types de matériaux. Ces images sont issues de la base de données [35]. L'image 3.11, montre deux dernières scènes étudiées, celle de gauche la même scène que précédemment, c'est à dire une route avec des voitures et de la végétation, contenant 6 *endmembers* principaux, et celle de droite, tirée de [35], correspondant à une partie de la scène Salinas-A.

Ces images sont de types très différents et représentent un bon échantillon d'images hyperspectrales prises en vue aérienne pour tester les algorithmes proposés. Les algorithmes testés sont les algorithmes robustes proposés 2.5.5, avec l'estimation de p notée $\hat{p}_{FP}$ et 2.5.6, dont l'estimation de p sera notée $\hat{p}_{TYL}$. Pour comparer les résultats obtenus sur ces images réelles, ils sont comparés ici à la méthode AIC par $\hat{p}_{AIC}$, mais aussi à la méthode Hysime avec $\hat{p}_{Hysime}$, méthode proposée dans [11] et décrite dans l'introduction. Les résultats obtenus sur les quatre images présentées sont résumés dans le tableau 3.2, sachant que p représente une vérité terrain (pour rappel, la vérité terrain reste une notion assez subjective).

Les résultats en rouge sont les résultats qui se rapprochent le plus de la vérité terrain. La méthode AIC, comme prévu, ne donne pas de bons résultats. La méthode Hysime donne

FIGURE 3.10 – Une bande spectrale des images Indian Pines (à gauche) et PaviaU (à droite)

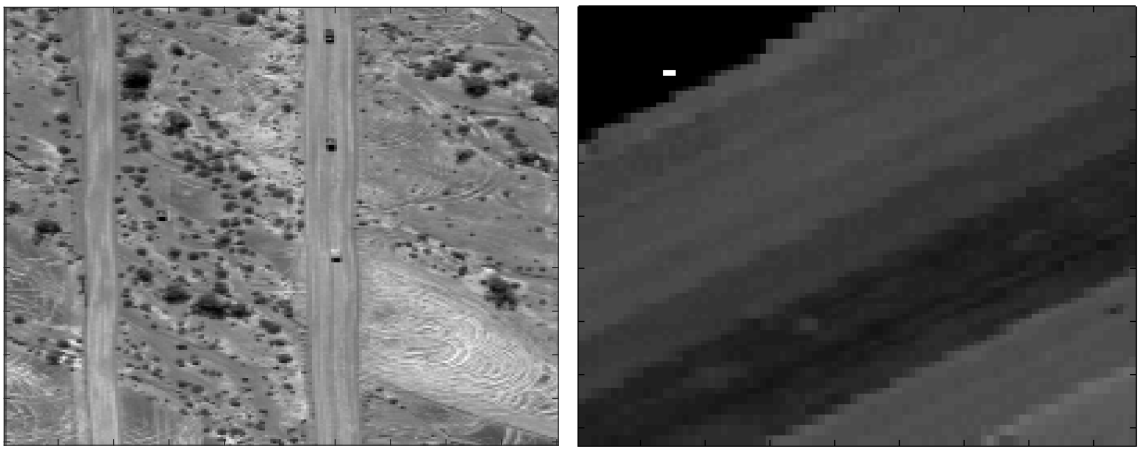

FIGURE 3.11 – Une bande spectrale des images "Cube" (à gauche) et SalinasA (à droite)

de bons résultats pour une image mais surestime souvent le nombre d'*endmembers*, tandis que la méthode robuste 2.5.5 a tendance à sous-estimer ce nombre. La méthode robuste cas de Tyler, quant à elle, semble plutôt stable et donne de relativement bons résultats.

L'hypothèse qui a été faite sur la non-gaussianité du bruit semble réaliste, puisque les résultats obtenus ne sont pas moins bons que ceux obtenus par des méthodes plus "classiques". Les images utilisées étant de types assez variés, les résultats sont plutôt concluants, et sont un bon indicateur de la pertinence du modèle et des méthodes proposées. Sur des images sans aucun a priori, il est ainsi possible de retrouver le nombre d'*endmembers*, ou, en tout cas, un ordre d'idée.

Le principal problème est bien entendu de déterminer des critères qui permettraient d'évaluer la pertinence du nombre d'*endmembers* trouvés. Á part comparer plusieurs méthodes ou appliquer une technique de démélange pour vérifier la cohérence de la répartition des *endmembers* sur l'image, comme il n'y a aucune information sur l'image, il ne semble pas y avoir de possibilités de contrôle des résultats.

Les applications qui recherchent un changement dans l'image reposent sur la connaissance de l'image avant ce changement, comme présenté ci-avant. Ce type d'application donne des résultats qu'il est possible d'imaginer plus vérifiables et plus fiables, puisqu'une grossière erreur de calcul sur le nombre d'*endmembers* estimé donnera un résultat qui ne semblera pas cohérent avec les images avant et après le changement.

3.2 Allocation de portefeuilles en finance

Les méthodes proposées sont des méthodes issues de résultats théoriques généraux, et même si cette thèse s'occupe avant tout de l'imagerie hyperspectrale, il semble intéressant d'appliquer les résultats obtenus sur des données d'un tout autre genre, et d'en démontrer l'utilité. Cela pourrait s'appliquer au problème de la détection et l'estimation en radar, du traitement d'antennes et de la localisation de sources, à tous les problèmes d'estimation d'ordre de modèle, et à d'autres applications de ce type ou utilisant des données de grande dimension.

Les résultats théorique précédents sont ainsi appliqués à l'allocation de portefeuille en finance et notamment à un cas pour lequel la fonction d'optimisation ne dépend que de la matrice de covariance des rendements des actifs financiers, à savoir l'allocation en variété maximale.

Cette partie de la thèse a donc pour vocation de prouver l'adaptabilité des résultats théoriques du second chapitre à d'autres types de données et en particulier à l'allocation de portefeuille.

3.2.1 Petite introduction du problème

L'allocation de portefeuilles est un problème central en finance. A partir d'un ensemble d'actifs (ou titres) financiers, l'objectif est de calculer les proportions à allouer sur chacun de ces actifs pour maximiser un gain ou rendement en fonction de risques choisis. Il s'agit d'un problème d'optimisation qui dépend d'une fonction choisie, fonction qui permet de déterminer le poids à donner à chaque actif pour maximiser les rendements. La performance d'un actif financier est d'ailleurs défini comme étant son rendement (défini mathématiquement ci-après) sur une période de temps fixe. Les méthodologies sont nombreuses, allant d'une allocation équi-pondérée à des allocations résultant de la maximisation (ou minimisation) de critères de risque ou de performance pouvant être assez complexes.

Un rendement, noté p_t, est le taux d'évolution d'un prix entre les instants $t-1$ et t, c'est à dire qu'un rendement $r(t)$ est défini par :

$$r(t) = \frac{p(t) - p(t-1)}{p(t-1)}\,. \tag{3.1}$$

Pour un portefeuille P d'actifs constitué de N actifs et dont les poids sont notés $w_i(t)$, $i \in [\![1, N]\!]$ pour une date t fixée, le rendement $r_P(t)$ s'écrit :

$$r_P(t) = \sum_{i=1}^{N} w_i(t)\, r_i(t)\,. \tag{3.2}$$

Le risque, quant à lui, est souvent mesuré par l'écart type des rendements des portefeuilles, appelé volatilité dans le domaine financier. Soit $\mathbf{r}_P = [r_P(1), \ldots, r_P(T)]^T$ le vecteur pour tout temps $t \in [\![1, T]\!]$ des rendements de P, alors ce vecteur est tel que $\mathbf{r}_P = \mathbf{R}\mathbf{w}$, $\mathbf{R}$ la matrice de taille $T \times N$ des N actifs observés sur la longueur de temps T et $\mathbf{w}$ les vecteurs des N poids supposés constants sur la période T . La volatilité des rendements de P s'écrit alors :

$$V_P = \sqrt{\mathbf{w}^H \mathbf{\Sigma} \mathbf{w}}\,, \tag{3.3}$$

où $\mathbf{\Sigma}$ est la matrice de covariance de taille $N \times N$ des rendements des actifs, c'est à dire que $\mathbf{\Sigma} = \mathbb{E}[\mathbf{R}^H \mathbf{R}]$.

Un des premiers développements d'optimisation de portefeuille a été proposée par H. Markowitz [73] dans les années 1950. La solution envisagée, connue sous le nom d'allocation *mean-variance*, peut se résumer comme suit : pour un rendement de portefeuille attendu (ou pour un niveau de volatilité fixé), l'objectif est de minimiser la variance du portefeuille (ou respectivement de maximiser son rendement). Il est donc nécessaire de déterminer les rendements des actifs pour déterminer le rendement attendu du portefeuille, ainsi que d'estimer la matrice de covariance des rendements des actifs. De nombreuses études montrent que l'estimation des rendements attendus conduisent à des erreurs importantes d'estimation, ce qui n'est pas forcément le cas pour l'estimation de la matrice de covariance. C'est pourquoi les méthodes telles que le Global Minimum Variance (GMV), *Equally-Risk Contribution* (ERC) ou encore la variété maximale (VarMax ou MDP pour *Most Diversified Portfolio*) sont également très étudiées.

Il est possible de choisir différents types de poids, en fonction de ce qui est recherché. Pour cette thèse, c'est la maximisation du ratio de diversification tel qu'utilisé et défini chez Fideas Capital qui est étudié. Ce ratio de diversification est la fonction f qui permet d'assurer une certaine diversité des actions. Elle se définit comme suit :

$$f(\mathbf{w}, \mathbf{\Sigma}) = \max_{\mathbf{w}} \frac{\sum_{i=1}^{N} w_i(t)\, r_i(t)}{\sqrt{\mathbf{w}^T\, \mathbf{\Sigma}\, \mathbf{w}}}\,. \tag{3.4}$$

C'est cette fonction qui sera choisie pour la suite des applications. Il est donc nécessaire d'estimer la matrice de covariance des rendements $\mathbf{\Sigma}$ pour optimiser le problème en question, car cette matrice permet de donner l'erreur d'estimation nécessaire à la fonction d'optimisation.

Les solutions proposées dans cette partie ont été développées grâce à une des méthodes précédemment exposée et ont été mises en place par E. Jay [53]. Les résultats qui suivent sont donc de sa contribution.

Pour estimer les poids à attribuer aux différentes actions, il est ainsi nécessaire de mettre en place une stratégie pour estimer les différents autres paramètres, dont la matrice de covariance $\mathbf{\Sigma}$. Cependant, la SCM, souvent utilisée, est optimale dans le cas gaussien, ce qui n'est pas forcément le cas pour les séries financières (une série financière représente l'évolution des prix d'une action), qui peuvent présenter des valeurs qui ont des queues de distribution plus larges que la loi gaussienne. De plus, le nombre d'échantillons N disponibles peut aussi être du même ordre de grandeur que la dimension des données, ce qui, comme présenté dans l'introduction, fait de la SCM un estimateur non-consistant pour la matrice de covariance. L'idée proposée par E. Jay est donc d'utiliser un des estimateurs proposés dans le chapitre précédent pour mieux estimer la matrice de covariance des rendements.

Soit la matrice $\mathbf{R} = [\mathbf{r}_1, \ldots, \mathbf{r}_N]$ la matrice contenant les rendements considérés comme les observations, de taille $m \times N$. Parmi ces observations, il existe K facteurs principaux générant l'espace des observations. Le modèle suivi par les observations est le même que celui des images hyperspectrales, c'est à dire :

$$\mathbf{r}_t = \mathbf{B}_t \, \mathbf{f}_t + \sqrt{\tau_t} \, \mathbf{C}^{1/2} \, \mathbf{x}_t \text{ pour tout } t \in [\![1, N]\!], \tag{3.5}$$

avec $\mathbf{B}_t$ une matrice de taille $m \times K$ contenant les coefficients qui définissent la sensibilité des actions à chaque facteur à un instant t, $\mathbf{f}_t$, vecteur de taille k, contenant les valeurs des facteurs à un instant t et supposé commun à toutes les actions. $\mathbf{x}_t$ est un vecteur aléatoire de taille m uniformément distribué sur une sphère de rayon un, $\mathbf{C}$ la matrice de dispersion du bruit, de taille $m \times m$, supposée de Toeplitz, et enfin τ_t des variables aléatoires positives représentant, à l'instant t, la variance du bruit.

Le problème peut donc se résumer à un problème d'estimation de la matrice $\mathbf{C}$, afin d'optimiser correctement la fonction f choisie.

3.2.2 Méthode suivie

L'estimateur choisi pour la matrice $\mathbf{C}$ est le M-estimateur robuste utilisant la méthode de Tyler (2.37), de telle façon que $\hat{\mathbf{C}}$ se définit comme l'unique solution de :

$$\mathbf{C} = \frac{m}{N} \sum_{i=1}^{N} \frac{\mathbf{r}_i \, \mathbf{r}_i^T}{\mathbf{r}_i^T \, \mathbf{C}^{-1} \, \mathbf{r}_i} \, . \tag{3.6}$$

Une fois $\hat{\mathbf{C}}$ calculée, cette matrice est aussi forcée à être de Toeplitz, par l'opérateur de rectification $\mathcal{T}$:

$$\check{\mathbf{C}} = \mathcal{T}\left(\hat{\mathbf{C}}\right) \, . \tag{3.7}$$

Les résultats du chapitre précédent s'appliquent et cet estimateur est un estimateur consistant. Les rendements sont ensuite blanchis : $\mathbf{r}_{w,i} = \check{\mathbf{C}}^{-1/2} \, \mathbf{r}_i$. Puis la matrice de covariance $\mathbf{\Sigma}$ est estimée grâce à l'estimateur de Tyler des données blanchies : $\check{\mathbf{\Sigma}}$ est l'unique solution de l'équation suivante en $\mathbf{\Sigma}$:

$$\mathbf{\Sigma} = \sum_{i=1}^{N} \frac{\mathbf{r}_{w,i} \, \mathbf{r}_{w,i}^T}{\mathbf{r}_{w,i}^T \, \mathbf{\Sigma}^{-1} \, \mathbf{r}_{w,i}} \, . \tag{3.8}$$

Une fois $\check{\mathbf{\Sigma}}$ calculée, il est ainsi possible de maximiser la fonction f choisie, et de définir les poids $\mathbf{w}$ nécessaires.

3.2.3 Comparaison des résultats

La méthode a ainsi été comparée par E. Jay et al [53] aux autres méthodes fréquemment utilisées. Des contraintes sont posées sur les poids w_i, qui correspondent aux contraintes

TABLE 3.3 – Performances du portefeuille d'actions

Méthode	Rendements Ann.	Volatilité Ann.	Rapport Rend/Vol	Max drawdown
RMT tyler	9.71 %	12.9 %	0.75	50.41%
SCM	8.51 %	13.80 %	0.62	55.02 %
Référence	4.92 %	15.19 %	0.32	58.36 %

posées sur les abondances des *endmembers* : la somme est supposée égale à un et les poids sont positifs, de valeur comprise entre 0 et 1. Pour tester la méthode, les prix de 40 actifs sont observés entre Juin 2000 et Janvier 2018. Les prix observés sont les prix à la fermeture de la bourse, dits prix de clôture. Ces actifs représentent des indices constitués d'actions européennes appartenant à un même secteur, à un même pays ou construit selon les valeurs d'un facteur identifié comme important sur les marchés financiers comme par exemple le facteur *Momentum* ou encore *Value* [6]. Toutes les quatre semaines, la matrice de covariance des actions est estimée sur la base des rendements de l'année précédente, et les poids qui permettent de maximiser l'équation (3.4) sont calculés sur cette période. Puis les poids sont fixés sur la période des quatre semaines suivantes. Les résultats sont alors comparés entre la méthode utilisant la SCM et la méthode décrite.

La figure 3.12 montre l'évolution de la richesse des portefeuilles, et compare les méthodes SCM et la méthode proposée notée "RMT Tyler blanchi". La méthode "*benchmark*" correspond aux valeurs de référence, qui représente l'évolution de l'indice Stoxx ®Europe 600 [105], constitué de 600 actions européennes de 17 pays d'Europe. Les actions incluses dans cet indice constituent la base de la construction des indices "pays", "secteurs" et "facteurs" sus-cités et donnent ainsi une indication de la valeur apportée par la méthode d'allocation en comparaison de la méthode d'allocation de l'indice lui-même.

Sur la figure 3.13 est représenté le "*turnover*" ou roulement des portefeuilles, c'est à dire l'ajustement qu'il a du être fait sur les poids entre deux périodes t et $t+1$, soit $\sum_{j=1}^{m} |w_{i,t+1} - w_{i,t}|$. Il permet de quantifier l'importance des changements à effectuer sur les compositions des portefeuilles entre deux périodes d'allocation consécutives (c'est à dire ici, toutes les quatre semaines). De plus, ces changements sont assujettis à des frais de transaction lors du passage des ordres sur le marché. Ce "*turnover*" est donc un indicateur de la stabilité de la méthode ainsi qu'une manière de quantifier les frais à payer pour modifier la composition du portefeuille à chaque rebalancement. Ces simulations et ces données ont été respectivement faites et rassemblées dans l'article [53].

Sur ces figures, les résultats sont en faveur de la méthode proposée, puisque celle-ci aboutit à des modifications moins brutales et donc à un "*turnover*" plus faible qu'avec la SCM. A noter cependant que les performances des portefeuilles sur la figure 3.12 ne tiennent pas compte des frais de transactions, l'écart entre les deux courbes bleue et verte serait donc légèrement plus important dans ce cas. De même pour la courbe du "*benchmark*" qui ne prend pas en compte les possible frais de transaction.

Enfin, le tableau suivant 3.3 donne quelques statistiques sur les performances du portefeuille résultant. Il compare sur la période totale, les rendements annuels, la volatilité annuelle, qui représente l'importance des fluctuations de valeurs d'une action, le rapport entre le rendement et la volatilité ainsi que le "*drawdown*" maximal c'est à dire la perte maximale enregistrée. La méthode proposée est notée RMT Tyler. La référence correspond aux valeurs de référence.

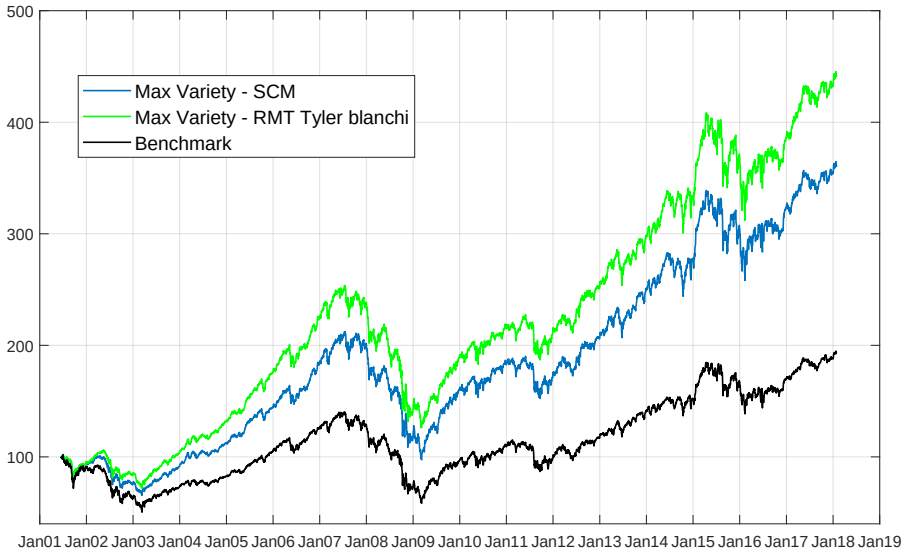

FIGURE 3.12 – richesse du portefeuille, débutant à 100 pour la première période, sur le début du mois de Janvier.

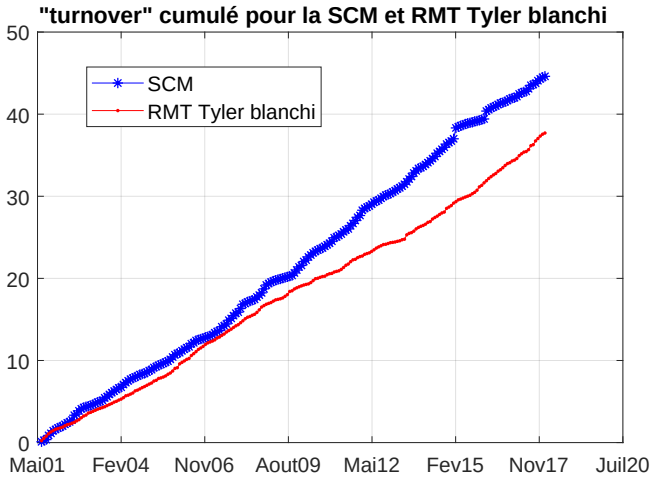

FIGURE 3.13 – "*Turnover*" entre les différentes périodes consécutives.

Une fois encore, c'est la méthode proposée qui donne les meilleurs résultats, pour tous les critères regardés. La meilleure estimation de la matrice de covariance entraîne donc une meilleure allocation des actifs pour les portefeuilles d'actions. L'amélioration vient, en grande partie, de l'estimation robuste de cette matrice, puisque la SCM ne prend pas bien en compte la non-gaussianité des signaux. De plus, le modèle choisi pour modéliser les données est une bonne représentation de celles-ci puisque les résultats obtenus avec des données réelles sont bons. La théorie ayant déjà été démontrée dans le chapitre précédent, la conformité du modèle est ainsi validée.

3.3 Synthèse

Le modèle du bruit CES corrélé a donc de nombreuses applications possibles, renforçant son intérêt. En ce qui concerne l'imagerie hyperspectrale, il a été montré que les algorithmes proposés donnent des résultats satisfaisants, et que la dernière méthode utilisant un M-estimateur de Tyler améliore les résultats obtenus par des méthodes plus "classiques", qui peuvent s'appliquer au cas du modèle choisi pour cette thèse.

Plusieurs applications des résultats sont envisageables, et les algorithmes présentés dans le chapitre précédent peuvent améliorer différents types de situation. C'est ce qu'il a été possible de montrer par l'application des résultats obtenus au cas de la constitution de portefeuilles d'actions. L'utilisation de la SCM, souvent systématique, peut être remplacée par le M-estimateur de Tyler rectifié proposé dans le chapitre précédent. Les résultats sont alors meilleurs, ce qu'il a été possible de montrer sur des données réelles.

Les applications des résultats théoriques du chapitre précédent confirment leur pertinence et la conformité du modèle choisi.

Chapitre 4

Le problème du démélange spectral

Ce chapitre s'attèle au problème du démélange spectral, c'est à dire, une fois que le nombre d'endmembers est connu, à déterminer quels sont leurs spectres et leurs positions dans l'image. Les méthodes proposées ici s'appuient sur des méthodes classiques de démélange spectral, qui sont ici rendues robustes ou modifiées pour permettre leur utilisation sous les hypothèses générales de la RMT. Une méthode limite est aussi proposée dans le cas où il y aurait un grand nombre d'endmembers, même si ceci ne correspond pas aux hypothèses initiales.

Le démélange spectral est un problème récurrent en imagerie hyperspectrale, où il s'agit d'identifier le spectre des *endmembers* ainsi que leurs positions dans l'image. Pour cela, il est nécessaire d'avoir, au préalable, au moins identifié le nombre d'*endmembers* présents dans l'image hyperspectrale. C'est d'ailleurs ce problème qui est traité dans les chapitres précédents de cette thèse. Bien entendu, connaître par avance le spectre des *endmembers* présents sur l'image ou disposer d'une bibliothèque spectrale rend la résolution plus facile. Ici, comme précédemment, il n'est supposé aucun a priori sur l'image hyperspectrale considérée, c'est à dire qu'il n'y a aucune information sur les spectres et les positions des *endmembers*. Il n'y a pas non plus de données secondaires. Le nombre d'*endmembers* est supposé connu ou correctement estimé. Le problème de la variabilité spectrale, c'est à dire la variabilité du spectre, due par exemple aux conditions climatiques ou à d'autres phénomènes (comme décrits en introduction), n'en est ici plus un, puisque les spectres ne sont pas déduits d'une librairie spectrale mais de l'image en elle-même. En effet, un spectre n'est pas recherché à partir d'un dictionnaire mais il est identifié grâce à l'image. L'étape suivante, c'est à dire l'identification de ce spectre, devra alors quant à elle tenir compte de la variabilité spectrale. Mais ce n'est pas le problème considéré ici.

Le problème peut se définir comme suit : à partir d'une image $\mathbf{Y} = \mathbf{M}\mathbf{S} + \mathbf{C}^{1/2}\mathbf{X}\mathbf{T}^{1/2}$, de l'estimation $\check{\mathbf{C}}$ et de p, la dimension du sous espace des *endmembers*, retrouver les matrices $\mathbf{M}$ et $\mathbf{S}$. Ces matrices estimées seront, par la suite, notées $\hat{\mathbf{M}}$ et $\hat{\mathbf{S}}$. La matrice $\mathbf{S}$ doit également vérifier l'hypothèse suivante, comme postulé dans le chapitre d'introduction :

Hypothèse 6

$$\begin{aligned} &\text{Non-négativité} : \forall i \in [1,p]\,,\ \forall j \in [1,N]\ \hat{\mathbf{s}}_{i,j} \geq 0\,, \\ &\text{Somme égale à un} : \forall j \in [1,N]\,,\ \sum_{i=1}^{p} \hat{\mathbf{s}}_{i,j} = 1\,. \end{aligned} \tag{4.1}$$

Les différents types d'images hyperspectrales qui existent sont divisés ici en deux catégories : des images qui contiennent un nombre modéré d'*endmembers* en comparaison au nombre de bandes spectrales retenues et au nombre de pixel, c'est à dire $p \ll m$, et des images qui contiennent un grand nombre d'*endmembers*, c'est à dire $p \sim m$ ou $p < m$. Le premier cas de figure se rencontre par exemple dans le cas de la détection de changement ou de la détection d'anomalies ou encore de cibles sur une image hyperspectrale. C'est aussi le cas de beaucoup de données géologiques prises en aéroporté, lorsque la résolution des capteurs n'est pas trop importante et que l'étendue spatiale reste raisonnable. Le second cas est moins souvent rencontré, c'est le cas de données de grande précision spatiale, ou encore lorsque les *endmembers* recherchés sont définis avec une grande précision spectrale, soit encore le cas d'images qui contiennent tout simplement un grand nombre d'*endmembers*.

4.1 Démélange spectral dans le cas où $p \ll m$

Ce cas correspond à un cas assez répandu où le nombre d'*endmembers* n'est pas du même ordre que le nombre de bandes spectrales. C'est d'ailleurs le cas étudié dans les chapitres précédents, c'est à dire que $p \ll m$. Dans ce cas, il est possible d'utiliser des méthodes de démélange spectral dîtes "classiques" et de les améliorer au cas d'un bruit non gaussien et de la grande dimension.

4.1.1 Problème et méthodes

Le type de méthode retenue ici est celui présenté en introduction, sur les méthodes géométriques de démélange spectral. C'est ce type d'algorithme amélioré qui sera utilisé par la suite. Avant de présenter les apports de cette thèse sur ces algorithmes, voici un rappel sur les méthodes qui serviront d'éléments de comparaison et de point de départ pour les nouveaux algorithmes.

La méthode MVC-NMF semble souvent utilisée pour des images hyperspectrales. Pour rappel, cette méthode, issue de l'article [77], minimise le critère suivant de façon itérative et alternative en fonction des matrices de mélange et des abondances $\hat{\mathbf{M}}$ et $\hat{\mathbf{S}}$:

$$f(\hat{\mathbf{M}}, \hat{\mathbf{S}}) = \frac{1}{2} J\left(\hat{\mathbf{M}}, \hat{\mathbf{S}}\right) + \lambda K\left(\hat{\mathbf{M}}\right), \tag{4.2}$$

sous l'hypothèse 4, où $K(\hat{\mathbf{M}})$ est une fonction de pénalité sur le volume du simplex de l'estimation de la matrice de $\mathbf{M}$, $\lambda \in \mathbb{R}$ un paramètre de régularisation à fixer. Ici, cette fonction $K(.)$ est choisie de la même façon que dans l'article [77], comme déjà précisé dans le chapitre d'introduction, c'est à dire : $K(\hat{\mathbf{M}}) = \frac{1}{2\,(c-1)!} \left| \begin{pmatrix} \mathbf{1}_p^T \\ \tilde{\mathbf{M}} \end{pmatrix} \right|^2$, avec $\tilde{\mathbf{M}} = \mathbf{U}^T \left(\hat{\mathbf{M}} - \boldsymbol{\nu} \mathbf{1}_p^T\right)$ et $\boldsymbol{\nu}$ un vecteur colonne contenant la moyenne sur les lignes de la matrice $\mathbf{Y}$, $\mathbf{U}$, de dimension $m \times (p-1)$ contenant les $p-1$ données les plus importantes de $\mathbf{Y}$ obtenues par PCA. Cette contrainte de volume ne représente pas exactement le volume du simplex des *endmembers*, seulement une approximation. Cependant, ceci n'est pas un problème, puisque, la contrainte de volume servant à restreindre et minimiser le volume pris par le simplex des *endmembers*, cette fonction K remplit parfaitement cette tâche. De plus, comme un calcul de volume par un déterminant nécessite une matrice carrée, même si ce calcul ne représente pas le véritable volume du simplex des *endmembers*, il a le mérite de pouvoir se calculer facilement.

Enfin, la contrainte sur les sommes des colonnes de la matrice $\hat{\mathbf{S}}$ qui doivent être égales à un est prise en compte en rajoutant une ligne dans les matrices $\hat{\mathbf{M}}$ et $\mathbf{Y}$ lors des étapes d'estimation de la matrice $\mathbf{S}$. Ces matrices sont notées $\bar{\mathbf{M}}$ et $\bar{\mathbf{Y}}$, et sont calculées comme suit : avec $\mathbf{1}_N$ et $\mathbf{1}_p$ deux vecteurs de dimension N et p respectivement, constitués de 1, tels

que

$$\bar{\mathbf{M}} = \begin{pmatrix} \hat{\mathbf{M}} \\ \delta \mathbf{1}_N \end{pmatrix} \text{ et } \bar{\mathbf{Y}} = \begin{pmatrix} \mathbf{Y} \\ \delta \mathbf{1}_p \end{pmatrix} .$$

Ce paramètre δ joue un rôle sur la vitesse de convergence vers une solution pour l'estimation de la matrice $\mathbf{S}$, et permet d'avoir une main mise sur l'importance donnée au critère. Le choisir entre 10 et 20 semble, d'après l'article [77], un bon compromis. L'algorithme est un algorithme de descente itératif, où, à chaque itération k, des nouvelles valeurs pour les matrices $\hat{\mathbf{M}}$ et $\hat{\mathbf{S}}$ sont calculées :

$$\hat{\mathbf{M}}_k \leftarrow \hat{\mathbf{M}}_{k-1} + \mu_{\hat{\mathbf{M}}_{k-1}} \frac{\partial J\left(\hat{\mathbf{M}}_{k-1}, \hat{\mathbf{S}}_{k-1}\right)}{\partial \hat{\mathbf{M}}} , \tag{4.3}$$

$$\hat{\mathbf{S}}_k \leftarrow \hat{\mathbf{S}}_{k-1} + \mu_{\hat{\mathbf{S}}_{k-1}} \frac{\partial J\left(\bar{\mathbf{M}}_{\mathbf{k-1}}, \hat{\mathbf{S}}_{k-1}\right)}{\partial \hat{\mathbf{S}}} . \tag{4.4}$$

Les pas de descente $\mu_{\mathbf{M}_{k-1}}$ et $\mu_{\mathbf{S}_{k-1}}$ peuvent être choisis de différentes façons. La méthode donnant de meilleurs résultats est de calculer le pas optimal qui vérifie la règle d'Armijo A.2. Cela signifie concrètement qu'à chaque itération, des itérations sont rajoutées dans l'algorithme pour calculer le pas optimal, qui se calcule à l'aide des dérivées partielles de J et pour un paramètre $\omega = 0.1$ (A.2). Ici, à chaque itération du calcul du pas, le pas est réduit de moitié en partant d'un pas de 1, jusqu'à ce que le critère soit vérifié. Quant aux directions de descente, elles correspondent ici aux directions des gradients des matrices. Quant à la fonction J, elle est donnée par l'équation (1.9), c'est à dire :

$$J\left(\hat{\mathbf{M}}, \hat{\mathbf{S}}\right) = \left\| \mathbf{Y} - \hat{\mathbf{M}}\hat{\mathbf{S}} \right\|_F^2 . \tag{4.5}$$

La méthode PGNMF, également présentée en introduction [64], qui, pour rappel, projette avant chaque itération les nouvelles valeurs sur un espace permettant de vérifier les contraintes (4.1), sera aussi utilisée comme un élément de comparaison. L'algorithme de cette méthode est décrit dans l'article [64]. Cet algorithme n'est pas rappelé ici puisqu'il peut être trouvé complet dans cet article et qu'il a été pris tel quel pour les simulations.

4.1.2 Nouvelles méthodes de démélange d'images hyperspectrales

Deux méthodes sont proposées. Les méthodes développées dans cette partie s'appuient toutes deux sur la méthode MVC-NMF. La première méthode, cherche à améliorer la robustesse de l'algorithme aux *outliers* et aux données qui ne correspondraient pas au modèle choisi. Le critère à optimiser cherche à minimiser l'impact des données qui pourraient avoir un effet néfaste sur ce critère. La seconde méthode est assez différente, et se concentre sur le blanchiment et la projection des données, rendu possible grâce à l'estimation correcte de la matrice $\mathbf{C}$, estimation proposée dans le Chapitre 2. Elle exploite la connaissance de la matrice $\mathbf{C}$ et de la dimension du sous espace signal, pour améliorer le critère à optimiser en le rendant théoriquement plus petit.

Méthode Robuste

La première méthode est une méthode robuste, qui utilise une pondération avec poids de Huber [50], comme proposé dans l'article [36]. C'est à dire qu'au lieu de chercher à minimiser la fonction $\| \mathbf{Y} - \hat{\mathbf{M}}\hat{\mathbf{S}} \|_F^2$, le problème se pose alors comme suit :

$$\arg\min_{\mathbf{M},\mathbf{S}} \sum_{i,j} \rho\left([\mathbf{Y}]_{i,j} - [\hat{\mathbf{M}}\hat{\mathbf{S}}]_{i,j} \right) \text{ , sous contraintes} , \tag{4.6}$$

les contraintes pouvant être diverses. Dans cette thèse sera considéré les contraintes (4.1). Avec la fonction ρ définie de la manière suivante, à l'aide d'un paramètre υ qui peut être fixé grâce à différents paramètres ou à une loi du χ^2, même si en pratique, la méthode est peu sensible à sa valeur exacte :

$$\rho(u) = \begin{cases} u^2 & \text{pour} \quad |u| \leq \upsilon/2 \\ \upsilon|u| - \upsilon^2/4 & \text{pour} \quad |u| > \upsilon/2 \end{cases} \tag{4.7}$$

C'est ici une application de la fonction de coût de Huber au cas matriciel. Il s'agit ensuite de minimiser la fonction suivante :

$$J\left(\hat{\mathbf{M}}, \hat{\mathbf{S}}\right) = \rho\left(\left(\mathbf{Y} - \hat{\mathbf{M}}\,\hat{\mathbf{S}}\right)_{i,j}\right) + \lambda\, K\left(\hat{\mathbf{M}}\right) . \tag{4.8}$$

Puis, soient les notations suivantes : pour tout $i \in [\![1, N]\!]$, $\mathbf{y}_i = \mathbf{M}\,\mathbf{s}_i + \mathbf{b}_i$, avec $\mathbf{s}_i$ un vecteur colonne de $\mathbf{S}$ et $\mathbf{M}(j,:)$, de taille $1 \times p$, un vecteur ligne de la matrice $\mathbf{M}$, pour tout $j \in [\![1, m]\!]$. Alors, l'équation précédente (4.8) revient à trouver les matrice $\hat{\mathbf{M}}$ et $\hat{\mathbf{S}}$ telles qu'à chaque itération, elles soient données par :

$$\forall i\,,\, i \in [\![1, N]\!]\ \ \hat{\mathbf{s}}_i = \arg\min_{\hat{\mathbf{s}}_i} \sum_{j=1}^{m} \rho\left(y_{i,j} - \hat{\mathbf{M}}(j,:)\,\hat{\mathbf{s}}_i\right) , \tag{4.9}$$

$$\forall j\,,\, j \in [\![1, m]\!]\ \ \hat{\mathbf{M}}(j,:) = \arg\min_{\hat{\mathbf{M}}(j,:)} \left(\sum_{k=1}^{N} \rho\left(y_{j,k} - \hat{\mathbf{M}}(j,:)\,\hat{\mathbf{s}}_k\right) + \lambda\, K\left(\hat{\mathbf{M}}\right) \right) . \tag{4.10}$$

Initialisation

L'algorithme est initialisé de la façon suivante : un premier vecteur colonne de la matrice des données $\mathbf{Y}$ est tiré pour constituer la première colonne de la matrice $\hat{\mathbf{M}}$. Un second vecteur colonne de $\mathbf{Y}$ est tiré et rajouté à la matrice $\hat{\mathbf{M}}$ à condition que la divergence de Kullback-Leibler (voir A.1) entre les deux colonnes de $\hat{\mathbf{M}}$ dépasse un certain seuil. Un troisième vecteur est ensuite tiré de la matrice $\mathbf{Y}$ et ainsi de suite jusqu'à obtenir p colonnes dans la matrice $\hat{\mathbf{M}}$, en vérifiant à chaque fois que la divergence de Kullback-Leibler entre le nouveau vecteur ajouté à la matrice $\hat{\mathbf{M}}$ et les autres colonnes de cette matrice dépasse le seuil. C'est d'ailleurs une méthode qui est déjà utilisée, par exemple, dans l'article [65]. Ce seuil est fixé arbitrairement et va dépendre des données. Cette méthode permet de s'assurer que les vecteurs pris comme spectres initiaux pour les *endmembers* sont assez différents les uns des autres, ceci permettant surtout à l'algorithme de converger plus vite vers une solution possible. Une fois la matrice $\hat{\mathbf{M}}$ initialisée, la matrice $\hat{\mathbf{S}}$ est obtenue en multipliant la pseudo inverse de la matrice $\hat{\mathbf{M}}$ par la matrice des données $\mathbf{Y}$.

Phase de calcul

En ce qui concerne le calcul des dérivées partielles, il ne sera développé ici que pour $\hat{\mathbf{S}}$, le calcul pour $\hat{\mathbf{M}}$ étant le même, à l'exception du terme $K\left(\hat{\mathbf{M}}\right)$. Ce terme peut être dérivé comme dans l'article [77], c'est à dire :

$$\frac{\partial K\left(\hat{\mathbf{M}}\right)}{\partial \hat{\mathbf{M}}} = \frac{1}{(p-1)!}\,|\mathbf{Z}|^2\ \mathbf{U}\,\mathbf{B}^T\,\mathbf{Z}^{-T}\,, \tag{4.11}$$

où $\mathbf{Z} = \begin{pmatrix} \mathbf{1}_p^T \\ \mathbf{0}_{(p-1)\times p} \end{pmatrix} + \begin{pmatrix} \mathbf{0}_{p-1\times 1}^T \\ \mathbf{I}_{p-1} \end{pmatrix} \tilde{\mathbf{M}}$.

Pour le calcul du reste de la dérivée partielle en fonction de $\hat{\mathbf{S}}$, il est en premier lieu nécessaire de regarder la valeur de $u = \left| y_{i,j} - \hat{\mathbf{M}}(j,:)\,\hat{\mathbf{s}}_i \right|^2$ pour tout $j \in [\![1, m]\!]$ et pour tout $i \in [\![1, N]\!]$.

- Si $u \leq \frac{\upsilon}{2}$: $\dfrac{\partial \rho(u)}{\partial \hat{\mathbf{s}}_i} = 2\,\hat{\mathbf{M}}(j,:)\,\left(y_{i,j} - \hat{\mathbf{M}}(j,:)\,\hat{\mathbf{s}}_i\right)$,
- Si $u > \frac{\upsilon}{2}$: $\dfrac{\partial \rho(u)}{\partial \hat{\mathbf{s}}_i} = \lambda\,\dfrac{-\hat{\mathbf{M}}(j,:)\,\left(y^{\star}_{i,j} - \hat{\mathbf{M}}(j,:)\,\hat{\mathbf{s}}_i\right) - \left(y_{i,j} - \hat{\mathbf{M}}(j,:)\,\hat{\mathbf{s}}_i\right)\,\hat{\mathbf{M}}(j,:)}{2\,\left|y_{i,j} - \hat{\mathbf{M}}(j,:)\,\hat{\mathbf{s}}_i\right|}$.

Il s'agit ensuite de sommer ces dérivées partielles pour donner, à chaque itération, les matrices $\hat{\mathbf{M}}$ et $\hat{\mathbf{S}}$ optimales, en les mettant à jour suivant les équations (4.4). Il est ensuite indispensable de transformer ce calcul pour qu'il ne se fasse non plus sur chaque élément des matrices $\hat{\mathbf{M}}$ et $\hat{\mathbf{S}}$ mais sur les matrices $\hat{\mathbf{M}}$ et $\hat{\mathbf{S}}$ directement, afin de diminuer fortement les temps de calculs. Ceci peut se faire par diverses manipulations et permet d'insérer plus facilement le terme $\dfrac{\partial K\left(\hat{\mathbf{M}}\right)}{\partial \hat{\mathbf{M}}}$ dans la dérivée partielle de la fonction J par rapport à $\hat{\mathbf{M}}$.

Arrêt

Le critère d'arrêt pour cet algorithme est la distance euclidienne entre $\mathbf{Y}$ et $\hat{\mathbf{M}}\,\hat{\mathbf{S}}$, mais aussi un nombre d'itérations minimal qui doit être respecté.

Ce nombre d'itérations minimal permet de ne pas donner comme solution un minimum local trop rapidement et d'être plus robuste à l'existence de ces minima locaux. En effet, la méthode n'étant pas convexe, des minima locaux peuvent exister. Tant que la méthode se rapproche d'un de ces minima, le critère diminue. Par contre, si le nombre d'itérations minimal n'est pas respecté, même si le critère calculé, c'est à dire la fonction J, est à son minimum, l'algorithme continue et le critère augmente. Dès que le critère est de nouveau une fonction décroissante, le nombre d'itérations compté est remis à zéro. De cette manière, si le critère converge vers un second minimum, c'est à dire si le critère diminue, le nombre d'itérations comparé au nombre d'itérations minimal à respecter est remis à zéro, et l'algorithme continue ; tandis que si le critère continue d'augmenter, c'est que d'autres minima ne sont pas à proximité, le nombre d'itérations compté va dépasser le nombre d'itérations minimal choisi et l'algorithme s'arrête. Ce nombre minimum ne doit pas permettre à l'algorithme de s'éloigner d'un minimum qui pourrait être le minimum global. C'est donc une difficulté supplémentaire à la méthode, mais celle-ci donne tout de même de meilleurs résultats que la technique qui consiste à s'arrêter dès que le critère ne décroit plus.

Cette méthode rend donc le critère plus robuste, et travaille sur la non gaussianité des données.

Méthode Projetée

Une autre idée qui fût testée est la suivante : connaissant la matrice $\check{\mathbf{C}}$ ainsi que la dimension du sous espace signal, il est possible de blanchir le signal pour le projeter ensuite sur le sous espace correspondant. Il s'agit en fait de tirer profit de la connaissance d'un bon estimateur de $\mathbf{C}$.

La méthode se décline donc comme suit : soient les données $\mathbf{Y}$:

$$\mathbf{Y} = \mathbf{M}\,\mathbf{S} + \mathbf{C}^{1/2}\,\mathbf{X}\,\mathbf{T}^{1/2}\,. \tag{4.12}$$

Alors, le blanchiment par la matrice $\check{\mathbf{C}}$ donne :

$$\begin{aligned} \check{\mathbf{C}}^{-1/2}\mathbf{Y} &= \check{\mathbf{C}}^{-1/2}\,\mathbf{M}\,\mathbf{S} + \check{\mathbf{C}}^{-1/2}\,\mathbf{C}^{-1/2}\,\mathbf{X}\,\mathbf{T}^{1/2}\,, \\ \text{et } \frac{1}{N}\,\left(\check{\mathbf{C}}^{-1/2}\,\mathbf{Y}\,\mathbf{Y}^{H}\,\check{\mathbf{C}}^{-H/2}\right) &= \mathbf{U}^{-1}\,\mathbf{\Delta}\,\mathbf{U}\,, \end{aligned}$$

avec $\mathbf{\Delta}$ une matrice diagonale contenant les valeurs propres de l'estimation de l'espérance du signal blanchi, et $\mathbf{U}$ la matrice contenant, en colonne, les vecteurs propres associés.

Il est alors possible de décomposer la matrice des vecteurs propres en deux parties, celle concernant le sous espace signal et celle concernant le sous espace bruit, la partie du sous espace signal correspondant aux vecteurs propres associés aux plus grandes valeurs propres : $\mathbf{U} = \mathbf{U}_s + \mathbf{U}_b$ où $\mathbf{U}_s$ est lié au sous espace signal et $\mathbf{U}_b$ au sous espace bruit.

La projection sur le sous espace signal peut ainsi se faire de la manière suivante :

$$\mathbf{U}_s \, \mathbf{U}_s^H \, \check{\mathbf{C}}^{-1/2} \, \mathbf{Y} = \check{\mathbf{C}}^{-1/2} \mathbf{M} \, \mathbf{S} + \mathbf{\Xi} \, ,$$
$$\text{puis : } \mathbf{Y}_p = \check{\mathbf{C}}^{1/2} \, \mathbf{U}_s \, \mathbf{U}_s^H \, \check{\mathbf{C}}^{-1/2} \, \mathbf{Y} = \mathbf{M} \, \mathbf{S} + \mathbf{\Xi}' \, ,$$

avec $\mathbf{\Xi}$ et $\mathbf{\Xi}'$ le bruit résiduel.

Ainsi, la méthode projetée proposée consiste à minimiser :

$$J\left(\hat{\mathbf{M}}, \hat{\mathbf{S}}\right) = \left\| \check{\mathbf{C}}^{1/2} \, \mathbf{U}_s \, \mathbf{U}_s^H \, \check{\mathbf{C}}^{-1/2} \, \mathbf{Y} - \hat{\mathbf{M}} \, \hat{\mathbf{S}} \right\|_F^2 + \lambda \, K\left(\hat{\mathbf{M}}\right) \, , \tag{4.13}$$

toujours en suivant la trame de la méthode proposée dans l'article [77].

La méthode reprend ensuite les mêmes étapes que la méthode précédente.

Initialisation
L'initialisation est la même que pour la méthode robuste, c'est à dire que la matrice $\hat{\mathbf{S}}$ initiale est obtenue par pseudo-inverse et la matrice $\hat{\mathbf{M}}$ initiale est obtenue en prenant des vecteurs colonnes de $\mathbf{Y}$ qui vérifient une condition sur la divergence de Kullback-Leibler.

Phase de calcul
L'algorithme en lui même suit le même cheminement que celui proposé dans [77], mis à part que les données utilisées sont projetées. Ainsi, à chaque itération est d'abord calculé :

$$\frac{\partial J\left(\hat{\mathbf{M}}_{k-1}, \hat{\mathbf{S}}_{k-1}\right)}{\partial \hat{\mathbf{M}}} = \left(\hat{\mathbf{M}}_{k-1} \, \hat{\mathbf{S}}_{k-1} - \mathbf{Y}_p\right) \, \hat{\mathbf{S}}_{k-1}^H + \frac{\lambda}{(p-1)!} \, |\mathbf{Z}|^2 \, \mathbf{U}\mathbf{B}^T \, \mathbf{Z}^{-T} \, , \tag{4.14}$$

$$\frac{\partial J\left(\bar{\mathbf{M}}_{\mathbf{k-1}}, \hat{\mathbf{S}}_{k-1}\right)}{\partial \hat{\mathbf{S}}} = \bar{\mathbf{M}}_{k-1} \, \left(\bar{\mathbf{M}}_{k-1} \, \hat{\mathbf{S}}_{k-1} - \bar{\mathbf{Y}}_p\right) \, . \tag{4.15}$$

Puis la mise à jour des matrices se fait grâce aux équations (4.4).

Arrêt
Le critère d'arrêt est là encore, le même que pour la partie précédente.

4.1.3 Applications

Une première approche intuitive

Pour permettre une comparaison avec les résultats proposés, une méthode simple est d'abord appliquée aux images qui seront utilisées par la suite. Il s'agit de la méthode PCA (*Pincipal Component Analysis*, déjà présentée en introduction et développée dans l'article [97]). Cette méthode consiste juste à regarder quelles sont les composantes principales de l'image hyperspectrale, et pourrait sembler, à première vue, être une bonne approche de démélange pour ce genre d'images.

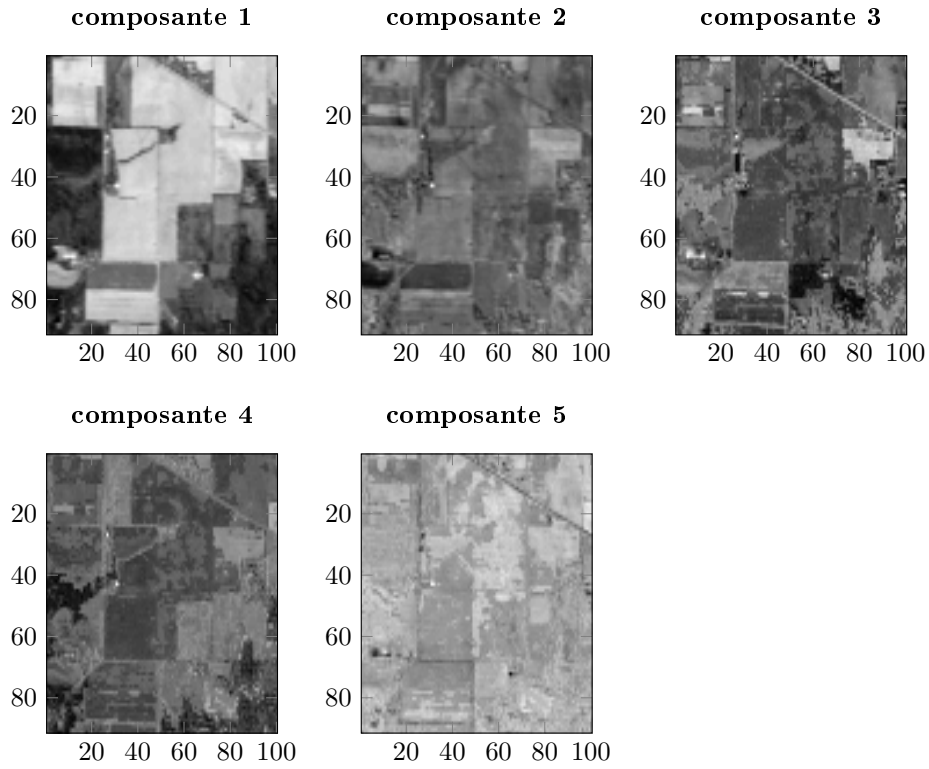

FIGURE 4.1 – 5 premières composantes de la PCA sur l'image Indian Pines [35].

Les figures 4.1, 4.3, 4.5 et 4.7 représentent les 5 principales composantes de la décomposition en composantes principales des images Indian Pines, Pavia U, Salinas A de [35] et l'image "Cube". Ces images ont déjà été utilisées et représentées partiellement dans le Chapitre 3, sur les figures 3.10 et 3.11. Les autres composantes correspondant à du bruit et n'apportant rien d'autre, elles n'ont pas été représentées. Sur certaines, seules les trois premières composantes semblent correspondre à des *endmembers*. C'est par exemple le cas de l'image 4.3 où les trois dernières composantes ne semblent correspondre à rien. Les parties appartenant à la composante ressortent en blanc, tandis que le reste est noir.

Á part pour la dernière image 4.7 où la voiture semble correctement identifiée sur la composante 3 et la végétation sur la composante 2, les résultats sur les autres images sont plus mitigés. Sur l'image 4.1, les différents types de champs ne ressortent pas en tant que différents composants. La figure 4.2 montre la vérité terrain pour cette image, où chaque couleur correspond à un *endmember* différent. Sur cette image, il n'y a pas de correspondance entre les différents éléments de la PCA et les *endmembers*, puisque les *endmembers* n'apparaissent pas en clair à tour de rôle sur les différents composants. En ce qui concerne les autres images, le constat est le même : la méthode PCA ne permet pas de correctement identifier les *endmembers* : comme sur les images 4.6 et 4.4 qui montrent la vérité terrain, il n'est pas possible de distinguer les différents *endmembers* entre les différentes composantes résultantes de la PCA.

Ces résultats montrent l'importance de développer des algorithmes de traitement des images hyperspectrales plus complexes. Considérer seulement des plus grandes valeurs propres et les vecteurs propres associés pour former la matrice des abondances ainsi que les méthodes telles que la PCA ne semblent pas adaptées.

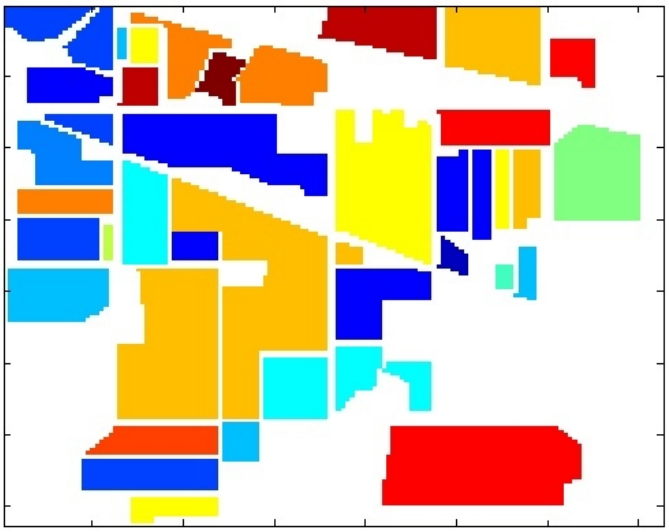

FIGURE 4.2 – Vérité terrain pour la figure Indian Pines.

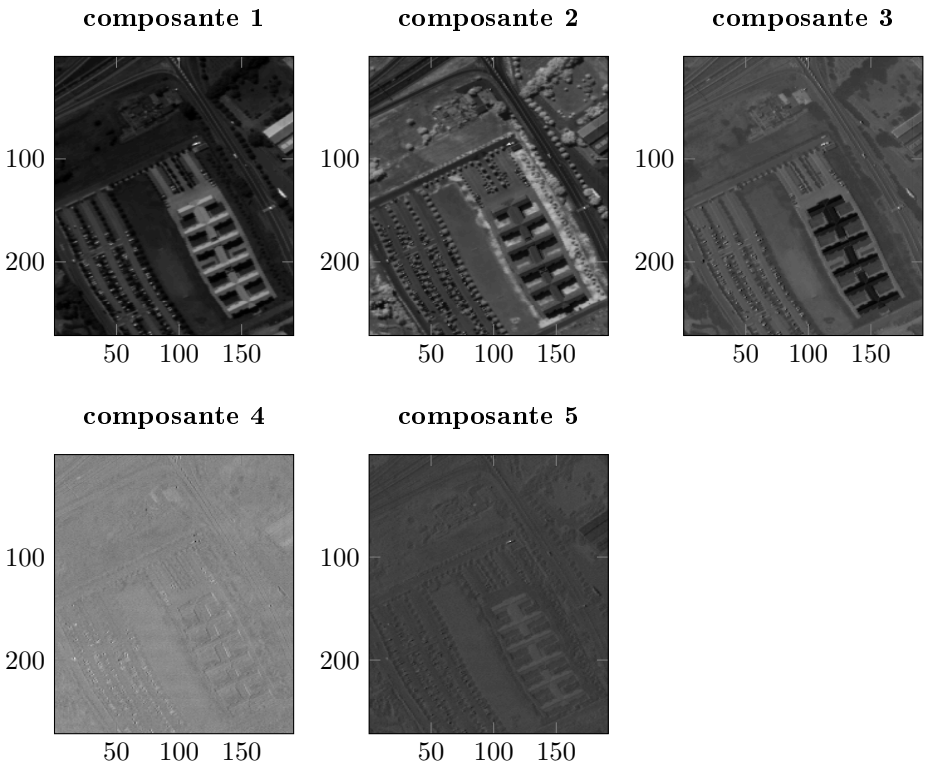

FIGURE 4.3 – 5 premières composantes de la PCA sur l'image Pavia U [35].

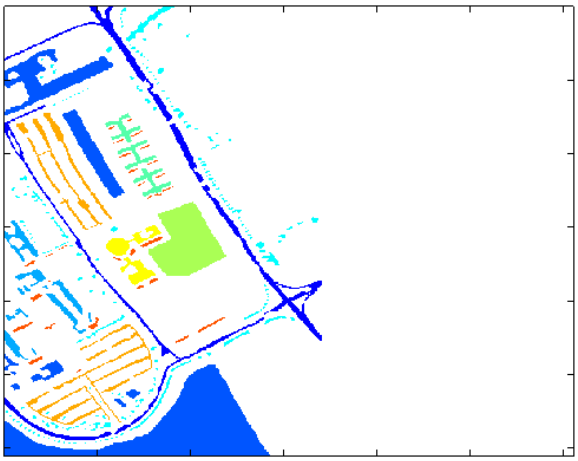

FIGURE 4.4 – Vérité terrain pour la figure Indian Pines.

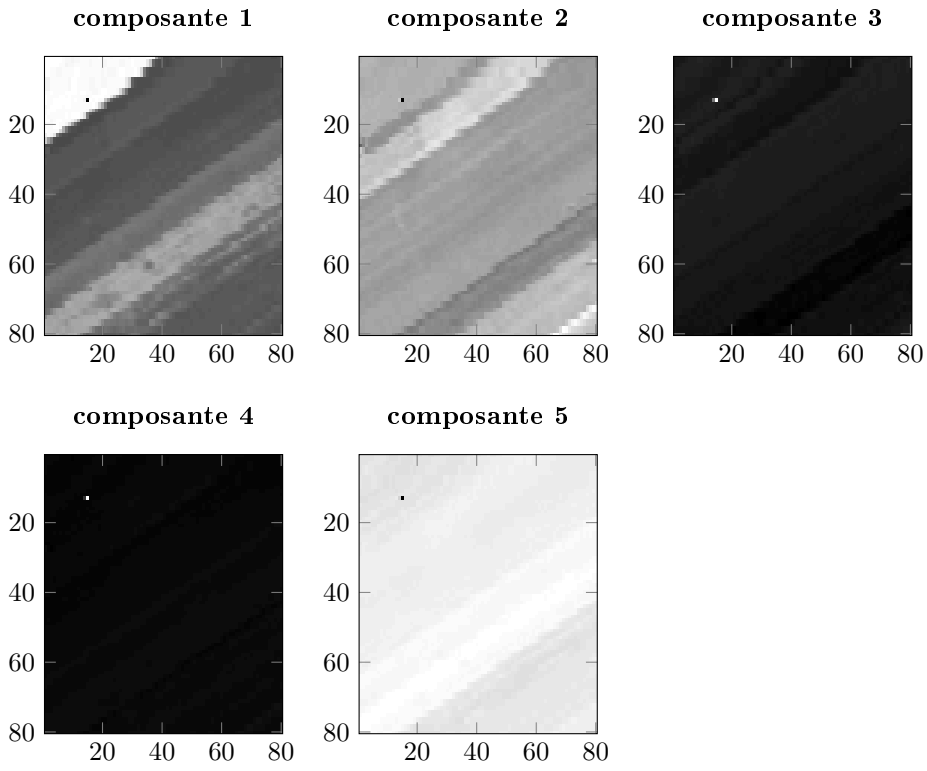

FIGURE 4.5 – 5 premières composantes de la PCA sur l'image Salinas A [35].

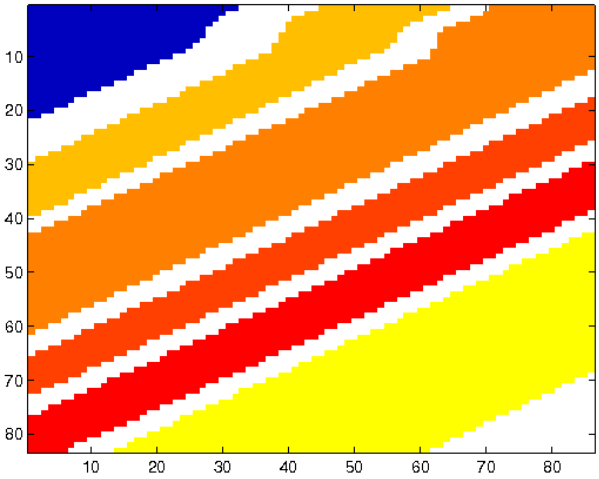

FIGURE 4.6 – Vérité terrain pour la figure Indian Pines, une couleur correspondant à un *endmember*.

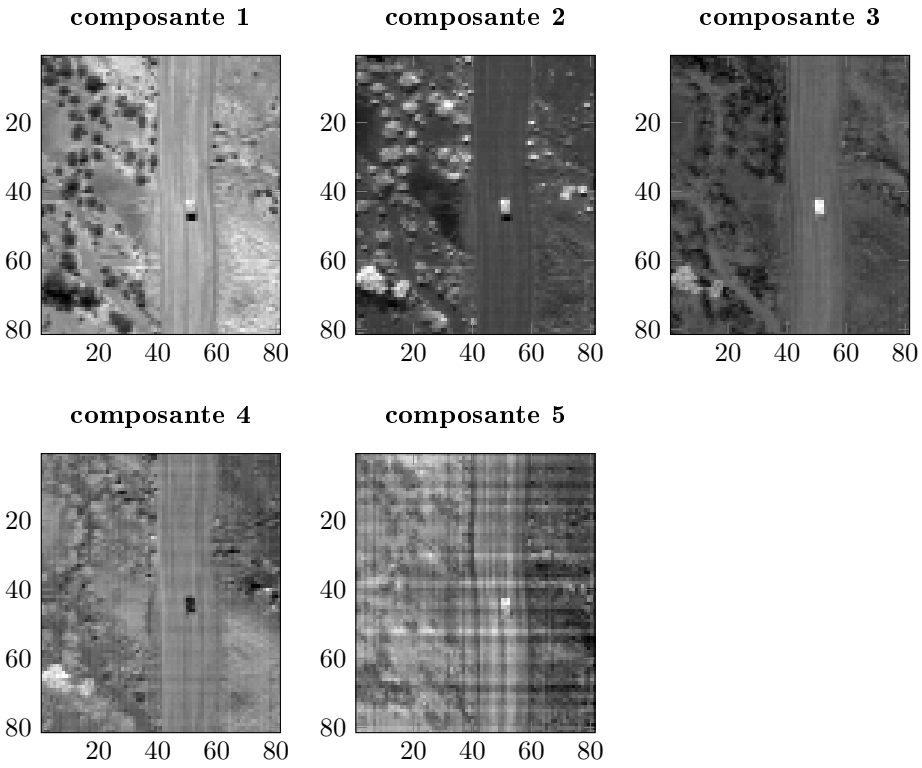

FIGURE 4.7 – 5 premières composantes de la PCA sur l'image contenant les voitures.

Quelques critères d'évaluation

Afin de comparer les résultats sur d'autres critères que la répartition visuelle des abondances, la littérature sur le sujet propose différents critères d'évaluations dont ceux qui vont être présentés. Certains sont aussi exploités dans l'article [77]. Ces critères reposent sur un postulat qui n'est pas négligeable : il suppose la connaissance des matrices $\mathbf{M}$ et $\mathbf{S}$ réelles. Cela n'est pas en contradiction ici avec les hypothèses générales de cette thèse puisque ce paragraphe cherche à démontrer l'efficacité des méthodes proposées pour pouvoir ensuite les utiliser sur des images dont on ne connaitrait pas la vérité terrain.

Même si la distance euclidienne entre $\mathbf{Y}$ et $\hat{\mathbf{M}}\hat{\mathbf{S}}$ est toujours utilisée par les algorithmes comme critère d'arrêt, il semble intéressant de regarder la concordance entre les matrices $\mathbf{M}$ et son estimée $\hat{\mathbf{M}}$ mais aussi entre $\mathbf{S}$ et son estimée $\hat{\mathbf{S}}$, afin d'évaluer la performance des algorithmes. Il est ainsi possible de regarder les distances euclidiennes entre ces matrices mais aussi de regarder d'autres critères qui peuvent s'avérer plus précis : soit $\mathbf{m} = [m_1, \ldots, m_m]^T$ un vecteur colonne de la matrice $\mathbf{M}$ et $\hat{\mathbf{m}} = [\hat{m}_1, \ldots, \hat{m}_m]^T$ le vecteur colonne de la matrice $\hat{\mathbf{M}}$ correspondant.

Remarque : Par "correspondant" s'entend correspondant à l'estimation de $\mathbf{m}$, et non à la position de $\mathbf{m}$. En effet, l'estimation de la matrice $\mathbf{M}$ (et de la matrice $\mathbf{S}$) est donnée à une rotation près puisque pour toute matrice de rotation $\mathbf{U}$, $\mathbf{M}\,\mathbf{U}$ et $\mathbf{U}^{-1}\,\mathbf{S}$ vérifient l'hypothèse 4 et sont également solutions du problème d'optimisation. Ainsi, chaque vecteur $\hat{\mathbf{m}}$ correspond à l'estimation d'un vecteur $\mathbf{m}$ qui n'a pas forcément la même position dans la matrice $\hat{\mathbf{M}}$ que $\mathbf{m}$ dans la matrice $\mathbf{M}$.

De même, soit $\mathbf{s}$ un vecteur ligne de la matrice $\mathbf{S}$ et $\hat{\mathbf{s}}$ le vecteur correspondant sur la matrice $\hat{\mathbf{S}}$. Soient, pour un vecteur $\mathbf{x} = [x_1, \ldots, x_k]^T$ de dimension k, $\mathbf{p}(\mathbf{x}) = \mathbf{x} / \sum_j x_j$, $\hat{\mathbf{p}} = \mathbf{p}(\hat{\mathbf{x}})$ et $D(\mathbf{x}|\hat{\mathbf{x}}) = \sum_j p_j(\mathbf{x}) \log\left(\frac{p_j(\mathbf{x})}{\hat{p}_j(\mathbf{x})}\right)$. Alors il est possible de considérer les critères d'évaluation suivants :

$$SAD = acos\left(\frac{\mathbf{m}^T\,\hat{\mathbf{m}}}{\|\mathbf{m}|\ \|\hat{\mathbf{m}}\|}\right), \tag{4.16}$$

$$SID = D\left(\mathbf{m}|\,\hat{\mathbf{m}}\right) + D\left(\hat{\mathbf{m}}|\mathbf{m}\right), \tag{4.17}$$

$$AID = D\left(\mathbf{s}|\,\hat{\mathbf{s}}\right) + D\left(\hat{\mathbf{s}}|\mathbf{s}\right). \tag{4.18}$$

Le premier critère représente l'écart angulaire entre le vecteur estimé et le vecteur original dans l'espace des spectres. Plus la méthode est bonne, plus ce critère doit se rapprocher de zéro. Les seconds vecteurs représentent la divergence de Kullback-Leibler (voir A.1) entre les vecteurs et leurs estimées. Plus la méthode est bonne, plus ces critères doivent être faibles.

En pratique cependant, comme précisé dans la remarque, il est nécessaire de faire les rotations nécessaires sur les matrices $\hat{\mathbf{M}}$ et $\hat{\mathbf{S}}$ pour associer à chaque vecteur estimé le vecteur original correspondant. Afin d'identifier correctement la rotation à effectuer, la méthode choisie est de calculer le critère SAD pour toutes les combinaisons possibles entre les vecteurs de la matrice $\mathbf{M}$ et ceux de la matrice $\hat{\mathbf{M}}$ pour ensuite choisir la rotation qui donne le critère minimal. Cette même rotation est ensuite appliquée aux lignes de la matrice $\hat{\mathbf{S}}$ pour associer chaque abondance estimée à l'abondance qui correspond à la vérité terrain.

TABLE 4.1 – Comparaison entre les méthodes MVC-NMF et PGNMF pour différents types de bruits

	Type de bruit	Bruit Gaussien	Bruit CES 1	Bruit CES 2
AID	MVC-NMF	5.06e-04	0.00350	9.30e-04
	PGNMF	5.30e-04	5.14e-04	5.27e-04
SAD	MVC-NMF	0.521	0.439	2.51
	PGNMF	0.446	0.405	0.454
SID	MVC-NMF	0.768	0.279	27.2
	PGNMF	1.18	0.774	0.916

Évaluation des méthodes existantes

Les résultats des méthodes MVC-NMF et PGNMF serviront d'éléments de comparaison pour les méthodes proposées. Il semble donc nécessaire d'introduire quelques uns de leurs résultats.

Le premier tableau 4.1 montre les résultats obtenus pour la méthode MVC-NMF et la méthode PGNMF sur des signaux gaussiens et CES. Les simulations sont menées sur des images simulées, et les valeurs sont les moyennes sur cinq itérations. Le nombre de pixels simulé est 64×64 et le nombre de bandes spectrales est de 224. Les images sont simulées de la manière suivante : à un bruit gaussien ou CES sont ajoutés 4 vrais spectres hyperspectraux issus de la base [59], avec des RSB de 100 définis de sorte que la puissance du bruit soit égale à $\frac{1}{\text{RSB}^2}$. Le bruit CES 1 fait référence à un bruit CES dont les variables τ_i suivent une loi inverse gamma, de paramètre $\nu = 2$. Le bruit CES 2 est un bruit pour lequel les variables $\sqrt{\tau_i}$ suivent une loi de Student-t, de paramètre $\nu = 3$.

Comme cela pouvait être prévisible, le changement du bruit gaussien en bruit non gaussien induit une baisse des performances de l'algorithme MVC-NMF, mais en ce qui concerne l'algorithme PGNMF, celui-ci semble plus robuste mais donne, pour un bruit gaussien, de moins bons résultats que l'algorithme MVC-NMF. Ce même genre de comparaisons est mené dans l'article [77], pour des simulations différentes. Même si ici, les paramètres ont été choisis pour ressembler au plus aux simulations menées dans cet article [77], les performances de la méthode PGNMF ont été trouvées bien meilleures que dans leurs simulations. Cela dit, certains paramètres tels que la puissance du bruit ou les spectres des *endmembers* ne sont pas donnés, et il semble donc possible que ceux-ci aient une influence non négligeable sur les résultats.

Enfin, le tableau 4.2 montre les mêmes résultats mais pour différentes dimensions, et pour un bruit gaussien. Ce tableau permet de voir l'influence de la dimension sur la performance des algorithmes déjà existants. Les simulations sont menées de la même façon que pour le tableau précédent. La différence entre les deux colonnes correspond aux dimensions des images simulées, $N = 1200$ et $m = 960$ pour la première, soit $c = 0.8$, $N = 4096$ et $m = 224$ pour la seconde, soit $c = 0.05$. Ces résultats permettent de conjecturer que la dimension des images peut avoir une influence sur le critère AID, c'est à dire sur l'estimation de la matrice des abondances.

TABLE 4.2 – Comparaison entre les méthodes MVC-NMF et PGNMF pour différents types de bruits

	Dimensions	$c = 0.8$	$c = 0.05$
AID	MVC-NMF	1.97	5.06e-04
	PGNFM	2.00	5.30e-04
SAD	MVC-NMF	0.356	0.521
	PGNFM	0.616	0.446
SID	MVC-NMF	0.272	0.768
	PGNFM	0.687	1.18

Ainsi, les hypothèses considérées pour cette thèse qui sont relatives à un bruit non gaussien et une grande dimension dégradent certains des résultats données par ces méthodes.

Évaluation des méthodes proposées

Les simulations suivantes permettent de comparer les résultats obtenus avec les méthodes décrites et ceux obtenus avec les méthodes ci-dessus. Elles sont obtenues à l'aide de spectres hyperspectraux issus de [60] qui sont les spectres de l'*Actinolite*, la *Staurolite*, l'*Ilmenite* et le spectre *StonwllPly2grnGrass*. Les dimensions de l'image simulée sont $m = 960$ et $N = 1200$. Le RSB est défini comme suit : chaque spectre est multiplié par une puissance P telle que $P = 10^{\text{RSB}/10}$. Á chaque spectre est ajouté un bruit CES dont la matrice $\mathbf{T}^{1/2}$ contient des données suivant une distribution de Student-t, ce qui, d'après la figure 3.3, donne un bruit fortement non-gaussien. La matrice des abondances est simulée aléatoirement, chaque pixel suivant une loi gaussienne. Sur chacune des figures qui vont suivre est représenté, pour chaque méthode, l'évolution du critère regardé en fonction du RSB. Chaque groupe d'histogrammes correspond à une méthode différente. Les RSB considérés varient de -50 dB à 50 dB (-50, -37, -24, -11, 2, 15, 14, 26, 38, 50)dB et sont représentés par différentes couleurs. Le nombre maximum d'itérations admis pour chaque algorithme est de 205. Pour chaque RSB, le tracé correspond à la moyenne sur 100 itérations. En ce qui concerne le critère SAD, il n'est pas représenté ici car pour la méthode MVC-NMF et pour certains RSB des autres méthodes, il manque certaines données car des itérations ne renvoyaient pas un résultat admissible.

La méthode PGNMF ne donne pas de bons résultats par rapport aux autres méthodes. En revanche, il est intéressant de remarquer que selon le critère évalué, les résultats des différentes méthodes ne sont pas les mêmes. Ainsi, les méthodes proposées donnent des résultats légèrement meilleurs que la méthode MVC-NFM pour le critère AID.

Sur la figure 4.9 sont tracés les mêmes résultats mais cette fois-ci pour des τ_i, $i \in [\![1, N]\!]$ suivant une loi inverse gamma. Les RSB considérés sont alors [-10 dB, 0 dB, 10 dB, 20 dB, 30 dB, 40 dB]. Sur ces figures, la méthode robuste donne de meilleurs résultats.

Pour terminer, avec les mêmes paramètres que pour la figure 4.9, les distances euclidiennes entre les différentes matrices peuvent également être étudiées. La figure 4.10 trace la distance euclidienne entre $\mathbf{S}$ et $\hat{\mathbf{S}}$. Les autres distances ont donné des résultats similaires pour toutes les méthodes.

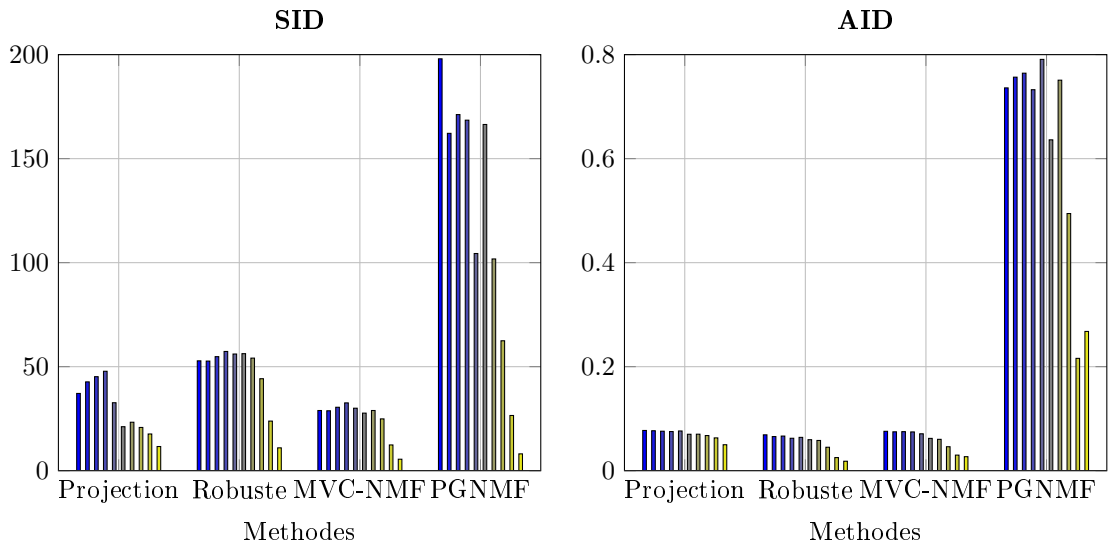

FIGURE 4.8 – Critères SID et SAD pour des images avec des spectres hyperspectraux réels, $m = 960$,$N = 1200$, pour différents RSB et différentes méthodes.

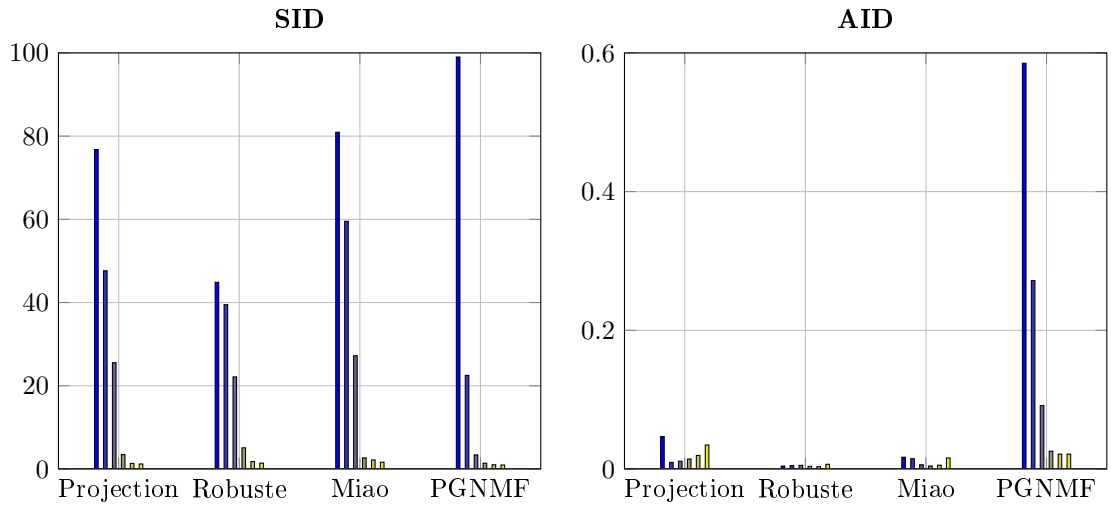

FIGURE 4.9 – Critères SID et AID pour des images avec des spectres hyperspectraux réels, $m = 960$,$N = 1200$, pour différents RSB et différentes méthodes.

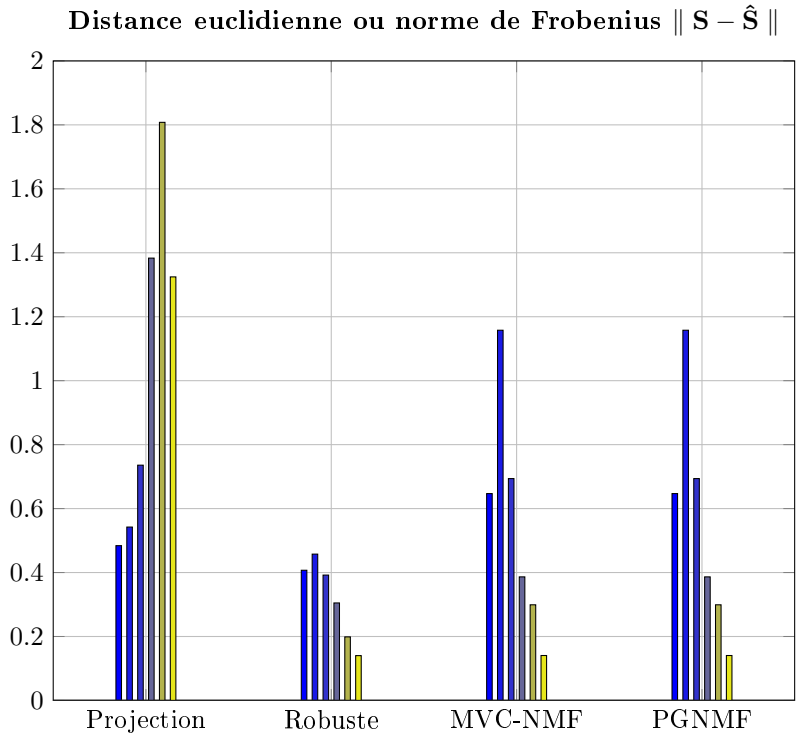

FIGURE 4.10 – Distance euclidienne entre $\mathbf{S}$ et $\hat{\mathbf{S}}$ pour $m = 960$,$N = 1200$ et différentes valeurs de RSB.

Sur cette figure, la méthode robuste est celle qui donne, encore une fois, de meilleurs résultats. Une simulation similaire a été menée pour des variables aléatoires τ_i, $i \in [\![1, N]\!]$ suivant une loi inverse gamma, donnant des résultats similaires.

Sur ces simulations, l'augmentation du RSB n'induit pas forcément une diminution de la valeur des critères. Ce résultat est commenté dans l'article [77], et provient de la structure des méthodes.

Simulations sur images hyperspectrales réelles

Pour les données réelles, le problème rencontré est de trouver une vérité terrain qui correspond bien à un bruit additif plus ou moins homogène, c'est à dire une image pour laquelle la vérité terrain disponible sur la base de donnée correspond bien à la vérité terrain et prend en compte tous les *endmembers*. Par exemple, voici à gauche sur la figure 4.11, pour une image hyperspectrale de la base de donnée [128] représentant une scène urbaine, la 100ième bande spectrale de $\mathbf{Y} - \mathbf{MS}$. Cette image devrait correspondre au bruit puisqu'il s'agit bien ici des vraies matrices $\mathbf{M}$ et $\mathbf{S}$ données comme vérité terrain. Or cette image de gauche de la figure 4.11 n'est pas la figure d'un bruit CES ni même d'un bruit puisqu'il est possible de distinguer en clair des types de sol particuliers. Cette image ressemble plus à la matrice des abondances pour un ou plusieurs *endmembers*. Il semble donc rester de l'information dans ce qui devrait être du bruit. De plus, sur cette figure 4.11 est aussi représentée à droite une bande spectrale de $\mathbf{MS}$ et la même bande spectrale au milieu de l'image projetée sur le sous espace signal. Ces deux images ne se ressemblent pas alors qu'elles devraient être théoriquement similaires.

Ainsi, pour avoir des résultats interprétables, il est nécessaire de trouver des images et leur vérité terrain. L'image Cuprite, issue de la base de données [103] donne de meilleurs

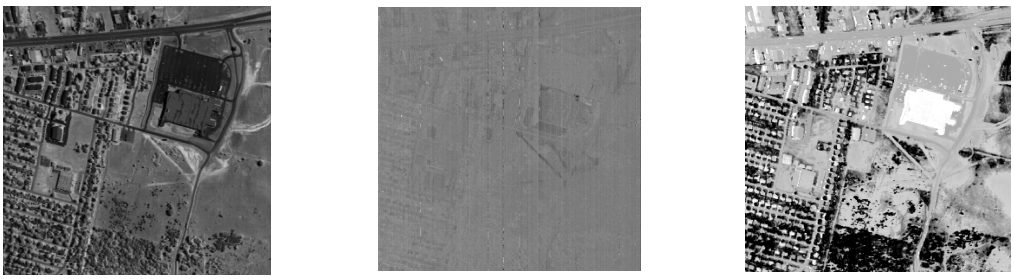

FIGURE 4.11 – Une bande spectrale d'une scène urbaine [128], à gauche $\mathbf{Y} - \mathbf{MS}$, au milieu $\mathbf{Y}$ projeté sur le sous espace signal et à droite $\mathbf{MS}$.

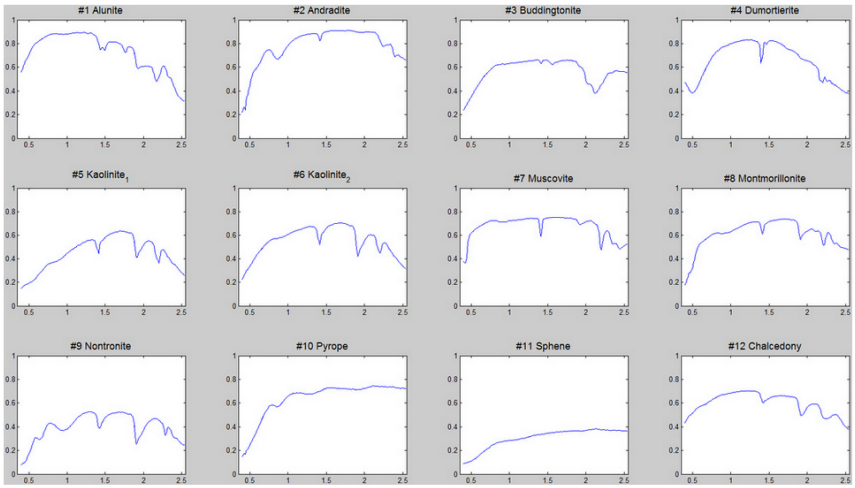

FIGURE 4.12 – Vérité terrain des spectres de l'image Cuprite

résultats, seulement les abondances ne sont pas disponibles sous un autre format qu'une image png. La moyenne de l'image n'est pas systématiquement retirée pour les simulations suivantes puisque retirer la moyenne revient à retirer une partie du signal.
Ainsi donc, sur les images 4.12 et 4.13 sont tracés les vérités terrain, en spectres et abondances. L'image 4.12 est issue de la base de donnée AVIRIS [80].

Les figures suivantes sont les résultats des différentes méthodes. Pour ces figures, le nombre maximal d'itération admis est de 200. Les figures 4.14 et 4.15 représentent les résultats de la méthode faisant appel à une projection des données. La figure 4.15 montre les spectres des différents *endmembers*. En ordonnée se trouvent les réflectances, en abscisse les longueurs d'ondes numérotées, c'est à dire qu'il ne s'agit pas de la valeur de la longueur d'onde à laquelle est photographiée la scène mais seulement de la position dans le vecteur $\mathbf{m}_j, j \in [\![1, p]\!]$ de la valeur de la réflectance correspondante. En effet, sans plus d'a priori sur les images, il n'est pas possible de retrouver la valeur de la longueur d'onde correspondant à telle ou telle valeur de réflectance. Il est possible, sur cette figure, de reconnaître certains spectres : la figure 4.16 montre les vrais spectres de deux *endmembers* et les résultats obtenus. Les deux spectres correspondent assez bien.

Les figures 4.17 et 4.18 présentent les résultats pour la méthode robuste proposée. Les échelles sont les mêmes que pour les figures précédentes.

Les figures 4.19, 4.21 et 4.20 représentent respectivement les résultats en appliquant les

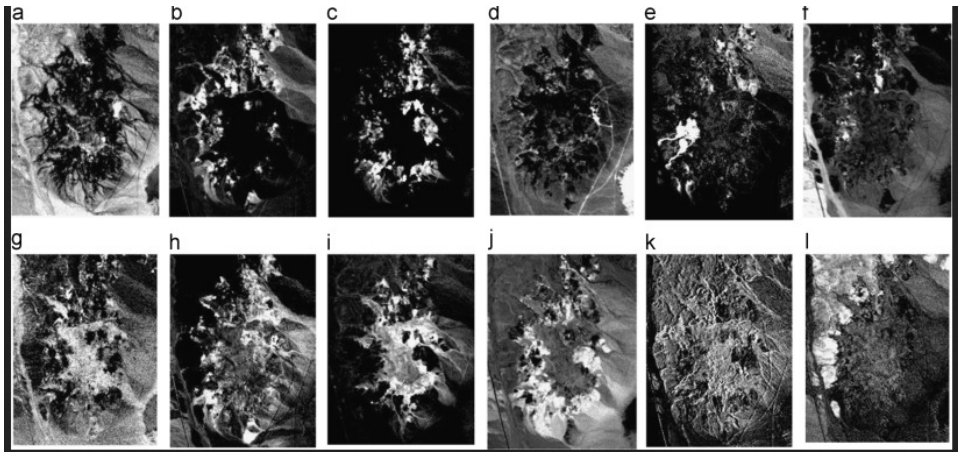

FIGURE 4.13 – Vérité terrain des abondances de l'image Cuprite , chaque image correspondant à un *endmember*.

méthodes MVC-NMF et PGNMF. Sur la figure 4.19 n'est représenté qu'un spectre et les abondances d'un seul *endmember* car les résultats sont les mêmes pour les autres *endmembers*. Enfin, les figures 4.23 et 4.22 représentent les résultats de la méthode MVC-NMF lorsque la contrainte de volume est enlevée, ce qui revient au même que pour les résultats de la méthode NMF.

D'après toutes ces figures, La méthode qui donne les meilleurs résultats est ainsi la méthode qui utilise la projection. La méthode MVC-NFM ne trouve aucun endmember, ni les abondances associées. Quant à la méthode PGNMF, elle permet d'estimer les spectres correspondant à la Kaolinite mais pas les autres spectres. En effet, tous les spectres de la figure 4.20 montrent une bonne correspondance avec celui de la Kaolinite. La méthode NMF, en revanche, donne de bons résultats également, puisqu'elle retrouve différents spectres ressemblant à ceux de la vérité terrain. Ces évaluations restent des évaluations "visuelles", c'est à dire que les spectres obtenus sont comparés à ceux de la vérité terrain non pas à partir de la valeur de la réflectance obtenue pour une longueur d'onde donnée mais par corrélation en mettant au mieux le spectre obtenu à l'échelle de celui de la vérité terrain. En effet, sans plus d'a priori sur l'image, il n'est pas possible de savoir à quelle longueur d'onde correspond telle valeur de réflectance obtenue, qui, elle même, peut être influencée par le RSB de l'image. Il est cependant possible d'évaluer les résultats autrement, comme proposé dans la partie précédente ou encore dans le tableau de celle qui suit.

Conclusion sur les méthodes proposées

Sur l'image hyperspectrale Cuprite, les méthodes qui fonctionnent le mieux ne sont pas celles qui ont montré les meilleures performances sur les images simulées avec les spectres réels. La seconde méthode proposée, qui utilise l'estimation de la matrice $\mathbf{C}$ et projette les données sur l'espace signal, est la méthode qui donne les meilleurs résultats, ainsi que la méthode NMF. La méthode MVC-NMF, pourtant présentée comme une amélioration de la méthode NMF, ne donne pas sur cette image de bons résultats. Il en est de même pour la méthode PGNMF qui ne trouve qu'un seul type d'*endmembers*.

Pour finir, selon le critère d'évaluation choisi, les résultats sont plus ou moins bons selon les méthodes. Par exemple, s'il est montré un intérêt pour la distance euclidienne $\|\mathbf{Y} - \mathbf{M}\,\mathbf{S}\|$, les résultats pour l'image Cuprite sont les suivants :

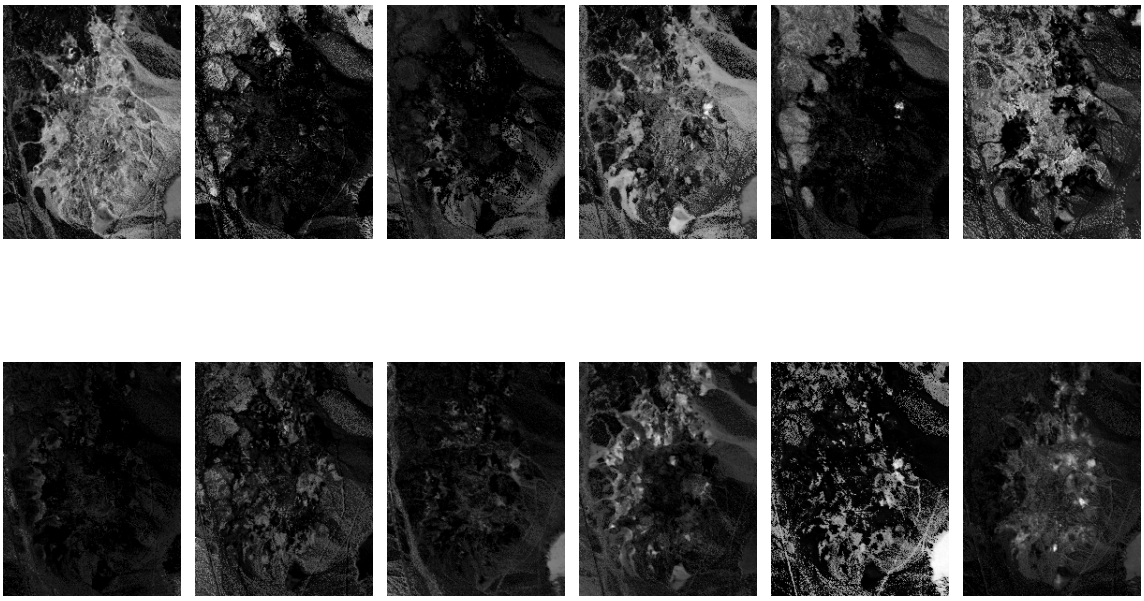

FIGURE 4.14 – Abondances obtenues après application sur les données Cuprite de la méthode projetée.

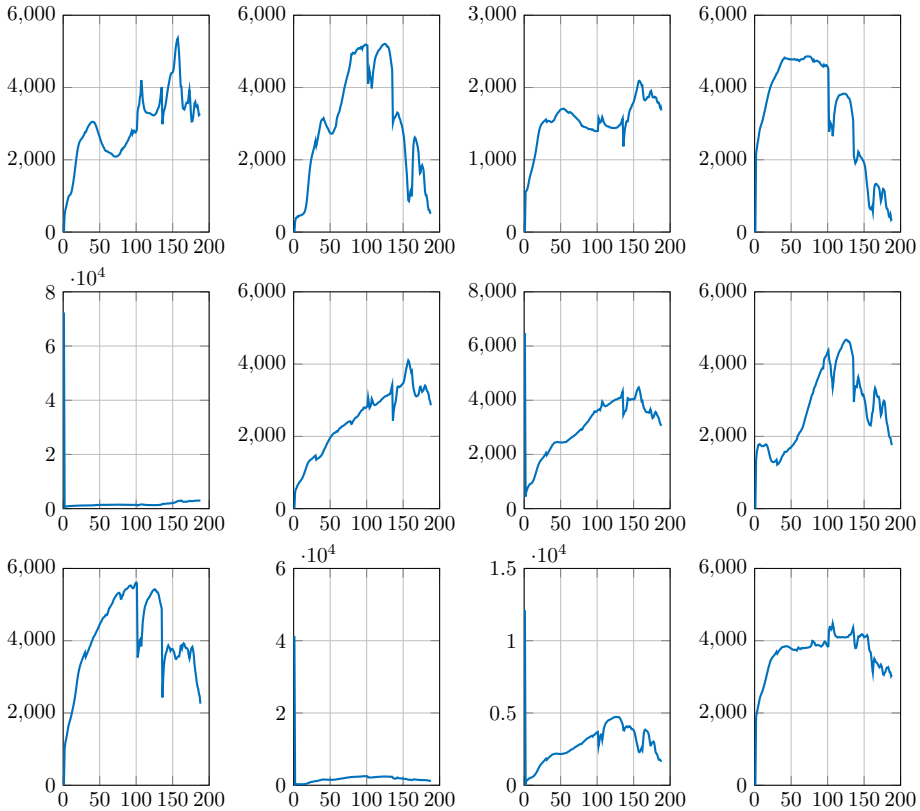

FIGURE 4.15 – Spectres obtenus après application sur les données Cuprite de la méthode projetée.

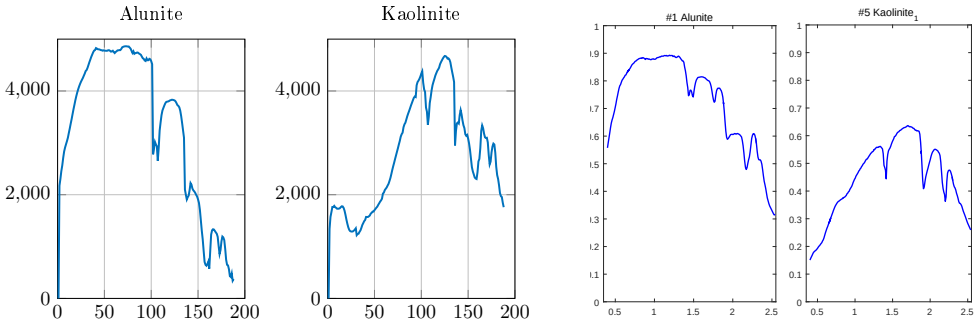

FIGURE 4.16 – Comparaison entre les spectres de l'Alunite et de la Kaolinite issus de la vérité terrain et de la méthode projetée, à gauche selon la méthode projetée proposée, à droite les vérités terrains.

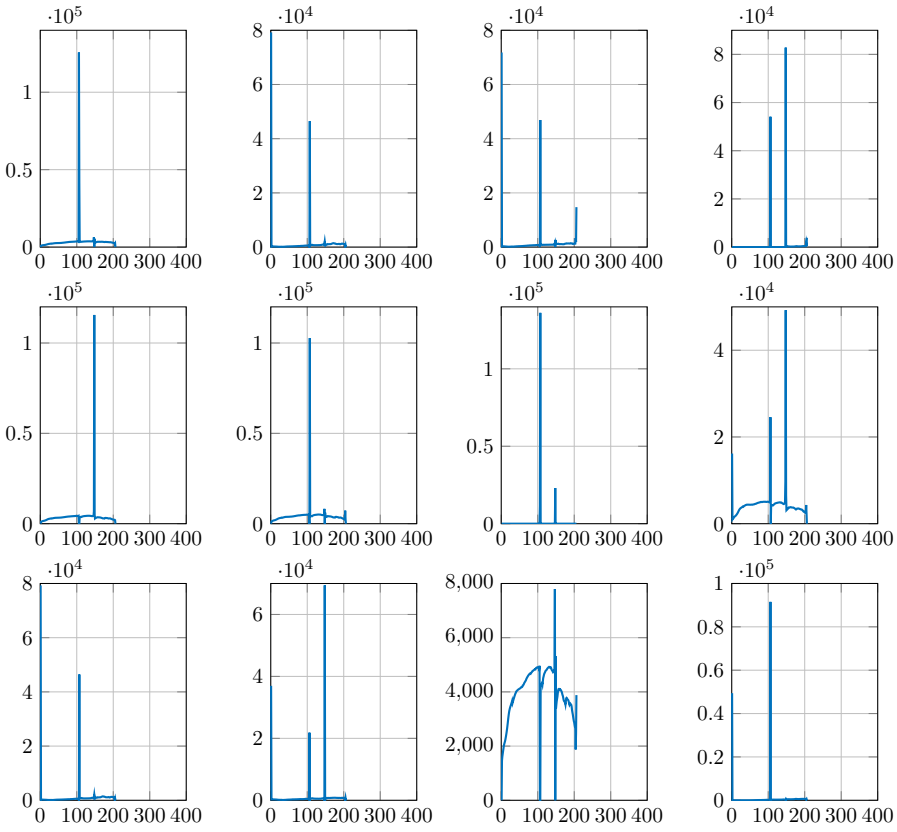

FIGURE 4.17 – Spectres issus de la méthode robuste appliquée sur les données Cuprite. Beaucoup de réflectances sont à zéro.

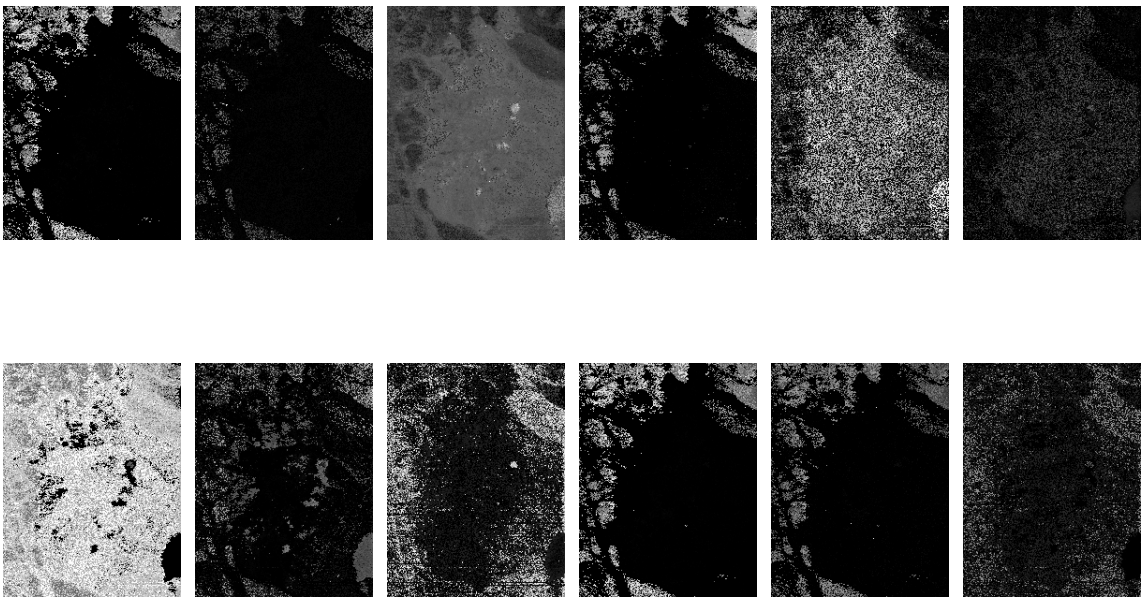

FIGURE 4.18 – Abondances issues de la méthode robuste appliquée sur les données Cuprite.

FIGURE 4.19 – Résultats de la méthode MVC-NMF sur les données Cuprite, résultats similaires pour tous les *endmembers*.

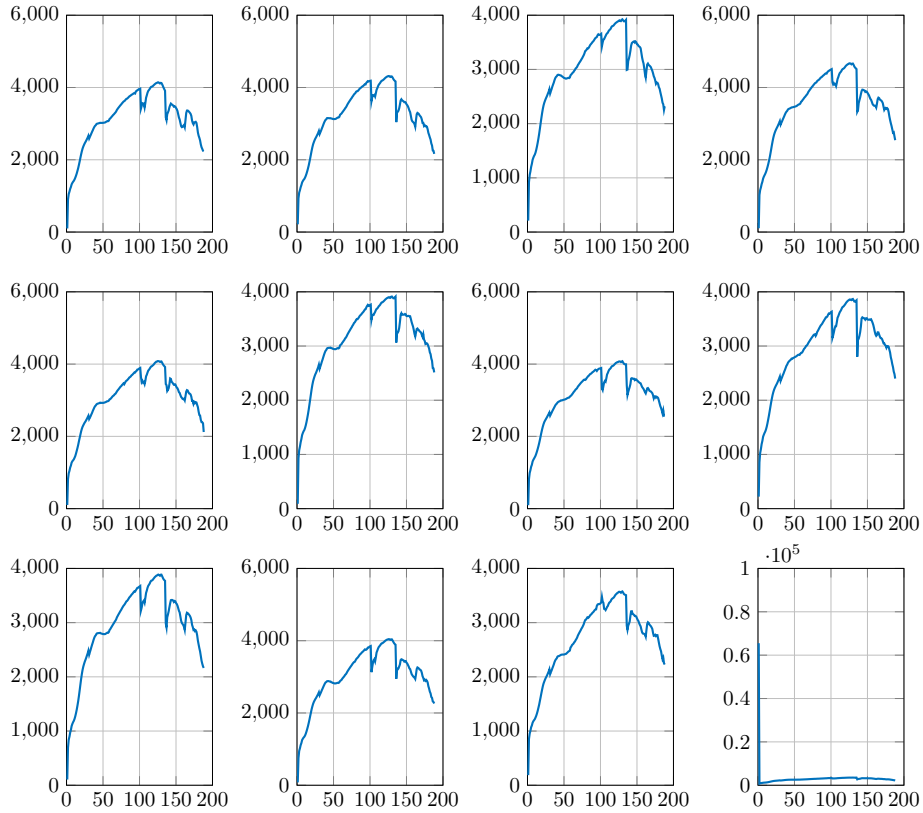

FIGURE 4.20 – Spectres issus de la méthode PGNMF appliquée sur les données Cuprite.

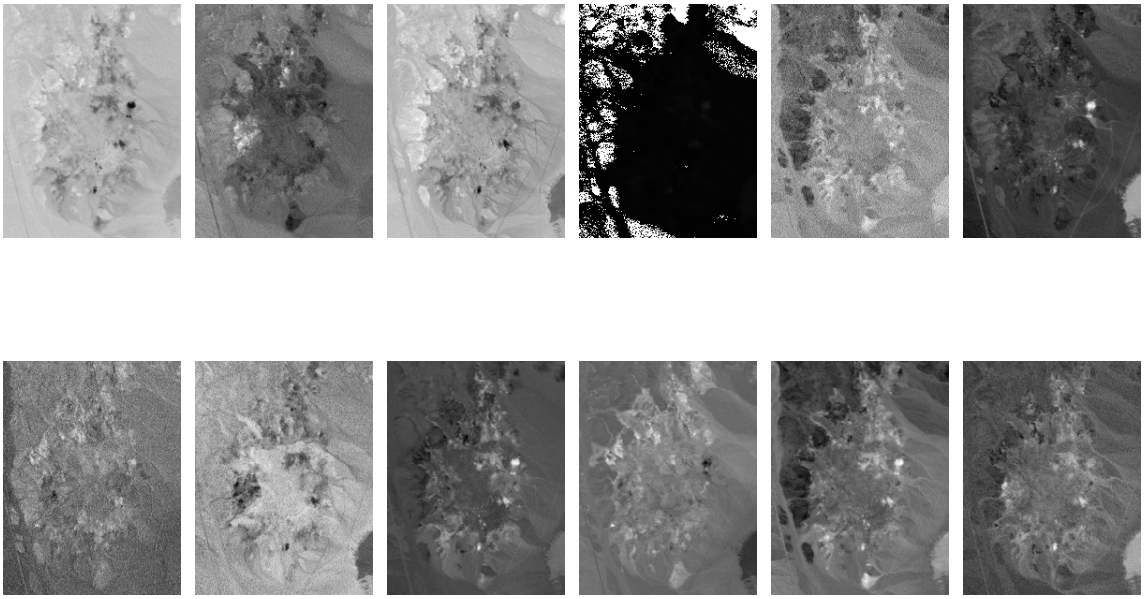

FIGURE 4.21 – Abondances issues de la méthode PGNMF appliquée sur les données Cuprite.

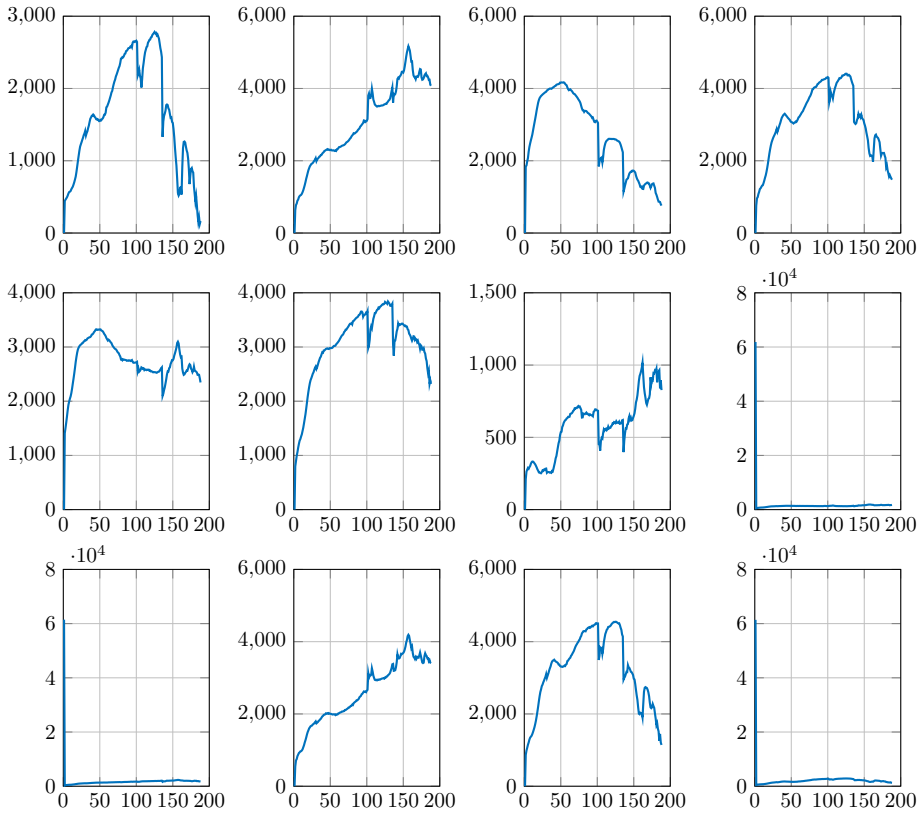

FIGURE 4.22 – Spectres issus de la méthode NMF appliquée sur les données Cuprite.

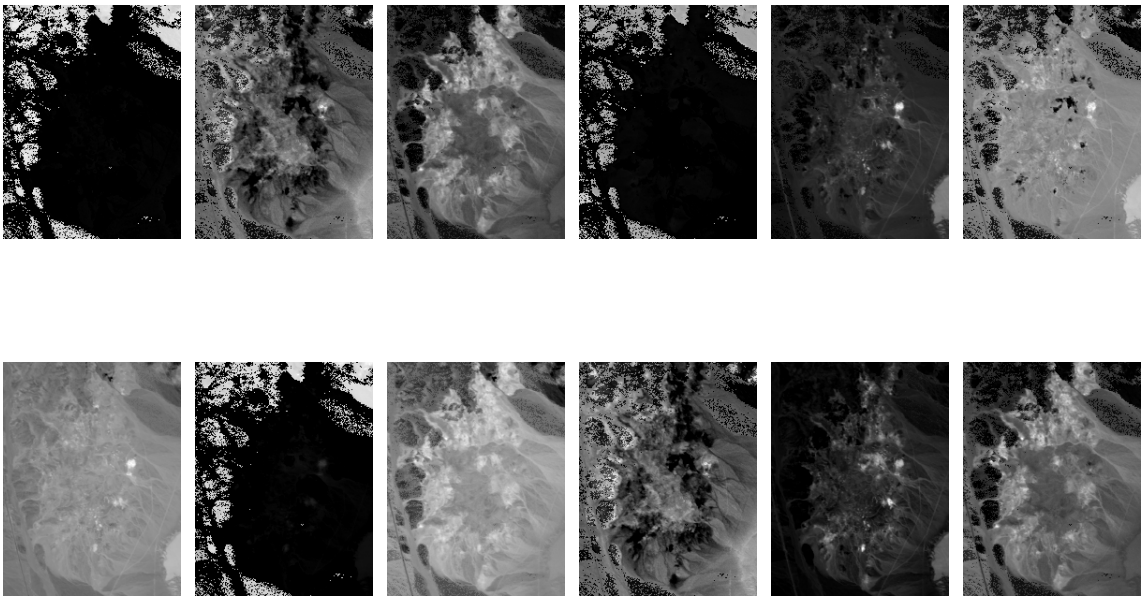

FIGURE 4.23 – Abondances issues de la méthode NMF appliquée sur les données Cuprite.

Méthode	PGNMF	NMF	Méthode Projetée	Méthode Robuste
$\|\mathbf{Y} - \mathbf{MS}\|$	9.7e2	7.78e3	1.9e6	4e4

La meilleure méthode est la méthode PGNFM alors que ce n'est pas cette méthode qui donne de meilleurs résultats pour les autres critères, notamment pour des critères plus "visuels" de la partie précédente. Ainsi, selon les attentes en matière de résultats, le choix de la méthode peut également changer la donne.

Ces simulations permettent de montrer que les méthodes utilisées sont plus ou moins efficaces selon l'image en question. Il semble donc difficile de trouver une méthode universelle qui permette de trouver les spectres des *endmembers* et leurs abondances. Cependant, les deux méthodes proposées semblent améliorer certains résultats dans différents cas. Ce sont donc des méthodes qui peuvent s'appliquer sur un spectre plus large d'images hyperspectrales et donnent, dans certains cas, une vraie amélioration du problème.

4.2 Limite du démélange spectral dans le cas où $p < m$

Ce cas correspond à d'autres cas de figure, où les dimensions de l'image hyperspectrale sont assez proches du nombre d'*endmembers*, par exemple dans le cas d'une image hyperspectrale plutôt petite, ou avec une grande définition, de telle manière que le nombre d'*endmembers* p ne vérifie plus les hypothèses considérées jusqu'ici. Cependant, considérer ce cas est intéressant car la RMT propose des solutions qui peuvent être adaptées au cas du démélange spectral, et d'autant plus qu'il est possible, selon l'image hyperspectrale considérée, de se trouver dans ce cas. Il est donc intéressant de proposer une méthode alternative qui améliore les résultats existants pour le cas où $p < m$. Cette méthode repose sur l'estimation G-MUSIC, comme résumé dans les articles [115] et [75].

4.2.1 Méthode

La méthode ressemble à la méthode projetée, et repose sur l'estimation des valeurs propres et des vecteurs propres de la matrice $\check{\mathbf{C}}$. La suite s'appuie sur les théorèmes 2.3.2 et 1.6.7. Pour rappel, ce théorème donne une estimation des valeurs propres de la matrice $\mathbf{C}$. Ce théorème peut être complété de la manière suivante, avec les mêmes notations que dans le chapitre 1 :

Théorème 4.2.1. *Sous les mêmes hypothèses que pour 2.3.2, si la valeur de* $c = \dfrac{m}{N}$ *permet d'assurer la consistance (voir [75]) les résultats suivants s'appliquent :*

- *soient* $[\check{\mathbf{e}}_1, \dots, \check{\mathbf{e}}_K]$ *les vecteurs propres de la matrice* $\check{\mathbf{C}}$ *associées aux* K *valeurs propres distinctes* $\check{\lambda}_1, \dots, \check{\lambda}_K$,
- *soient* $\mathbf{s}_1$ *et* $\mathbf{s}_2$ *deux vecteurs de taille* $m \times 1$,
- *soit, pour tout* $i \in [\![1, m]\!]$

$$\theta_k(i) = \begin{cases} -\phi_k(i) & \text{si } i \notin \mathcal{M}_k \\ 1 + \psi_k(i) & \text{si } i \in \mathcal{M}_k \end{cases}$$

avec

$$\phi_k(i) = \sum_{r \in \mathcal{M}_k} \left(\frac{\check{\lambda}_r}{\check{\lambda}_k - \check{\lambda}_r} - \frac{\hat{\mu}_r}{\check{\lambda}_k - \hat{\mu}_r} \right)$$

$$\psi_k(i) = \sum_{r \notin \mathcal{M}_k} \left(\frac{\check{\lambda}_r}{\check{\lambda}_k - \check{\lambda}_r} - \frac{\hat{\mu}_r}{\check{\lambda}_k - \hat{\mu}_r} \right),$$

et pour tout r, $\hat{\mu}_r$ les uniques solutions réelles de

$$\frac{1}{m}\sum_{k=1}^{m}\frac{\check{\lambda}_k}{\check{\lambda}_k-\hat{\mu}}=\frac{1}{c}\,, \tag{4.19}$$

alors, pour tout $k\in[\![1,K]\!]$:

$$\check{\eta}_k=\sum_{i=1}^{m}\theta_k(i)\,\mathbf{s}_1^H\,\check{\mathbf{e}}_i\,\check{\mathbf{e}}_i^H\,\mathbf{s}_2\,, \tag{4.20}$$

sont des estimateurs consistants des formes quadratiques $\eta=\mathbf{s}_1^H\,\mathbf{e}_i\,\mathbf{e}_i^H\,\mathbf{s}_2$ caractérisant les vecteurs propres $\mathbf{e}_i$ de $\mathbf{C}$.

Avec ce théorème, il est ainsi possible de modifier légèrement l'étape de projection de l'algorithme projeté, en y ajoutant les coefficients θ_k. La projection ne se fait alors plus sur l'espace signal représenté par les vecteurs propres associés aux plus grandes valeurs propres $\check{\lambda}$ mais sur l'espace caractérisé par les $\check{\eta}$. La suite de la méthode est ainsi la même que pour la méthode projetée.

L'article [115] propose également une amélioration de l'estimation du sous espace signal, en ajoutant des coefficients sur les vecteurs propres.

La première méthode décrite, issue de 1.6.7, a été testée sur des données simulées et donne de meilleurs résultats que la méthode PGNMF. Ces simulations ont été menées pour des signaux de dimensions $m=30$, $N=90$ et $p=24$, pour des raisons de temps de calculs mais aussi pour se positionner dans un cas favorable pour la méthode limite proposée. Les résultats ne sont en revanche pas concluants en ce qui concerne la comparaison avec la méthode projetée sans ces coefficients RMT. Cependant, dans le cas limite où $p\ll m$, les coefficients calculés sont égaux à 1, ce qui signifie que les deux méthodes sont bien équivalentes. La méthode limite change donc le sous-espace de projection uniquement dans le cas $p<m$.

Il semblerait intéressant, pour des travaux futurs, de reconsidérer cette amélioration de la projection en comparant les résultats de [75] et de [115] sur des images hyperspectrales réelles contenant un grand nombre d'*endmembers*, et non plus uniquement sur des signaux qui ne correspondent pas à la réalité physique d'une image hyperspectrale puisque le nombre de bandes et de pixels considérés sont trop faibles.

4.3 Synthèse

Le démélange spectral sans dictionnaire ni données secondaires n'est pas un problème évident. La première étape étant l'estimation du nombre d'*endmembers*, si celle-ci n'est pas effectuée correctement, le démélange spectral s'en ressentira. Grâce aux méthodes proposées dans le chapitre précédent, le démélange spectral a pu être globalement amélioré puisque le nombre d'*endmembers* est mieux estimé. Si ce nombre est supposé connu a priori, les méthodes proposées donnent des résultats plutôt bons, mais n'améliorent les résultats existants que sous certains aspects : la méthode robuste, par exemple, est particulièrement adaptée aux images ayant un bruit CES et donnera pour l'estimation de la matrice des abondances de meilleurs résultats que les autres méthodes. La méthode projetée, quant à elle, donne de meilleurs résultats sur l'image Cuprite. En pratique, pour évaluer les performances des méthodes proposées, et comme l'image hyperspecrale est supposée inconnue, la seule solution semble d'exécuter plusieurs méthodes et de rester critique par rapport aux résultats obtenus en comparant par exemple les résultats des différentes méthodes appliquées. Cependant, les

méthodes proposées donnent les résultats les plus robustes au type d'image hyperspectrale considéré. Ainsi, les appliquer donne déjà une bonne idée des spectres et de leur position dans l'image sans avoir à se soucier du type de l'image étudiée. Dans le cas d'un grand nombre d'*endmembers*, la méthode proposée change l'espace de projection. Il est donc raisonnable d'envisager une amélioration des résultats, puisque les éléments théoriques vont dans cette direction.

Ce chapitre propose donc des méthodes qui donnent des résultats au moins aussi bons que ceux obtenus par les méthodes existantes, mais même en prenant en compte ces dernières, chaque image hyperspectrale pouvant correspondre à un modèle différent, il semble pour le moment difficile de discerner, sans a priori sur l'image, une méthode universelle donnant de meilleurs résultats que toutes les autres méthodes.

Conclusion et Perspectives

Conclusion

Cette thèse a permis de développer des méthodologies permettant d'améliorer à la fois l'estimation du nombre d'endmembers et le démélange spectral pour le traitement des images hyperspectrales mais aussi l'amélioration en général des méthodes d'estimation d'ordre de modèle, problème largement rencontré en finance, radar ou encore traitement d'antenne. Ces méthodes s'appliquent au cas d'un bruit non-gaussien corrélé, pour des données de grande dimension, et permettent une généralisation de techniques déjà existantes comme la minimisation de critères telles que les méthodes AIC ou encore MDL. Ces techniques développées dans cette thèse s'appliquent ainsi aussi aux signaux gaussiens blancs et à ceux corrélés, aux signaux non-gaussien blanc et corrélés. Cet élargissement du cadre applicatif en imagerie hyperspectrale de méthodes déjà existantes s'appuie sur des preuves théoriques développées durant cette thèse.

Le développement de ces nouvelles méthodes a été permis grâce à l'utilisation de la théorie de matrices aléatoires, particulièrement à la page en ces temps où de nombreux problèmes nécessitent le traitement de données de grandes dimensions. Cette théorie n'a pas encore été populairement exploitée dans des domaines où elle est pourtant d'un grand intérêt, comme montré dans cette thèse avec les images hyperspectrales ou avec les données issues du domaine de la finance.

Une des plus grandes contributions de cette thèse est ainsi l'amélioration de l'estimation de la dimension du sous espace signal ou de l'estimation de l'ordre du modèle. Cette étape, indispensable à l'étude des images hyperspectrales, si elle est correctement menée, permet d'amender le démélange spectral en offrant une meilleure estimation du nombre d'*endmembers*. Des méthodes dîtes "classiques" ont été montrées inefficaces et inconsistantes tandis que celles proposées, issues de la théorie des matrices aléatoires, donnent de relativement bons résultats.

Trois nouveaux algorithmes ont été proposés, tous trois reposant sur le même principe :

— blanchiment des données avec une estimation de la matrice de covariance du bruit avec : soit la SCM, soit un M-estimateur, soit l'estimateur de Tyler,

— estimation de la matrice de dispersion avec : soit un M-estimateur soit l'estimateur de Tyler,

— estimation des valeurs propres de cette matrice,

— calcul d'un seuil en fonction de la méthode choisie, celle utilisant un estimateur de Tyler permettant d'utiliser un seuil indépendant des données et donc plus facilement calculable et robuste sans *a priori* sur les données,

— seuillage des valeurs propres et estimation de la dimension du sous espace signal.

Ces trois méthodes se sont montrées meilleures que les méthodes auxquelles elles ont été comparées. Cependant, la dernière méthode proposée, reposant sur l'estimateur de Tyler, est à la fois plus facile d'utilisation mais aussi plus robuste, puisque ne dépend pas d'un hypothétique modèle choisi voire connu pour l'image hyperspectrale étudiée. Cette méthode est donc à utiliser en priorité par rapport aux autres. Cette dernière méthode s'est également montrée particulièrement utile pour l'allocation de portefeuilles en finance. La robustesse de cette méthode est par ailleurs un atout par rapport aux autres, puisque, non seulement adaptée aux données CES, elle peut, par sa forme, s'adapter à d'autres types de bruits.

La seconde partie s'occupe du problème du démélange spectral, qui est un problème assez général d'optimisation et pourrait également être appliqué à d'autres domaines. Cependant, les spécificités des images hyperspectrales, la non-convexité du problème et le manque d'information en font un problème difficile à résoudre.

Deux types de méthodes sont ici proposées, qui utilisent deux idées différentes :

— La première repose sur la projection des données sur le sous espace signal, puisque celui-ci, grâce aux étapes des algorithmes des méthodes précédentes, est a priori connu. Cette projection permet ensuite d'optimiser un problème classique de factorisation matricielle d'une façon plus précise.

— La seconde se fonde sur une approche robuste de l'optimisation et repose sur les poids de Huber. Cette méthode a ses avantages, puisqu'elle est plus robuste que la méthode précédente, mais elle n'utilise pas la connaissance de la matrice de covariance du bruit estimée.

Les méthodes proposées permettent d'avoir de meilleurs résultats selon les critères d'évaluations considérés. Une conclusion a déjà été tirée à ce sujet dans la partie synthèse du dernier chapitre. Ce qui semble important est que les méthodes proposées ne sont pas moins bonnes que les méthodes déjà existantes, et semblent donner des résultats plus constants, c'est à dire plus robustes à l'image hyperspectrale étudiée. Elles donnent même de meilleurs résultats au regard de certains critères. Par contre, aucune des méthodes évaluées ne donne des résultats systématiquement bons ou systématiquement mauvais. Ainsi, même parmi les méthodes déjà existantes qui sont utilisées couramment, il n'est pas possible de les classifier et de donner la meilleure méthode.

Une des difficultés supplémentaires pour l'évaluation du nombre d'*endmembers* et du démélange spectral est la subjectivité de la définition d'un *endmember*. Selon ce qui est recherché, un *endmember* pourra être un regroupement d'*endmembers* ou une partie d'un *endmember*. Ainsi, l'évaluation des méthodes est rendue plus complexe puisque les vérités terrains données avec les images hyperspectrales ne sont pas les mêmes selon ce qui est recherché. Le nombre d'*endmembers* sur une image peut ainsi varier presque du simple au double. Seule la matrice des abondances permet de trancher et de comprendre à quoi correspondent les *endmembers* trouvés, mais une fois encore, la méthode pourra être jugée insuffisante si la précision sur l'image des abondances estimée est trouvée trop grande ou trop faible.

Finalement, un des freins principal de cette thèse aura été l'obtention d'images hyperspectrales et les vérités terrains associées.

L'étude des images hyperspectrales sans aucun *a priori* n'est donc pas un problème simple, mais les résultats prometteurs de cette thèse permettent d'apporter certains outils et solutions pour ce genre de problème.

Perspectives

Les images hyperspectrales pouvant être de tout type, il semblerait intéressant d'élargir le problème à l'étude des images ayant un type de bruit encore plus général ou de type différent. Cette thèse s'est appliquée à généraliser le bruit au bruit de type CES sous certaines hypothèses (matrice de corrélation de Toeplitz par exemple), mais il pourrait être utile de généraliser à tout bruit CES, ce qui semble possible, même si très complexe, à l'aide de nouveaux résultats en théorie des matrices aléatoires.

Plus précisément, une perspective intéressante serait de relier les trois méthodes proposées pour l'estimation du nombre d'*endmembers* : montrer que le calcul du seuil dans le cas d'utilisation de la SCM et de l'estimateur de Tyler correspond à un cas particulier du seuil calculé dans le cas général des M-estimateurs. Les M-estimateurs devant vérifier des hypothèses particulières que ne vérifient ni la SCM ni l'estimateur de Tyler, il faudrait pouvoir relaxer ces hypothèses sans que les conséquences n'aient un impact sur les résultats d'utilisation de ces méthodes. De plus, il serait utile de développer la preuve des derniers algorithmes proposés dans le cas où des *endmembers* sont présents dans la matrice $\mathbf{Y}$.

En ce qui concerne le démélange spectral, des méthodes efficaces dans le cas où $p < m$ sont encore à développer à l'aide de la théorie des matrices aléatoires. Une ébauche a été ici proposée, il faudrait cependant mieux étudier le problème. La variabilité spectrale pourrait par exemple être prise en compte dans les ordres de multiplicités des sous espaces propres.

Il semblerait également intéressant de mener des statistiques sur un grand nombre d'images hyperspectrales pour désigner la méthode qui, en moyenne, donnerait les meilleurs résultats, pour un critère choisi. Cette méthode doit être à la fois robuste aux différents types de données (comme les méthodes proposées) mais aussi aux différents type d'images hyperspectrales (nombre d'*endmembers*, répartition sur l'image entre autres). La principale difficulté sera de regrouper un nombre suffisant d'images hyperspectrales mais surtout avec la vérité terrain correspondante.

Il est également possible d'appliquer la méthode robuste aux données projetées. Les premiers résultats obtenus par des simulations menées durant cette thèse, plutôt moins bons que ceux déjà obtenus avec les autres méthodes, ne sont ainsi pas en faveur de cette méthode. Cependant, des améliorations doivent pouvoir être apportées.

Plus généralement, il n'a été fait ici aucune étude sur le temps de traitement des images hyperspectrales. Certains algorithmes étant plus rapides que d'autres, il serait possible de les classer selon ce critère également. Pour le moment, aucune des méthodes proposées ne permettrait un traitement des images en temps réel, c'est à dire par exemple sur un grand nombre d'images hyperspectrales d'une même scène. Ce problème n'est peut être pas pertinent puisque les images hyperspectrales sont, pour le moment, trop volumineuses pour être prises en grande quantité pour une seule même scène : chaque image est une photographie de la scène et non une suite d'images type vidéo.

Enfin, toutes les méthodes développées ici peuvent trouver de nombreuses applications. Il pourrait être fructueux de les appliquer à d'autres domaines, comme cela à été fait ici pour la finance, par exemple au traitement d'antennes, au traitement de signaux radar, STAP ou encore signaux MIMO.

Annexe A

Généralités

A.1 Définitions

• Dimensions intrinsèque, virtuelle et effective

La dimension intrinsèque est la dimension du sous espace signal ou le nombre de paramètres nécessaires pour prendre en compte toutes les propriétés des données.

La dimension virtuelle est le nombre d'endmembers nécessaire pour permettre le démélange spectral. Elle dépend ainsi de la méthode choisie pour le démélange.

Quant à la dimension effective, c'est la dimension minimale du sous-espace affine permettant de décrire correctement chaque pixel.

• Divergence de Kullback-Leibler

La divergence de Kullback-Leibler entre deux variables aléatoires X et Y se définit de la manière suivante, pour le cas discret :

$$KL(p_X, p_Y) = \sum_{x \in \mathcal{S}} p_X(x) \log\left(\frac{p_X(x)}{p_Y(x)}\right) , \tag{A.1}$$

où x et y sont deux réalisations de X et Y, p_X et p_Y les distributions de probabilités de X et Y, et $\mathcal{S}$ l'ensemble des réalisations possibles qui doit être commun à X et Y.

Remarque : cette divergence est nulle lorsque $P_X = P_Y$.

• Matrice semi-définie positive

Soit une matrice symétrique $\mathbf{A}$. Elle est dite semi-définie positive si et seulement si, pour tout vecteur $\mathbf{x}$ non nul, $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$.

• Modes de convergence

Le mode de convergence le plus utilisé dans cette thèse est la convergence presque sûre notée $p.s.$ c'est à dire :

la suite de variable aléatoires X_n converge presque sûrement vers X si

$$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1 . \tag{A.2}$$

• Simplexe

D'une façon générale, un p-simplexe est un ensemble (ici polyèdre) d'un espace vectoriel de dimension p construit à partir d'un ensemble dont on considère l'enveloppe convexe.

En imagerie hyperspectrale, soit $\mathbf{y} \in \mathbb{R}^m$ tel que $\mathbf{y} = \mathbf{M}\mathbf{s} + \mathbf{b}$, $\mathbf{M} \in \mathbb{R}^{m,p}$, $\mathbf{s} \in \mathbb{R}^p$ tel que $\forall i \in [\![1, p]\!]$, $s_i \geq 0$ et $\sum_{i=1}^{p} s_i = 1$. Alors l'enveloppe convexe des colonnes de $\mathbf{M}$ definit le $p-1$ simplexe de $\mathbb{R}^m$.

A.2 Lemmes et théorèmes

• Règle d'Armijo

Pour un problème d'optimisation sans contrainte $\min_{\mathbf{x}} f(\mathbf{x})$, avec f une fonction régulière, soient $\mathbf{d}_k$ la direction de descente à l'itération k et α_k le pas de descente. Soit $\omega \in]0, 1[$ et $\mathbf{\Delta} f$ le gradient de la fonction f. Alors la règle d'Armijo permet de trouver le pas optimal α_k pour lequel f va forcément décroitre avec la proportion ω choisie. Dans ce cas, l'équation suivante sera vérifiée :

$$f(\mathbf{x}_k + \alpha_k \mathbf{d}_k) \leq f(\mathbf{x}_k) + \omega \langle \mathbf{\Delta} f(\mathbf{x}_k), \mathbf{d}_k \rangle \,, \tag{A.3}$$

la notation $\langle \cdot, \cdot \rangle$ désignant le produit scalaire.

• Inégalité de Markov

Soit z une variable aléatoire réelle positive, définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. Alors, pour tout $a > 0$:

$$\mathbb{P}(z \geq a) \leq \frac{\mathbb{E}(z)}{a} \,. \tag{A.4}$$

• Lemme de Borel Cantelli

Si la somme des probabilités d'une suite $(\mathcal{A}_n)_{n \geq 0}$ d'événement d'un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ est finie, alors la probabilité qu'une infinité d'entre eux se réalisent simultanément est nulle. Autrement écrit : soit $(\mathcal{A}_n)_{n \geq 0}$ une suite d'éléments de $\mathcal{A}$. Si $\sum_{n \geq 0} \mathbb{P}(\mathcal{A}_n) < \infty$ alors $\mathbb{P}(\limsup \mathcal{A}_n) = 0$.

Annexe B

Preuves des théorèmes

B.1 Preuve du Théorème 2.4.1

La preuve de ce théorème s'inspire de [119]. Le théorème ne s'applique rigoureusement que pour $\mathbf{Y} = \mathbf{C}^{1/2}\,\mathbf{X}\,\mathbf{T}^{1/2}$.

Tout d'abord, soit le lemme suivant, qui est le lemme 4.1 de [41] : pour une matrice hermitienne et de Toeplitz $\mathbf{A} = \mathcal{L}\left((a_0, \ldots, a_{m-1})^T\right)$ de taille $m \times m$ avec $\{a_k\}_{k\in[\![0,m-1]\!]}$ des coefficients absolument sommables ($a_{-k} = a_k^*$), Il est possible de définir la fonction $f(.)$ telle que, pour tout $\lambda \in [0, 2\,\pi]$, $f(\lambda) = \sum_{k=1-m}^{m-1} a_k\, e^{i\lambda k}$ et M_f caractérise son sup essentiel :

$$\|\mathbf{A}\| \leq M_f = \sup_{\lambda\in[0,2\pi)} \left| \sum_{k=1-m}^{m-1} a_k\, e^{i\lambda k} \right| . \tag{B.1}$$

Rappel des notations : soient $\mathbf{Y} = \mathbf{C}^{1/2}\,\mathbf{X}\,\mathbf{T}^{1/2}$, et soit $\mathcal{T}$ l'opérateur de Toeplitz. $\boldsymbol{\delta}$ est supposée gaussienne. Alors, sous les hypothèses **1**, **2** (1.4.1), **3** (1.4.2) et **4** (2.4), et comme les matrices $\mathcal{T}\left(\frac{1}{N}\,\mathbf{Y}\,\mathbf{Y}^H\right)$ et $\mathbb{E}[\tau]\,\mathbf{C}$ sont de Toeplitz, d'après (B.1) :

$$\left\| \mathcal{T}\left(\frac{1}{N}\,\mathbf{Y}\,\mathbf{Y}^H\right) - \mathbb{E}[\tau]\,\mathbf{C} \right\| \leq \sup_{\lambda\in[0,2\pi)} |\hat{\gamma}_m(\lambda) - \mathbb{E}[\tau]\,\gamma_m(\lambda)| \ , \tag{B.2}$$

avec $\gamma_m(\lambda) = \sum_{k=1-m}^{m-1} c_{k,m}\, e^{i\,k\,\lambda}$ où $c_{-k} = c_k^\star$ et $\hat{\gamma}_m(\lambda) = \sum_{k=1-m}^{m-1} \check{c}_{k,m}\, e^{i\,k\,\lambda}$ où $\check{c}_{-k} = \check{c}_k^\star$.

Le lemme suivant est également utile pour le reste de la preuve :

Lemme B.1.1. La quantité $\hat{\gamma}_m(\lambda)$ peut s'écrire comme :

$$\hat{\gamma}_m(\lambda) = \mathbf{d}_m^H(\lambda)\,\frac{\mathbf{Y}\,\mathbf{Y}^H}{N}\,\mathbf{d}_m(\lambda)\,, \tag{B.3}$$

avec $\mathbf{d}_m(\lambda) = \frac{1}{\sqrt{m}}\,\left(1, e^{-i\,\lambda}, \ldots, e^{-i\,(m-1)\,\lambda}\right)^T$.

Démonstration.

La preuve s'inspire de celle de l'Annexe A1 de [116]. L'équation (B.3) peut se réécrire de la manière suivante :

$$\begin{aligned}
\mathbf{d}_m^H(\lambda)\,\frac{\mathbf{Y}\,\mathbf{Y}^H}{N}\,\mathbf{d}_m(\lambda) &= \frac{1}{m\,N}\sum_{l,l'=0}^{m-1} e^{-i\,(l'-l)\,\lambda}\,\left[\mathbf{Y}\,\mathbf{Y}^H\right]_{l,l'}\,,\\
\mathbf{d}_m^H(\lambda)\,\frac{\mathbf{Y}\,\mathbf{Y}^H}{N}\,\mathbf{d}_m(\lambda) &= \sum_{k=1-m}^{m-1} e^{-i\,k\,\lambda}\,\frac{1}{m\,N}\sum_{i=0}^{m-1}\sum_{j=0}^{N-1} y_{i,j}\,y^{\star}_{i+k,j}\mathbb{1}_{0\le i+k\le m}\,,\\
\mathbf{d}_m^H(\lambda)\,\frac{\mathbf{Y}\,\mathbf{Y}^H}{N}\,\mathbf{d}_m(\lambda) &= \sum_{k=1-m}^{m-1} \check{c}_k\,e^{-i\,k\,\lambda}\,.
\end{aligned}$$

□

Ainsi, il est possible d'écrire :

$$\begin{aligned}
\hat{\gamma}_m(\lambda) &= \mathbf{d}_m^H(\lambda)\,\frac{\mathbf{Y}\,\mathbf{Y}^H}{N}\,\mathbf{d}_m(\lambda)\,,\\
\hat{\gamma}_m(\lambda) &= \mathbf{d}_m^H(\lambda)\,\frac{\mathbf{C}^{1/2}\,\mathbf{X}\,\mathbf{T}\,\mathbf{X}^H\,\mathbf{C}^{1/2}}{N}\,\mathbf{d}_m(\lambda)\,. \qquad \text{(B.4)}
\end{aligned}$$

Soit :
$\hat{\gamma}_m^{noise}(\lambda) = \mathbf{d}_m^H(\lambda)\,\dfrac{\mathbf{C}^{1/2}\,\mathbf{X}\,\mathbf{T}\,\mathbf{X}^H\,\mathbf{C}^{1/2}}{N}\,\mathbf{d}_m(\lambda).$

Alors (B.2) devient :

$$\left\|\mathcal{T}\left(\frac{1}{N}\,\mathbf{Y}\,\mathbf{Y}^H\right) - \mathbb{E}[\tau]\,\mathbf{C}\right\| \le \sup_{\lambda\in[0,2\pi)} |\hat{\gamma}_m^{noise}(\lambda) - \mathbb{E}[\tau]\,\gamma_m(\lambda)|\,. \qquad \text{(B.5)}$$

$$\begin{aligned}
\left\|\mathcal{T}\left(\frac{1}{N}\mathbf{Y}\,\mathbf{Y}^H\right) - \mathbb{E}[\tau]\,\mathbf{C}\right\| \quad &\le \sup_{\lambda\in[0,2\pi)} \left|\hat{\gamma}_m^{noise}(\lambda) - \mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda)\right]\right|\\
&+ \sup_{\lambda\in[0,2\pi)} \left|\mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda)\right] - \mathbb{E}(\tau)\,\gamma_m(\lambda)\right|.
\end{aligned}$$

La suite de la preuve consiste à analyser chaque terme de (B.6).

B.1.1 Analyse de $\sup\left|\mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda)\right] - \mathbb{E}(\tau)\,\gamma_m(\lambda)\right|$ pour $\lambda \in [0, 2\pi)$

Le lemme suivant permet de développer $\mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda)\right]$:

Lemme B.1.2.

$$\mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda)\right] = \mathbb{E}[\tau]\,\mathbf{d}_m^H(\lambda)\,\mathbf{C}\,\mathbf{d}_m(\lambda) = \mathbb{E}[\tau]\,\gamma_m(\lambda)\,. \qquad \text{(B.6)}$$

Démonstration.

L'équation (B.4) donne : $\mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda)\right] = \mathbf{d}_m^H(\lambda)\,\mathbb{E}\left[\dfrac{\mathbf{C}^{1/2}\mathbf{X}\,\mathbf{T}\,\mathbf{X}^H\mathbf{C}^{1/2}}{N}\right]\,\mathbf{d}_m(\lambda).$

Soient $\mathbf{V} = \mathbf{C}^{1/2}\mathbf{X}$ et $(\mathbf{V})_{i,j} = v_{i,j}$. Alors, $\mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda)\right] = \mathbf{d}_m^H(\lambda)\,\mathbb{E}\left[\dfrac{\mathbf{V}\,\mathbf{T}\,\mathbf{V}^H}{N}\right]\,\mathbf{d}_m(\lambda).$

De plus, comme $\left(\mathbb{E}\left[\mathbf{V}\,\mathbf{T}\,\mathbf{V}^H\right]\right)_{ij} = \sum_{k=1}^{N}\mathbb{E}[\tau]\,\mathbb{E}\left[v_{j,k}^*\,v_{ik}\right]$ et $c_{k'} = \mathbb{E}\left[v_{p,n}\,v_{p+k',n}^*\right]$, alors $\left(\mathbb{E}\left[\mathbf{V}\,\mathbf{T}\,\mathbf{V}^H\right]\right)_{ij} = \sum_{k=1}^{N}\mathbb{E}[\tau]\,c_{j-i} = N\,\mathbb{E}[\tau]\,c_{j-i}$.

Et ceci mène à :

$$\mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda)\right] = \frac{N\,\mathbb{E}(\tau)}{N}\,\mathbf{d}_m^H(\lambda)\,\mathbf{C}\,\mathbf{d}_m(\lambda) = \mathbb{E}[\tau]\,\gamma_m(\lambda)\,.$$

□

Ainsi, le second terme de (B.6) se développe de la manière suivante :

$$\begin{aligned}&\sup_{\lambda\in[0,2\pi)} \left|\mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda)\right] - \mathbb{E}(\tau)\,\gamma_m(\lambda)\right| \\ &= \sup_{\lambda\in[0,2\pi)} |\,\mathbb{E}(\tau)\,\gamma_m(\lambda) - \mathbb{E}(\tau)\,\gamma_m(\lambda)\,| = 0\,.\end{aligned}$$

Ce terme est donc égal à zero.

B.1.2 Analyse de $\sup|\hat{\gamma}_m^{noise}(\lambda) - \mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda)\right]|$ pour $\lambda \in [0, 2\pi)$

Pour l'analyse de ce terme, plusieurs méthodes sont envisageables : la méthode des moments ou une méthode utilisant des inégalités de concentration. La méthode des moments utilise le lemme de Borel Cantelli et nécessite des calculs de vitesse de convergence. Or, pour avoir une bonne vitesse de convergence, des calculs de moments d'ordre élevé sont nécessaires. La méthode retenue est ainsi la même que celle de [116], c'est à dire une méthode utilisant des inégalités de concentration. Elle consiste à montrer, pour un $\lambda_i \in [0, 2\pi)$ et un réel $x > 0$, que $\mathbb{P}\left[\left|\hat{\gamma}_m^{noise}(\lambda_i) - \mathbb{E}\left[\gamma_m^{noise}(\lambda_i)\right]\right| > x\right] \to 0$. Par la suite, il restera à montrer que $\mathbb{P}\left[\sup_{\lambda\in[\lambda_i,\lambda_{i+1})} \left|\gamma_m^{noise}(\lambda) - \gamma_m^{noise}(\lambda_i)\right| > x\right] \to 0$ et que $\mathbb{P}\left[\sup_{\lambda\in[\lambda_i,\lambda_{i+1})} \left|\mathbb{E}\gamma_m^{noise}(\lambda_i) - \mathbb{E}\gamma_m^{noise}(\lambda)\right| > x\right] \to 0$ pour obtenir la convergence du terme étudié vers zero.

Soit $\lfloor\cdot\rfloor$ la fonction qui renvoie la partie entière d'un réel, soit un $\beta > 2$ et soient $\mathcal{I} = \left[0, \ldots, \lfloor N^\beta\rfloor - 1\right]$, $\lambda_i = 2\,\pi\,\frac{i}{\lfloor N^\beta\rfloor}$, $i \in \mathcal{I}$. Alors :

$$\begin{aligned}\sup_{\lambda\in[0,2\pi)} \left|\hat{\gamma}_m^{noise}(\lambda) - \mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda)\right]\right| &\leqslant \max_{i\in\mathcal{I}} \sup_{\lambda\in[\lambda_i\lambda_{i+1}]} \left|\hat{\gamma}_m^{noise}(\lambda) - \hat{\gamma}_m^{noise}(\lambda_i)\right| \\ &\quad + \max_{i\in\mathcal{I}} \left|\hat{\gamma}_m^{noise}(\lambda_i) - \mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda_i)\right]\right| \\ &\quad + \max_{i\in\mathcal{I}} \sup_{\lambda\in[\lambda_i\lambda_{i+1}]} \left|\mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda_i)\right] - \mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda)\right]\right|, \\ &\triangleq \chi_1 + \chi_2 + \chi_3\,. \qquad \text{(B.7)}\end{aligned}$$

Le paramètre β prend ici toute son importance : si plus de λ_i sont pris en compte, soit si le paramètre β est grand, la convergence de $\sup_{\lambda\in[\lambda_i\lambda_{i+1}]} \left|\hat{\gamma}_m^{noise}(\lambda) - \hat{\gamma}_m^{noise}(\lambda_i)\right|$ vers zero sera plus rapide (par continuité, comme montré par la suite) mais celle de $\max_{i\in\mathcal{I}} \left|\hat{\gamma}_m^{noise}(\lambda_i) - \mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda_i)\right]\right|$ sera moins rapide. Un bon compromis est donc à trouver sur ce paramètre β.

L'idée de la preuve dans l'article [116] est de donner des inégalités de concentration pour les deux termes χ_1 et χ_2 qui sont bien des termes aléatoires, et une borne sur χ_3.

Les deux différences avec l'article [116] sont la présence de la matrice $\mathbf{T}$ dans $\hat{\gamma}_m^{noise}(\lambda)$ et la corrélation à gauche du bruit.

Soit $\|\gamma\|_\infty$ la norme supérieure de la fonction $\gamma : \lambda \longrightarrow \sum_{k=-\infty}^{\infty} c_k\, e^{-ik\lambda}$ pour tout $\lambda \in [0\,2\pi)$. Alors la convergence du premier terme χ_1 est montrée dans le lemme suivant.

Lemme B.1.3. Une constante $A > 0$ peut être trouvée, telle que, pour tout $x > 0$ et pour tout N suffisamment grand,

$$\mathbb{P}\left[\chi_1 > x\right] \leq \exp\left(-\frac{c\,N^2}{a_N}\left(\frac{x\,N^{\beta-2}}{A\,\|\gamma\|_\infty} - \log\left(\frac{x\,N^{\beta-2}}{A\,\|\gamma\|_\infty}\right) - 1\right)\right),$$

presque sûrement, $a_N \in \mathbb{R}^+$ tel que $\frac{-cN^2}{a_N} \underset{N\to\infty}{\to} 0$ en probabilité.

Démonstration.

Comme déjà mentionné, la preuve suit celle de [116] sauf sur deux points : la présence de la matrice $\mathbf{T}$ et la corrélation à gauche du bruit au lieu d'une corrélation à droite.

Pour N fixé, l'inégalité suivante existe : $\left\|\mathbf{V}_N\,\mathbf{T}\,\mathbf{V}_N^H\right\| \leqslant \|\mathbf{T}\|_\infty \left\|\mathbf{V}_N\,\mathbf{V}_N^H\right\| \leq \|\mathbf{T}\|_\infty \|\mathbf{C}\| \left\|\mathbf{X}\,\mathbf{X}^H\right\|$ et permet d'écrire :

$$\left|\hat{\gamma}_m^{noise}(\lambda) - \hat{\gamma}_m^{noise}(\lambda_i)\right| = \left|\mathbf{d}_m^H(\lambda)\,\frac{\mathbf{V}\,\mathbf{T}\,\mathbf{V}^H}{N}\,\mathbf{d}_m(\lambda) - \mathbf{d}_m^H(\lambda_i)\,\frac{\mathbf{V}\,\mathbf{T}\,\mathbf{V}^H}{N}\,\mathbf{d}_m(\lambda_i)\right|,$$

$$\text{soit :}\quad \left|\hat{\gamma}_m^{noise}(\lambda) - \hat{\gamma}_m^{noise}(\lambda_i)\right| \leq \frac{2}{N}\,\left|\mathbf{d}_m(\lambda) - \mathbf{d}_m(\lambda_i)\right|\,\|\mathbf{C}\|\,\|\mathbf{T}\|_\infty \left\|\mathbf{X}\,\mathbf{X}^H\right\|.$$

Une fois cette inégalité établie, la suite de la preuve est la même que celle du Lemme 4 de l'article [116] en remplaçant c par $\frac{c}{\|\mathbf{T}\|_\infty}$ dans le terme de l'exponentielle, c'est à dire :

$$\mathbb{P}\left[\chi_1 > x|\,\|\mathbf{T}\|_\infty\right] \leq \exp\left(-\frac{c\,N^2}{\|\mathbf{T}\|_\infty}\left(\frac{x\,N^{\beta-2}}{A\,\|\gamma\|_\infty} - \log\left(\frac{x\,N^{\beta-2}}{A\,\|\gamma\|_\infty}\right) - 1\right)\right).$$

Quant à la corrélation à gauche, elle n'a pas de conséquence pour la preuve.

Pour N non fixé, il est nécessaire de rajouter une condition supplémentaire sur la distribution des τ, à savoir qu'ils suivent une loi parmi les suivantes : gaussienne, exponentielle, gamma, Student, de Pareto ou encore Weibull, c'est à dire une distribution à queue plus ou moins lourde. Cette condition supplémentaire n'est donc pas très restrictive et pourrait même s'adapter à d'autres lois avec d'autres travaux de recherche.

Dans ce cas, la théorie des distributions des valeurs extrêmes (Generalized Extreme Value Distribution) [44] donne alors la distribution de la plus grande valeur propre de $\mathbf{T}$. L'équation suivante est alors vraie, pour tout entier $\alpha > 1$:

$$\mathbb{P}\left(\|\mathbf{T}\|_\infty > N^\alpha\right) = 1 - \exp\left(-\left(1 + \zeta\,N^\alpha\right)^{-1/\zeta}\right)\ \text{si}\ \zeta > 0\,, \tag{B.8}$$

$$\text{ou}\ \mathbb{P}\left(\|\mathbf{T}\|_\infty > N^\alpha\right) = 1 - \exp\left(-\exp\left(-N^\alpha\right)\right)\ \text{si}\ \zeta = 0\,, \tag{B.9}$$

avec ζ un paramètre dépendant de la loi des τ, c'est à dire positif pour les lois mentionnées ci-dessus. Alors, pour tout $i \in [\![0, m-1]\!]$, $\mathbb{P}\left(\max(\lambda_i(\mathbf{T})) > N^\alpha\right) \underset{N\to\infty}{\to} 0$ pour tout $\zeta \geq 0$.

Il est alors toujours possible de trouver un réel a_N tel que $\max(\lambda_i(\mathbf{T})) < a_N$ avec $\frac{-c\,N^2}{a_N} \underset{N\to\infty}{\to} 0$ en probabilité. □

La convergence du second terme χ_2 est démontrée dans le lemme suivant.

Lemme B.1.4. Pour tout N, il existe $a_N \in \mathbb{R}^+$ avec $\frac{-c\,N}{a_N} \underset{N\to\infty}{\to} 0$ en probabilité et tel que :

$$\mathbb{P}\left[\chi_2 > x\right] \leq 2\,N^\beta \exp\left(-\frac{c\,N}{a_N}\left(\frac{x}{\|\gamma\|_\infty} - \log\left(\frac{x}{\|\gamma\|_\infty} + 1\right)\right)\right).$$

Démonstration.

La preuve est la même que celle du lemme B.1.3 ou que du lemme 5 de [116], avec $\frac{c}{a_N}$ au numérateur. □

La convergence de χ_3 est donnée par le lemme suivant :

Lemme B.1.5. χ_3 vérifie l'équation suivante : $\chi_3 \leq A \left\|\gamma\right\|_\infty N^{-\beta+1}$.

Démonstration.

Une fois encore, la preuve est presque identique à celle du lemme 6 de l'article [116], an ajoutant le terme $\frac{c}{\|\mathbf{T}\|_\infty}$ au numérateur. □

Ces inégalités prouvent ainsi que

$$\forall x \in \mathbb{R}^{+*},\ \mathbb{P}\left[\sup_{\lambda\in[0,2\pi)} \left|\hat{\gamma}_m^{noise}(\lambda) - \mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda)\right]\right| > x\right] \xrightarrow{p.s.} 0.$$

Ainsi, comme le terme de droite de (B.5) tend vers zero lorsque N tend vers l'infini, à un facteur près, ceci conclut la preuve de ce théorème.

B.2 Preuve du Théorème 2.4.2

Comme la preuve suit le même cheminement pour $\check{\mathbf{\Sigma}}_{SCM}$ et $\check{\mathbf{\Sigma}}_{FP}$, la notation $\check{\mathbf{\Sigma}}$ sera adoptée pour l'une ou l'autre de ces matrices, et la preuve n'est écrite qu'une seule fois. D'après les équations (2.10) et (2.33), avec $\check{\mathbf{y}}_{wi} = \check{\mathbf{C}}^{-1/2}\,\mathbf{y}_i$, $\check{\mathbf{\Sigma}}$ est l'unique solution de :

$$\mathbf{\Sigma} = \frac{1}{N}\sum_{i=0}^{N-1} u\left(\frac{1}{m}\,\mathbf{y}_i^H\,\check{\mathbf{C}}^{-1/2}\,\mathbf{\Sigma}^{-1}\,\check{\mathbf{C}}^{-1/2}\,\mathbf{y}_i\right) \times \check{\mathbf{C}}^{-1/2}\,\mathbf{y}_i\,\mathbf{y}_i^H\,\check{\mathbf{C}}^{-1/2}.$$

En réécrivant cette équation, avec , cette fois-ci, les variables $\{\check{\mathbf{y}}_{wi}\}_i$:

$$\mathbf{C}^{-1/2}\check{\mathbf{C}}^{1/2}\,\mathbf{\Sigma}\,\check{\mathbf{C}}^{1/2}\,\mathbf{C}^{-1/2} = \frac{1}{N}\sum_{i=0}^{N-1} u\left(\frac{1}{m}\,\check{\mathbf{y}}_{wi}^H\,\left(\mathbf{C}^{-1/2}\,\check{\mathbf{C}}^{1/2}\,\mathbf{\Sigma}\,\check{\mathbf{C}}^{1/2}\,\mathbf{C}^{-1/2}\right)^{-1}\,\check{\mathbf{y}}_{wi}\right) \times \check{\mathbf{y}}_{wi}\,\check{\mathbf{y}}_{wi}^H.$$

Alors, la relation suivante entre $\check{\mathbf{\Sigma}}$ et $\hat{\mathbf{\Sigma}}$ est établie :

$$\check{\mathbf{\Sigma}} = \check{\mathbf{C}}^{-1/2}\,\mathbf{C}^{1/2}\,\hat{\mathbf{\Sigma}}\,\mathbf{C}^{1/2}\,\check{\mathbf{C}}^{-1/2}. \tag{B.10}$$

Ainsi, les équations (2.27) et (2.34) peuvent se réécrire :

$$\left\|\check{\mathbf{\Sigma}} - \hat{\mathbf{S}}\right\| \leq \left\|\check{\mathbf{\Sigma}} - \hat{\mathbf{\Sigma}}\right\| + \left\|\hat{\mathbf{\Sigma}} - \hat{\mathbf{S}}\right\|. \tag{B.11}$$

En ce qui concerne le second terme de (B.11), l'article [29] prouve que la matrice $\hat{\mathbf{S}}$ définie par (2.13) est telle que :

$$\left\|\hat{\mathbf{\Sigma}} - \hat{\mathbf{S}}\right\| \xrightarrow{p.s.} 0. \tag{B.12}$$

L'équation (B.10) permets de réécrire le premier terme de droite de l'équation (B.11) de la manière suivante :

$$\left\|\check{\mathbf{\Sigma}} - \hat{\mathbf{\Sigma}}\right\| \leq \left\|\check{\mathbf{C}}^{-1/2}\,\mathbf{C}^{1/2}\,\hat{\mathbf{\Sigma}}\,\mathbf{C}^{1/2}\,\check{\mathbf{C}}^{-1/2} - \hat{\mathbf{\Sigma}}\,\mathbf{C}^{1/2}\,\check{\mathbf{C}}^{-1/2}\right\| + \left\|\hat{\mathbf{\Sigma}}\,\mathbf{C}^{1/2}\,\check{\mathbf{C}}^{-1/2} - \hat{\mathbf{\Sigma}}\right\|.$$

Puis, après factorisation à droite et à gauche, comme $\left\|\check{\mathbf{\Sigma}}\right\|$ est *p.s.* bornée [29], ceci donne :

$$\left\|\check{\mathbf{\Sigma}} - \hat{\mathbf{\Sigma}}\right\| \leq \left\|\check{\mathbf{C}}^{-1/2}\,\mathbf{C}^{1/2} - \mathbf{I}_m\right\| \left\|\hat{\mathbf{\Sigma}}\right\| \left(\left\|\mathbf{C}^{1/2}\,\check{\mathbf{C}}^{-1/2}\right\| + 1\right).$$

Comme le support de $\|\mathbf{C}\|$ est borné, $\|\check{\mathbf{C}}\|$ est aussi borné puisque le support de la distribution de ses valeurs propres converge presque sûrement vers la vraie distribution. De plus, les Théorèmes 2.4.2 et 2.4.4 ont montré la consistance de $\left\|\mathbf{C} - \check{\mathbf{C}}\right\| \xrightarrow{p.s.} 0$.

Ceci termine la preuve.

B.3 Preuve du Théorème 2.4.3

La preuve suit la même idée que la preuve du Théorème 2.4.1. Avec la notation suivante : $\check{\mathbf{C}}_{FP} = \mathcal{T}(\hat{\mathbf{C}}_{FP})$ où $\mathcal{T}$ est l'opérateur de Toeplitz défini dans l'introduction, l'équation à démontrer devient :

$$\left\| \mathcal{T}(\hat{\mathbf{C}}_{FP}) - \mathbb{E}\left[v(\tau\xi)\,\tau\right]\,\mathbf{C} \right\| \xrightarrow{p.s.} 0\,. \tag{B.13}$$

Cette équation peut être découpée en deux parties :

$$\left\| \mathcal{T}\left(\hat{\mathbf{C}}_{FP}\right) - \mathbb{E}\left[v(\tau\xi)\,\tau\right]\,\mathbf{C} \right\| \leq \left\| \mathcal{T}\left(\hat{\mathbf{C}}_{FP} - \hat{\mathbf{S}}\right) \right\| + \left\| \mathcal{T}\left(\hat{\mathbf{S}}\right) - \mathbb{E}\left[v(\tau\,\xi)\,\tau\right]\,\mathbf{C} \right\|.$$

Et en considérant les notations suivantes :

- $\hat{\mathbf{S}}$ la matrice telle que $\left\| \check{\mathbf{\Sigma}} - \hat{\mathbf{S}} \right\| \xrightarrow{p.s.} 0$. Pour rappel, $\hat{\mathbf{S}}$ est la matrice définie par :

$$\hat{\mathbf{S}} = \frac{1}{N}\sum_{i=1}^{N} v\left(\tau_i\,\xi\right)\,\mathbf{y}_{wi}\,\mathbf{y}_{wi}^{H},$$

  où chaque ξ est l'unique solution (si elle existe) de :

$$1 = \frac{1}{N}\sum_{i=1}^{N} \frac{\psi(\tau_i\,\xi)}{1 + c\,\psi(\tau_i\,\xi)},$$

  où $g : x \mapsto \dfrac{x}{1 - c\,\phi(x)}$, $v : x \mapsto u\,o\,g^{-1}(x)$ et $\psi : x \mapsto x\,v(x)$. Enfin, pour la définition de $\hat{\mathbf{S}}$, $\check{\mathbf{\Sigma}}$ représente indifféremment $\check{\mathbf{\Sigma}}_{SCM}$ ou $\check{\mathbf{\Sigma}}_{FP}$.
- Pour toute matrice $\mathbf{A} = \mathcal{T}\left((a_0, \ldots, a_{m-1})^T\right)$ de Toeplitz ($a_{-k} = a_k^*$), alors $\hat{\gamma}^{\mathbf{A}}(\lambda)$ sera défini comme sa densité spectrale estimée :

$$\gamma^{\mathbf{A}}(\lambda) \triangleq \sum_{k=1-m}^{m-1} a_k\,e^{i\,k\,\lambda}\,.$$

Afin de prouver la consistance, il est possible de séparer $\left\| \mathcal{T}\left(\hat{\mathbf{C}}_{FP}\right) - \mathbb{E}[v(\tau\,\xi)\,\tau]\,\mathbf{C} \right\|$ en deux parties. Comme les matrices $\mathcal{T}\left(\hat{\mathbf{C}}_{FP}\right)$ et $\mathbf{C}$ sont de Toeplitz, l'équation donnée par (B.1) introduite pour la preuve du théorème 2.4.1 au début de cette annexe permet d'écrire :

$$\begin{aligned} \left\| \mathcal{T}\left(\hat{\mathbf{C}}_{FP}\right) - \mathbb{E}[v(\tau\,\xi)\,\tau]\,\mathbf{C} \right\| &\leq \sup_{\lambda\in[0,2\,\pi)} \left| \hat{\gamma}^{\hat{\mathbf{S}}}(\lambda) - \gamma^{\mathbb{E}[v(\tau\,\xi)\,\tau]\,\mathbf{C}}(\lambda) \right| + \left\| \mathcal{T}\left(\hat{\mathbf{C}}_{FP} - \hat{\mathbf{S}}\right) \right\|, \\ &\leq \chi_1 + \chi_2, \end{aligned} \tag{B.14}$$

Avec : $\chi_1 = \sup_{\lambda\in[0,2\,\pi)} \left| \hat{\gamma}^{\hat{\mathbf{S}}}(\lambda) - \gamma^{\mathbb{E}[v(\tau\,\xi)\,\tau]\,\mathbf{C}}(\lambda) \right|$ et $\chi_2 = \left\| \mathcal{T}(\hat{\mathbf{C}}_{FP} - \hat{\mathbf{S}}) \right\|$.

Part 1 : convergence de χ_1

Le terme χ_1 peut être décomposé comme suit :

$$\begin{aligned} \sup_{\lambda\in[0,2\,\pi)} \left| \hat{\gamma}^{\hat{\mathbf{S}}}(\lambda) - \gamma^{\mathbb{E}[v(\tau\,\xi)\,\tau]\,\mathbf{C}}(\lambda) \right| &\leq \sup_{\lambda\in[0,2\,\pi)} \left| \hat{\gamma}^{\hat{\mathbf{S}}}(\lambda) - \mathbb{E}\left[\hat{\gamma}^{\hat{\mathbf{S}}}(\lambda)\right] \right| \\ &+ \sup_{\lambda\in[0,2\,\pi)} \left| \mathbb{E}\left[\hat{\gamma}^{\hat{\mathbf{S}}}(\lambda)\right] - \gamma^{\mathbb{E}[v(\tau\,\xi)\,\tau]\,\mathbf{C}}(\lambda) \right|, \\ &\leq \chi_{11} + \chi_{12}\,, \end{aligned}$$

où $\chi_{11} = \sup_{\lambda\in[0,2\,\pi)} \left| \hat{\gamma}^{\hat{\mathbf{S}}}(\lambda) - \mathbb{E}\left[\hat{\gamma}^{\hat{\mathbf{S}}}(\lambda)\right] \right|$ et $\chi_{12} = \sup_{\lambda\in[0,2\,\pi)} \left| \mathbb{E}\left[\hat{\gamma}^{\hat{\mathbf{S}}}(\lambda)\right] - \gamma^{\mathbb{E}[v(\tau\,\xi)\,\tau]\,\mathbf{C}}(\lambda) \right|$.

Part 1.1 : convergence de χ_{11}

Pour montrer la convergence de χ_{11}, le lemme suivant est nécessaire :

Lemme B.3.1. Les résultats suivants sont vérifiés :

$$\hat{\gamma}^{\hat{\mathbf{S}}}(\lambda) = \mathbf{d}_m^H(\lambda)\,\hat{\mathbf{S}}\,\mathbf{d}_m(\lambda)\,, \tag{B.15}$$

$$\text{et :} \quad \mathbb{E}\left[\hat{\gamma}^{\hat{\mathbf{S}}}(\lambda)\right] = \mathbb{E}\left[v(\tau\,\xi)\,\tau\right]\,\mathbf{d}_m^H(\lambda)\,\mathbf{I}_m\,\mathbf{d}_m(\lambda)\,, \tag{B.16}$$

avec $\mathbf{d}_m(\lambda) = \dfrac{1}{\sqrt{m}}\left[1, e^{-i\,\lambda}, \ldots, e^{-i\,(m-1)\,\lambda}\right]^T$.

Démonstration. La preuve suit le même déroulement que le Lemme B.1.1. Tout d'abord, il est possible d'écrire :

$$\begin{aligned}
\hat{\gamma}^{\hat{\mathbf{S}}}(\lambda) &= \sum_{k=1-m}^{m-1} \check{s}_k\, e^{i\,k\,\lambda}\,, \\
&= \frac{1}{m\,N}\sum_{k=1-m}^{m-1} e^{i\,k\,\lambda}\sum_{j=0}^{m-1}\sum_{n=0}^{N-1}\hat{s}_{j,n}\,\hat{s}^{\star}_{j+k,n}\,\mathbb{1}_{0\leq j+k<m}\,, \\
&= \frac{1}{m\,N}\sum_{l,l'=0}^{m-1} e^{-i\,(l'-l)\,\lambda}\sum_{n=0}^{N-1}\hat{s}_{l,n}\,\hat{s}^{\star}_{l',n} = \mathbf{d}_m^H(\lambda)\,\hat{\mathbf{S}}\,\mathbf{d}_m(\lambda)\,.
\end{aligned}$$

De cette façon, la première partie du lemme est montrée.

En ce qui concerne $\mathbb{E}\left[\hat{\gamma}^{\hat{\mathbf{S}}}(\lambda)\right]$, il est possible d'introduire une matrice $\mathbf{D}$ diagonale, contenant les $\{v(\tau_i\,\xi)\}_{i\in[\![0,N-1]\!]}$. Puis :

$$\begin{aligned}
\mathbb{E}\left[\hat{\gamma}^{\hat{\mathbf{S}}}(\lambda)\right] &= \mathbf{d}_m^H(\lambda)\,\mathbb{E}\left[\hat{\mathbf{S}}\right]\,\mathbf{d}_m(\lambda)\,, \\
&= \mathbf{d}_m^H(\lambda)\,\mathbb{E}\left[\frac{\mathbf{Y}_w\,\mathbf{D}\,\mathbf{Y}_w^H}{N}\right]\,\mathbf{d}_m(\lambda)\,.
\end{aligned}$$

En développant chaque terme de $\mathbb{E}\left[\mathbf{Y}_w\,\mathbf{D}\,\mathbf{Y}_w^H\right]$, cela donne :

$$\begin{aligned}
\left(\mathbb{E}\left[\mathbf{Y}_w\,\mathbf{D}\,\mathbf{Y}_w^H\right]\right)_{i,j} &= \mathbb{E}\left[\sum_{n=0}^{N-1} v(\tau_n\,\xi)\,y_{w\,i,n}\,y^{\star}_{w\,j,n}\right]\,, \\
&= \sum_{n=0}^{N-1}\mathbb{E}\left[v(\tau_n\,\xi)\,\tau_n\right] = N\,\mathbb{E}\left[v(\tau_n\,\xi)\,\tau_n\right]\,.
\end{aligned}$$

Et ainsi : $\mathbb{E}\left[\hat{\gamma}^{\hat{\mathbf{S}}}(\lambda)\right] = \mathbb{E}\left[v(\tau\,\xi)\,\tau\right]\,\mathbf{d}_m^H(\lambda)\,\mathbf{I}_m\,\mathbf{d}_m(\lambda)$. □

Le reste de la preuve concernant χ_{11} est la même que pour $\hat{\gamma}^{noise}$ du Théorème 2.4.1, sauf pour le terme $\|\mathbf{T}\|_\infty$ qui sera remplacé par $\|\mathbf{T}\|_\infty\,\|\mathbf{D}\|_\infty$ pour la fin de la preuve, sans conséquences sur les résultats.

Ainsi, $\chi_{11} \xrightarrow{p.s.} 0$ lorsque $m \longrightarrow \infty$.

Part 1.2 : convergence de χ_{12}

Le Lemme B.3.1 et l'équation (B.15) permettent d'écrire :

$$\mathbb{E}\left[\hat{\gamma}^{\hat{\mathbf{S}}}(\lambda)\right] = \mathbb{E}\left[v(\tau\,\xi)\,\tau\right]\,\mathbf{d}_m^H(\lambda)\,\mathbf{C}\,\mathbf{d}_m(\lambda)\,.$$

et

$$\mathbb{E}\left[v(\tau\,\xi)\,\tau\right]\,\gamma^{\mathbf{C}}(\lambda) = \mathbb{E}\left[v(\tau\,\xi)\,\tau\right]\,\mathbf{d}_m^H(\lambda)\,\mathbf{C}\,\mathbf{d}_m(\lambda)\,.$$

Ainsi, $\chi_{12} = 0$.

Part 2 : convergence de $\chi_2 = \left\|\mathcal{T}\left(\hat{\mathbf{C}}_{FP} - \hat{\mathbf{S}}\right)\right\|$

Il a déjà été montré, dans l'article [34], que $\left\|\hat{\mathbf{C}}_{FP} - \hat{\mathbf{S}}\right\| \xrightarrow{p.s.} 0$. Soit $\mathbf{J}$ une matrice telle que $(\mathbf{J})_{j-i=1} = 1$ et 0 partout ailleurs. $\mathbf{J}^k$ contient des 1 seulement sur la k^{ieme} diagonale. Comme précédemment, grâce à (B.1), il est possible d'écrire :

$$\left\|\mathcal{T}\left(\hat{\mathbf{C}}_{FP} - \hat{\mathbf{S}}\right)\right\| \leq \sup_{\lambda\in[0,2\pi)}\left|\sum_{k=1-m}^{m-1}\left(\check{fp}_k - \check{s}_k\right)\,e^{i\,k\,\lambda}\right|\,.$$

Soit $\mathcal{T}\left(\hat{\mathbf{C}}_{FP}\right) = \mathcal{L}\left(\left(\check{fp}_0,\ldots,\check{fp}_{m-1}\right)^T\right)$ où $\check{fp}_{-k} = \check{fp}_k^*$ et $\mathcal{T}\left(\hat{\mathbf{S}}\right) = \mathcal{L}\left(\left(\check{s}_0,\ldots,\check{s}_{m-1}\right)^T\right)$ avec $\check{s}_{-k} = \check{s}_k^*$. Alors :

$$\begin{aligned}\sup_{\lambda\in[0,2\,\pi)}\left|\sum_{k=1-m}^{m-1}\left(\check{fp}_k - \check{s}_k\right)\,e^{i\,k\,\lambda}\right| &= \sup_{\lambda\in[0,2\,\pi)}\left|\sum_{k=1-m}^{m-1}\frac{1}{m}\sum_{p-1}^{m}\left(\check{fp}_k - \check{s}_k\right)\,e^{i\,k\,\lambda}\,\mathbb{1}_{0\leq p+k\leq m}\right|\,,\\ &= \sup_{\lambda\in[0,2\,\pi)}\left|\mathrm{Tr}\left(\left(\hat{\mathbf{C}}_{FP} - \hat{\mathbf{S}}\right)\,\frac{1}{m}\sum_{k=1-m}^{m-1}\left(\mathbf{J}^T\right)^k\,e^{i\,k\,\lambda}\right)\right|\,.\end{aligned}$$

De plus, $\dfrac{1}{m}\displaystyle\sum_{k=1-m}^{m-1}\left(\mathbf{J}^T\right)^k\,e^{i\,k\,\lambda} = \mathbf{d}_m(\lambda)\,\mathbf{d}_m^H(\lambda)$. Ceci permet d'écrire :

$$\begin{aligned}\left\|\mathcal{T}\left(\hat{\mathbf{C}}_{FP} - \hat{\mathbf{S}}\right)\right\| &\leq \sup_{\lambda\in[0,2\,\pi)}\left|\mathrm{Tr}\left(\left(\hat{\mathbf{C}}_{FP} - \hat{\mathbf{S}}\right)\,\mathbf{d}_m(\lambda)\,\mathbf{d}_m^H(\lambda)\right)\right|,\\ &= \sup_{\lambda\in[0,2\,\pi)}\left|\mathbf{d}_m^H(\lambda)\,\left(\hat{\mathbf{C}}_{FP} - \hat{\mathbf{S}}\right)\,\mathbf{d}_m(\lambda)\right|\,.\end{aligned}$$

Pour tout vecteur $\mathbf{x}$, la dernière équation devient :

$$\sup_{\lambda\in[0,2\,\pi)}\left|\mathbf{d}_m^H(\lambda)\,\left(\hat{\mathbf{C}}_{FP} - \hat{\mathbf{S}}\right)\,\mathbf{d}_m(\lambda)\right| \leq \sup_{\|\mathbf{x}\|_2=1}\left|\mathbf{x}^H\,\left(\hat{\mathbf{C}}_{FP} - \hat{\mathbf{S}}\right)\,\mathbf{x}\right|\,,$$

Soit :

$$\sup_{\lambda\in[0,2\,\pi)}\left|\mathbf{d}_m^H(\lambda)\,\left(\hat{\mathbf{C}}_{FP} - \hat{\mathbf{S}}\right)\,\mathbf{d}_m(\lambda)\right| \leq \sup_{\|\mathbf{x}\|_2=1}\left\|\left(\hat{\mathbf{C}}_{FP} - \hat{\mathbf{S}}\right)\,\mathbf{x}\right\|_2 \leq \left\|\hat{\mathbf{C}}_{FP} - \hat{\mathbf{S}}\right\|\,.$$

Finalement, $\left\|\mathcal{T}\left(\hat{\mathbf{C}}_{FP} - \hat{\mathbf{S}}\right)\right\| \leq \left\|\hat{\mathbf{C}}_{FP} - \hat{\mathbf{S}}\right\|$. Comme $\left\|\hat{\mathbf{C}}_{FP} - \hat{\mathbf{S}}\right\| \xrightarrow{p.s.} 0$ alors $\chi_2 \xrightarrow{p.s.} 0$ et la preuve tu théorème est ainsi achevée.

B.4 Preuve du Théorème 2.4.5

La preuve du Théorème 2.4.5 s'appuie sur des résultats précédents et sur l'article [33]. La preuve est menée sous les hypothèses de Maronna (1.8), où le théorème 2.4.1 s'applique. Soit $\hat{C}_{SCM} = \frac{1}{N}\mathbf{Y}\mathbf{Y}^H$. Alors :

$$\| \check{\mathbf{C}}_{TYL} - \mathbf{C} \| \leq \| \check{\mathbf{C}}_{TYL} - \frac{1}{\mathbb{E}(\tau)}\mathcal{T}(\hat{C}_{SCM}) \| + \| \frac{1}{\mathbb{E}(\tau)}\mathcal{T}(\hat{C}_{SCM}) - \mathbf{C} \| \ . \tag{B.17}$$

De plus, il a été montré précédemment, d'après le Théorème (2.4.1), que :

$$\| \frac{1}{\mathbb{E}(\tau)}\mathcal{T}(\hat{C}_{SCM}) - \mathbf{C} \| \overset{p.s.}{\to} 0 \, .$$

Pour un bruit de moyenne nulle, il a été montré dans l'article [33] que : $\| \hat{\mathbf{C}_{TYL}} - \hat{\mathbf{C}}_{SCM} \| \overset{p.s.}{\to} 0$.
Comme le bruit peut ici s'écrire $\mathbf{y}_i = \sqrt{\tau_i}\mathbf{C}^{1/2}\mathbf{x}_i$, les $\mathbf{x}_i$ de moyenne nulle et indépendants des τ_i, il est ainsi possible d'adapter ce résultat de la manière suivante : $\| \hat{\mathbf{C}}_{TYL} - \frac{1}{\mathbb{E}(\tau)}\hat{\mathbf{C}}_{SCM} \| \overset{p.s.}{\to} 0$.
En effet, la multiplication des $\mathbf{x}_i$ par des scalaires τ_i ne modifie pas l'estimateur $\hat{\mathbf{C}}_{TYL}$ puisque ces termes s'équilibrent au numérateur et au dénominateur. Il ne reste donc qu'à diviser par la moyenne des $(\sqrt{\tau_i})^2$ le second terme de l'équation, pour se retrouver avec un bruit de moyenne nulle.

L'opérateur de Toeplitz $\mathcal{T}$ étant un opérateur linéaire,

$$\| \mathcal{T}(\hat{\mathbf{C}}_{TYL}) - \frac{1}{\mathbb{E}(\tau)}\mathcal{T}(\hat{\mathbf{C}}_{SCM}) \| \overset{p.s.}{\to} 0 \, .$$

Ceci termine la preuve du Théorème (2.4.5).

B.5 Preuve du Théorème 2.4.6

Soit $\mathbf{Y}_w = \mathbf{C}^{-1/2}\mathbf{M}\,\mathbf{S} + \mathbf{X}\,\mathbf{T}^{1/2}$, et $\mathbf{Y}_w = [\mathbf{y}_{w1}, \ldots, \mathbf{Y}_{wN}]$. Soit la matrice $\check{\mathbf{\Sigma}}_w$, l'unique solution de :

$$\check{\mathbf{\Sigma}}_w = \frac{m}{N}\sum_{i=1}^{N} \frac{\mathbf{y}_{wi}\mathbf{y}_{wi}^H}{\mathbf{y}_{wi}^H \check{\mathbf{\Sigma}}_w^{-1} \mathbf{y}_{wi}} \, . \tag{B.18}$$

De plus, pour rappel,

$$\check{\mathbf{\Sigma}}_{TYL} = \frac{m}{N}\sum_{i=1}^{N} \frac{\mathbf{y}_{wi}\mathbf{y}_{wi}^H}{\mathbf{y}_{wi}^H \check{\mathbf{\Sigma}}_{TYL}^{-1} \mathbf{y}_{wi}} \, .$$

Ainsi, en remplaçant $\check{\mathbf{y}}_{wi}$ par $\mathbf{C}^{-1/2}\ \check{\mathbf{C}}_{TYL}^{1/2}\ \mathbf{y}_{wi}$ pour tout i tel que $1 \leq i \leq N$, il reste :

$$\mathbf{C}^{-1/2}\ \check{\mathbf{C}}_{TYL}^{1/2}\ \check{\mathbf{\Sigma}}_{TYL}\ \check{\mathbf{C}}_{TYL}^{1/2}\mathbf{C}^{-1/2} = \frac{m}{N}\sum_{i=1}^{N} \frac{\mathbf{y}_{wi}\mathbf{y}_{wi}^H}{\mathbf{y}_{wi}^H \left(\mathbf{C}^{-1/2}\check{\mathbf{C}}_{TYL}^{1/2}\check{\mathbf{\Sigma}}_{TYL}\check{\mathbf{C}}_{TYL}^{1/2}\mathbf{C}^{-1/2}\right)^{-1} \mathbf{y}_{wi}} \, . \tag{B.19}$$

Cette précédente équation permet ainsi de déduire que :

$$\check{\mathbf{\Sigma}}_{TYL} = \check{\mathbf{C}}_{TYL}^{-1/2}\ \mathbf{C}^{1/2}\ \check{\mathbf{\Sigma}}_w\ \mathbf{C}^{1/2}\ \check{\mathbf{C}}_{TYL}^{-1/2} \, . \tag{B.20}$$

Puis, l'équation (2.41) peut se développer de la manière suivante :

$$\| \check{\mathbf{\Sigma}}_{TYL} - \hat{\mathbf{S}}_w \| \leq \left\| \check{\mathbf{\Sigma}}_{TYL} - \check{\mathbf{\Sigma}}_w \right\| + \left\| \check{\mathbf{\Sigma}}_w - \hat{\mathbf{S}}_w \right\| \, . \tag{B.21}$$

En ce qui concerne le second terme de l'équation, il a déjà été montré dans l'article [127] que $\left\|\check{\mathbf{\Sigma}}_w - \hat{\mathbf{S}}_w\right\| \overset{p.s.}{\to} 0$.

Quant au premier terme de l'équation, il peut être réécrit de la manière suivante :

$$\left\|\check{\mathbf{\Sigma}}_{TYL} - \check{\mathbf{\Sigma}}_w\right\| \leq \left\|\check{\mathbf{C}}_{TYL}^{-1/2}\mathbf{C}^{1/2}\check{\mathbf{\Sigma}}_w\mathbf{C}^{1/2}\check{\mathbf{C}}_{TYL}^{-1/2} - \check{\mathbf{\Sigma}}_w\mathbf{C}^{1/2}\check{\mathbf{C}}_{TYL}^{-1/2}\right\| + \left\|\check{\mathbf{\Sigma}}_w\mathbf{C}^{1/2}\check{\mathbf{C}}_{TYL}^{-1/2} - \check{\mathbf{\Sigma}}_w\right\| .$$

Puis, en factorisant à droite et à gauche :

$$\left\|\check{\mathbf{\Sigma}}_{TYL} - \check{\mathbf{\Sigma}}_w\right\| \leq \left\|\check{\mathbf{C}}_{TYL}^{-1/2}\mathbf{C}^{1/2} - \mathbf{I}_m\right\| \left\|\check{\mathbf{\Sigma}}_w\right\| \left(\left\|\mathbf{C}^{1/2}\check{\mathbf{C}}_{TYL}^{-1/2}\right\| + 1\right) .$$

Or, comme $\|\mathbf{C}\|$ a un support borné, le Théorème 2.4.5 assure que $\left\|\check{\mathbf{C}}_{TYL}\right\|$ est aussi à support borné. De plus, ce théorème montre la consistance puisque $\left\|\mathbf{C} - \check{\mathbf{C}}_{TYL}\right\| \xrightarrow{p.s.} 0$. Ceci termine la preuve.

Annexe C

Publications

C.1 Conférences

E. Terreaux, J.P. Ovarlez and F. Pascal. Anomaly detection and estimation in hyperspectral imaging using random matrix theory tools. *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, IEEE 6th International Workshop on. IEEE, p. 169-172, 2015.

E. Terreaux, J.P. Ovarlez and F. Pascal. New model order selection in large dimension regime for complex elliptically symmetric noise. *Signal Processing Conference (EUSIPCO)*, 2017 25th European. IEEE, p. 1090-1094, 2017.

E. Terreaux, J.P. Ovarlez and F. Pascal. A Toeplitz-Tyler Estimation of the Model Order in Large Dimensional Regime, *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, p. 4489-4493, 2018.

E. Terreaux, J.P. Ovarlez and F. Pascal. Estimation de l'Ordre de Modele pour un Bruit Elliptique Symétrique Complexe en Grande Dimension. *GRETSI 2017*, 2017.

E. Jay, E. Terreaux, J.P. Ovarlez, and al. Improving Portfolios Global Performance with Robust Covariance Matrix Estimation : Application to the Maximum Variety Portfolio. arXiv preprint arXiv :1804.00191, 2018.

C.1.1 Soumis

E. Jay [et al.]. Improving portfolios global performance using a cleaned and robust covariance matrix estimate, (sciencesconf.org :dyses2018 :221020), 2018.

C.2 Journaux

E. Terreaux, J.P. Ovarlez and F. Pascal. Robust model order selection in large dimensional elliptically symmetric noise. arXiv preprint arXiv :1710.06735, 2017.

Bibliographie

[1] Y. I. Abramovich, D. A. Gray, A. Y. Gorokhov, and N. K. Spencer. Positive-definite toeplitz completion in DOA estimation for nonuniform linear antenna arrays. i. fully augmentable arrays. *IEEE Transactions on Signal Processing*, 46(9) :2458–2471, Sep 1998.

[2] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6) :716–723, Dec 1974.

[3] T. W. Anderson. *Statistical inference in elliptically contoured and related distributions.* Allerton Pr, 1990.

[4] N. Arkind and B. Nadler. Parametric joint detection-estimation of the number of sources in array processing. In *2010 Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop, Jerusalem, Israel*, pages 269–272, Oct 2010.

[5] L. Arnold, V. M. Gundlach, and L. Demetrius. Evolutionary formalism for products of positive random matrices. *The Annals of Applied Probability*, pages 859–901, 1994.

[6] C. S. Asness, T. J. Moskowitz, and L. H. Pedersen. Value and momentum everywhere. 68(3) :929–985, june 2013.

[7] Z.-D. Bai and J. W. Silverstein. No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Annals of probability*, pages 316–345, 1998.

[8] J. Baik, G. B. Arous, S. Péché, et al. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability*, 33(5) :1643–1697, 2005.

[9] J. Baik and J. W. Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, 97(6) :1382–1408, 2006.

[10] J. Bernardo and A. F. M. Smith. *Bayesian Theory*, volume 15. 01 2000.

[11] J. M. Bioucas-Dias and J. M. Nascimento. Estimation of signal subspace on hyperspectral data. In *Image and Signal Processing for Remote Sensing XI*, volume 5982, page 59820L. International Society for Optics and Photonics, 2005.

[12] J. M. Bioucas-Dias and J. M. Nascimento. Hyperspectral subspace identification. *IEEE Transactions on Geoscience and Remote Sensing*, 46(8) :2435–2445, 2008.

[13] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot. Hyperspectral unmixing overview : Geometrical, statistical, and sparse regression-based approaches. *IEEE journal of selected topics in applied earth observations and remote sensing*, 5(2) :354–379, 2012.

[14] J. P. Burg, D. G. Luenberger, and D. L. Wenger. Estimation of structured covariance matrices. *Proceedings of the IEEE*, 70(9) :963–974, Sept 1982.

[15] T. T. Cai, Z. Ren, and H. H. Zhou. Optimal rates of convergence for estimating toeplitz covariance matrices. *Probability Theory and Related Fields*, 156(1-2) :101–143, 2013.

[16] S. Cambanis, S. Huang, and G. Simons. On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, 11(3) :368–385, 1981.

[17] K. Cawse, A. Robin, and M. Sears. The effect of noise whitening on methods for determining the intrinsic dimension of a hyperspectral image. In *Proc. Workshops on Hyperspectral Image and Signal Processing : Evolution in Remote Sensing (WHISPERS), 2011, Lisbon, Portugal*, pages 1–4. IEEE, 2011.

[18] K. Cawse, M. Sears, A. Robin, S. B. Damelin, K. Wessels, F. van den Bergh, and R. Mathieu. Using random matrix theory to determine the number of endmembers in a hyperspectral image. In *2010 2nd Workshop on Hyperspectral Image and Signal Processing : Evolution in Remote Sensing, Reykjavik, Iceland*, pages 1–4, June 2010.

[19] K. Cawse-Nicholson, S. B. Damelin, A. Robin, and M. Sears. Determining the intrinsic dimension of a hyperspectral image using random matrix theory. *IEEE Transactions on Image Processing*, 22(4) :1301–1310, 2013.

[20] S. Chandrasekhar. *Radiative transfer*. Courier Corporation, 2013.

[21] C.-I. Chang. *Hyperspectral imaging : techniques for spectral detection and classification*, volume 1. Springer Science & Business Media, 2003.

[22] C.-I. Chang and Q. Du. Estimation of number of spectrally distinct signal sources in hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 42(3) :608–619, 2004.

[23] F. Chapon, R. Couillet, W. Hachem, and X. Mestre. The outliers among the singular values of large rectangular random matrices with additive fixed rank deformation. *arXiv preprint arXiv :1207.0471*, 2012.

[24] C. H. Chen and X. Zhang. Independent component analysis for remote sensing study. In *Image and Signal Processing for Remote Sensing V*, volume 3871, pages 150–159. International Society for Optics and Photonics, 1999.

[25] C. Y. Chong, F. Pascal, J.-P. Ovarlez, and M. Lesturgie. Mimo radar detection in non-gaussian and heterogeneous clutter. *IEEE Journal of Selected Topics in Signal Processing*, 4(1) :115–126, 2010.

[26] M. Chu, F. Diele, R. Plemmons, and S. Ragni. Optimality, computation, and interpretation of nonnegative matrix factorizations. In *SIAM Journal on Matrix Analysis*. Citeseer, 2004.

[27] A. Combernoux, F. Pascal, M. Lesturgie, and G. Ginolhac. Performances of low rank detectors based on random matrix theory with application to stap. In *Proc. 2014 International Radar Conference*, pages 1–4, Oct 2014.

[28] P. Comon. Independent component analysis, a new concept ? *Signal processing*, 36(3) :287–314, 1994.

[29] R. Couillet. Robust spiked random matrices and a robust G-MUSIC estimator. *Journal of Multivariate Analysis*, 140 :139–161, 2015.

[30] R. Couillet and M. Debbah. *Random matrix methods for wireless communications*. Cambridge University Press, 2011.

[31] R. Couillet and W. Hachem. Fluctuations of spiked random matrix models and failure diagnosis in sensor networks. *IEEE Transactions on Information Theory*, 59(1) :509–525, 2013.

[32] R. Couillet, F. Pascal, and J. W. Silverstein. A joint robust estimation and random matrix framework with application to array processing. In *Proc. Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), Vancouver, BC, Canada, 2013*, pages 6561–6565. IEEE, 2013.

[33] R. Couillet, F. Pascal, and J. W. Silverstein. Robust estimates of covariance matrices in the large dimensional regime. *IEEE Transactions on Information Theory*, 60(11) :7269–7278, 2014.

[34] R. Couillet, F. Pascal, and J. W. Silverstein. The random matrix regime of Maronna's M-estimator with elliptically distributed samples. *Journal of Multivariate Analysis*, 139 :56–78, 2015.

[35] R. database. Hyperspectral remote sensing scenes.

[36] T. Diskin, G. Draskovic, F. Pascal, and A. Wiesel. Deep robust regression. In *IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP),Curaçao, Dutch Antilles*, pages 1–5. IEEE, 2017.

[37] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A. O. Hero. Joint bayesian endmember extraction and linear unmixing for hyperspectral imagery. *IEEE Transactions on Signal Processing*, 57(11) :4355–4368, 2009.

[38] N. Dobigeon, J.-Y. Tourneret, C. Richard, J. C. M. Bermudez, S. McLaughlin, and A. O. Hero. Nonlinear unmixing of hyperspectral images : Models and algorithms. *IEEE Signal Processing Magazine*, 31(1) :82–94, 2014.

[39] A. Goetz, G. Vane, J. E. Solomon, and B. N. Rock. Imaging spectrometry for earth remote sensing. 228 :1147–53, 07 1985.

[40] R. Gray. On the asymptotic eigenvalue distribution of Toeplitz matrices. *IEEE Transactions on Information Theory*, 18(6), Nov 1972.

[41] R. M. Gray et al. Toeplitz and circulant matrices : A review. *Foundations and Trends® in Communications and Information Theory*, 2(3) :155–239, 2006.

[42] A. A. Green, M. Berman, P. Switzer, and M. D. Craig. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing*, 26(1) :65–74, Jan 1988.

[43] R. O. Green, M. L. Eastwood, C. M. Sarture, T. G. Chrien, M. Aronsson, B. J. Chippendale, J. A. Faust, B. E. Pavri, C. J. Chovit, M. Solis, M. R. Olah, and O. Williams. Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (aviris). *Remote Sensing of Environment*, 65(3) :227 – 248, 1998.

[44] E. J. Gumbel. *Statistics of extremes*. Courier Corporation, 2012.

[45] W. Hachem, P. Loubaton, X. Mestre, J. Najim, and P. Vallet. A subspace estimator for fixed rank perturbations of large random matrices. *Journal of Multivariate Analysis*, 114 :427–447, 2013.

[46] A. Halimi, P. Honeine, M. Kharouf, C. Richard, and J. Y. Tourneret. Estimating the intrinsic dimension of hyperspectral images using a noise-whitened eigengap approach. *IEEE Transactions on Geoscience and Remote Sensing*, 54(7) :3811–3821, July 2016.

[47] J. Haupt, W. U. Bajwa, G. Raz, and R. Nowak. Toeplitz compressed sensing matrices with applications to sparse channel estimation. *IEEE Transactions on Information Theory*, 56(11) :5862–5875, Nov 2010.

[48] D. C. Heinz et al. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 39(3) :529–545, 2001.

[49] R. Heylen and P. Scheunders. Hyperspectral intrinsic dimensionality estimation with nearest-neighbor distance ratios. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(2) :570–579, 2013.

[50] P. J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, pages 73–101, 1964.

[51] A. Huck, M. Guillaume, and J. Blanc-Talon. Minimum dispersion constrained nonnegative matrix factorization to unmix hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 48(6) :2590–2602, 2010.

[52] E. Jay. *Détection en environnement non gaussien*. PhD thesis, Université de Cergy Pontoise, 2002.

[53] E. Jay, E. Terreaux, J. Ovarlez, and F. Pascal. Improving portfolios global performance with robust covariance matrix estimation : Application to the maximum variety portfolio. *Porc. European Signal Processing Conference (EUSIPCO), Rome, Italy*, 2018.

[54] I. Johnstone. *On the distribution of the largest principal component.* Department of Statistics, Stanford University, 2000.

[55] D. Kelker. Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā : The Indian Journal of Statistics, Series A*, pages 419–430, 1970.

[56] J. T. Kent and D. E. Tyler. Redescending M-estimates of multivariate location and scatter. *The Annals of Statistics*, pages 2102–2119, 1991.

[57] N. Keshava, J. Kerekes, D. Manolakis, and G. A. Shaw. Algorithm taxonomy for hyperspectral unmixing [4049-04]. 01 2000.

[58] N. Keshava and J. F. Mustard. Spectral unmixing. *IEEE Signal Processing Magazine*, 19(1) :44–57, Jan 2002.

[59] R. Kokaly. Usgs spectral library version 7.

[60] R. Kokaly, R. Clark, G. Swayze, K. Livo, T. Hoefen, N. Pearson, R. Wise, W. Benzel, H. Lowers, R. Driscoll, et al. USGS spectral library version 7 data : US geological survey data release, 2017.

[61] P. Krishnaiah and J. Lin. Complex elliptically symmetric distributions. *Communications in Statistics-Theory and Methods*, 15(12) :3693–3718, 1986.

[62] S. Kritchman and B. Nadler. Non-parametric detection of the number of signals : Hypothesis testing and random matrix theory. *IEEE Transactions on Signal Processing*, 57(10) :3930–3941, 2009.

[63] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*, pages 556–562, 2001.

[64] C.-J. Lin. Projected gradient methods for nonnegative matrix factorization. *Neural computation*, 19(10) :2756–2779, 2007.

[65] X. Liu, W. Xia, B. Wang, and L. Zhang. An approach based on constrained nonnegative matrix factorization to unmix hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 49(2) :757–772, 2011.

[66] P. Loubaton. On the almost sure location of the singular values of certain gaussian block-hankel large random matrices. *Journal of Theoretical Probability*, 29(4) :1339–1443, 2016.

[67] M. Mahot. *Estimation robuste de la matrice de covariance en traitement du signal.* PhD thesis, École normale supérieure de Cachan-ENS Cachan, 2012.

[68] E. R. Malinowski. Determination of the number of factors and the experimental error in a data matrix. *Analytical Chemistry*, 49(4) :612–617, 1977.

[69] D. Manolakis and G. Shaw. Detection algorithms for hyperspectral imaging applications. *IEEE Signal Processing Magazine*, 19(1) :29–43, 2002.

[70] D. Manolakis, E. Truslow, M. Pieper, T. Cooley, and M. Brueggeman. Detection algorithms in hyperspectral imaging systems : An overview of practical algorithms. *IEEE Signal Processing Magazine*, 31(1) :24–33, Jan 2014.

[71] D. G. Manolakis and D. B. Marden. Dimensionality reduction of hyperspectral imaging data using local principal components transforms. In *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery X*, volume 5425, pages 393–402. International Society for Optics and Photonics, 2004.

[72] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4) :457, 1967.

[73] H. Markowitz. Portfolio selection. *The journal of finance*, 7(1) :77–91, 1952.

[74] R. A. Maronna and V. J. Yohai. Robust estimation of multivariate location and scatter. *Wiley StatsRef : Statistics Reference Online*, 1976.

[75] X. Mestre. Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates. *IEEE Transactions on Information Theory*, 54(11) :5113–5129, 2008.

[76] X. Mestre. On the asymptotic behavior of the sample estimates of eigenvalues and eigenvectors of covariance matrices. *IEEE Transactions on Signal Processing*, 56(11) :5353–5368, 2008.

[77] L. Miao and H. Qi. Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization. *IEEE Transactions on Geoscience and Remote Sensing*, 45(3) :765–777, 2007.

[78] R. R. Nadakuditi and J. W. Silverstein. Fundamental limit of sample generalized eigenvalue based detection of signals in noise using relatively few signal-bearing and noise-only samples. *IEEE Journal of Selected Topics in Signal Processing*, 4(3) :468–480, 2010.

[79] B. Nadler, F. Penna, and R. Garello. Performance of eigenvalue-based signal detectors with known and unknown noise level. In *2011 IEEE International Conference on Communications (ICC)*, pages 1–5, June 2011.

[80] J. P. L. Nasa. Aviris airborne visible infrared imaging spectrometer.

[81] J. M. Nascimento and J. M. Dias. Does independent component analysis play a role in unmixing hyperspectral data ? *IEEE Transactions on Geoscience and Remote Sensing*, 43(1) :175–187, 2005.

[82] J. M. Nascimento and J. M. Dias. Vertex component analysis : A fast algorithm to unmix hyperspectral data. *IEEE transactions on Geoscience and Remote Sensing*, 43(4) :898–910, 2005.

[83] N. M. Nasrabadi. Hyperspectral target detection : An overview of current and future challenges. *IEEE Signal Processing Magazine*, 31(1) :34–44, 2014.

[84] E. Ollila and V. Koivunen. Robust antenna array processing using M-estimators of pseudo-covariance. In *Personal, Indoor and Mobile Radio Communications, 2003. PIMRC 2003. 14th IEEE Proceedings on*, volume 3, pages 2659–2663. IEEE, 2003.

[85] E. Ollila and V. Koivunen. Generalized complex elliptical distributions. In *Sensor Array and Multichannel Signal Processing Workshop Proceedings, 2004*, pages 460–464. IEEE, 2004.

[86] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor. Complex elliptically symmetric distributions : Survey, new results and applications. *IEEE Transactions on Signal Processing*, 60(11) :5597–5625, Nov 2012.

[87] F. Pascal, Y. Chitour, J.-P. Ovarlez, P. Forster, and P. Larzabal. Covariance structure maximum-likelihood estimates in compound Gaussian noise : Existence and algorithm analysis. *IEEE Transactions on Signal Processing*, 56(1) :34–48, 2008.

[88] F. Pascal, Y. Chitour, and Y. Quek. Generalized robust shrinkage estimator–Application to STAP data. *Submitted for publication. URL http ://arxiv. org/pdf/1311.6567*, 2013.

[89] F. Pascal and A. Kammoun. Théorie des matrices aléatoires robustes et applications à la détection radar. *Traitement du Signal*, 2016.

[90] D. Passemier and J.-F. Yao. On determining the number of spikes in a high-dimensional spiked population model. *Random Matrices : Theory and Applications*, 1(01) :1150002, 2012.

[91] L. A. Pastur and M. Shcherbina. *Eigenvalue distribution of large random matrices*. Number 171. American Mathematical Soc., 2011.

[92] J. Plaza, A. Plaza, R. Perez, and P. Martinez. On the use of small training sets for neural network-based characterization of mixed pixels in remotely sensed hyperspectral images. *Pattern Recognition*, 42(11) :3032–3045, 2009.

[93] J. M. F. Pons. *Robust target detection for Hyperspectral Imaging.* PhD thesis, Supélec, 2014.

[94] E. Radoi and A. Quinquis. A new method for estimating the number of harmonic components in noise with application in high resolution radar. *EURASIP Journal on Applied Signal Processing*, 2004 :1177–1188, 2004.

[95] J. Rissanen. Modeling by shortest data description. *Automatica*, 14(5) :465–471, Sept 1978.

[96] A. Robin, K. Cawse-Nicholson, A. Mahmood, and M. Sears. Estimation of the intrinsic dimension of hyperspectral images : Comparison of current methods. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6) :2854–2861, 2015.

[97] C. Rodarmel and J. Shan. Principal component analysis for hyperspectral image classification. *Surveying and Land Information Science*, 62(2) :115, 2002.

[98] F. Rubio, X. Mestre, and D. P. Palomar. Performance analysis and optimal selection of large minimum variance portfolios under estimation risk. *IEEE Journal of Selected Topics in Signal Processing*, 6(4) :337–350, Aug 2012.

[99] G. Schwarz et al. Estimating the dimension of a model. *The Annals of Statistics*, 6(2) :461–464, 1978.

[100] G. A. Shaw and H.-h. K. Burke. Spectral imaging for remote sensing. *Lincoln laboratory journal*, 14(1) :3–28, 2003.

[101] J. W. Silverstein and S.-I. Choi. Analysis of the limiting spectral distribution of large dimensional random matrices. *Journal of Multivariate Analysis*, 54(2) :295–309, 1995.

[102] M. O. Smith, P. E. Johnson, and J. B. Adams. Quantitative determination of mineral types and abundances from reflectance spectra using principal components analysis. *Journal of Geophysical Research : Solid Earth*, 90(S02), 1985.

[103] Spectir. Spectir hyperspectral and remote sensing solutions.

[104] M. R. Stites, T. K. Moon, J. H. Gunther, and G. P. Williams. A bayesian framework for abundance estimation in hyperspectral data using markov random fields. In *Proc. Conf. Signals, Systems and Computers, ACSSC Conference Record of the Forty-First Asilomar, Pacific Grove, CA, USA*, pages 725–729. IEEE, 2007.

[105] Stoxx. Stoxx innovate. global. indices.

[106] L. Sun. `http://lesun.weebly.com/hyperspectral-data-set.html`, consulté en 2018.

[107] E. Terreaux, J.-P. Ovarlez, and F. Pascal. A Toeplitz-Tyler estimation of the model order in large dimensional regime.

[108] E. Terreaux, J.-P. Ovarlez, and F. Pascal. Anomaly detection and estimation in hyperspectral imaging using random matrix theory tools. In *Proc. Conf. on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), Cancun, Mexico*, pages 169–172. IEEE, 2015.

[109] E. Terreaux, J.-P. Ovarlez, and F. Pascal. New model order selection in large dimension regime for complex elliptically symmetric noise. In *Proc. Conf. on Signal Processing Conference (EUSIPCO), 2017 25th European, Kos Island, Greece*, pages 1090–1094. IEEE, 2017.

[110] E. Terreaux, J.-P. Ovarlez, and F. Pascal. Robust model order selection in large dimensional elliptically symmetric noise. *arXiv preprint arXiv :1710.06735*, 2017.

[111] D. E. Tyler. A distribution-free M-estimator of multivariate scatter. *The Annals of Statistics*, pages 234–251, 1987.

[112] D. E. Tyler. Some results on the existence, uniqueness, and computation of the M-estimates of multivariate location and scatter. *SIAM Journal on Scientific and Statistical Computing*, 9(2) :354–362, 1988.

[113] P. Vallet. Matrices aléatoires et applications au traitement statistique du signal synthèse des résultats en français. 2012.

[114] P. Vallet, P. Loubaton, and X. Mestre. On the consistency of likelihood penalization methods in large sensor networks. In *Proc. Workshop on Sensor Array and Multichannel Signal Processing Workshop (SAM), Hoboken, NJ, USA*, pages 109–112, June 2012.

[115] P. Vallet, X. Mestre, and P. Loubaton. Performance analysis of an improved music doa estimator. *IEEE Transactions on Signal Processing*, 63(23) :6407–6422, Dec 2015.

[116] J. Vinogradova. *Random matrices and applications to detection and estimation in array processing.* PhD thesis, Telecom ParisTech Paris, 2014.

[117] J. Vinogradova, R. Couillet, and W. Hachem. A new method for source detection, power estimation, and localization in large sensor networks under noise with unknown statistics. In *Proc. Conf. on Acoustics, Speech and Signal Processing (ICASSP), Vancouver, BC, Canada*, pages 3943–3946. IEEE, 2013.

[118] J. Vinogradova, R. Couillet, and W. Hachem. Statistical inference in large antenna arrays under unknown noise pattern. *IEEE Transactions on Signal Processing*, 61(22) :5633–5645, 2013.

[119] J. Vinogradova, R. Couillet, and W. Hachem. Estimation of toeplitz covariance matrices in large dimensional regime with application to source detection. *IEEE Transactions on Signal Processing*, 63(18) :4903–4913, 2015.

[120] M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(2) :387–392, 1985.

[121] E. P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics*, 62(3) :548–564, 1955.

[122] R. M. Willett, M. F. Duarte, M. A. Davenport, and R. G. Baraniuk. Sparsity and structure in hyperspectral imaging : Sensing, reconstruction, and target detection. *IEEE Signal Processing Magazine*, 31(1) :116–126, 2014.

[123] M. E. Winter. N-findr : An algorithm for fast autonomous spectral end-member determination in hyperspectral data. In *Imaging Spectrometry V*, volume 3753, pages 266–276. International Society for Optics and Photonics, 1999.

[124] J. Wishart. The generalized product moment distribution in samples from a normal multivariate population. *Biometrika*, 31261, 1928.

[125] Y. Xu, Z. Wu, J. Li, A. Plaza, and Z. Wei. Anomaly detection in hyperspectral images based on low-rank and sparse representation. *IEEE Transactions on Geoscience and Remote Sensing*, 54(4) :1990–2000, April 2016.

[126] B. Yang, M. Yang, A. Plaza, L. Gao, and B. Zhang. Dual-mode fpga implementation of target and anomaly detection algorithms for real-time hyperspectral imaging. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6) :2950–2961, June 2015.

[127] T. Zhang, X. Cheng, and A. Singer. Marchenko-pastur law for Tyler's and Maronna's M-estimators. *arXiv preprint*, 2014.

[128] F. Zhu, Y. Wang, S. Xiang, B. Fan, and C. Pan. Structured sparse method for hyperspectral unmixing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 88 :101–118, 2014.

Titre : Théorie des matrices aléatoires pour l'imagerie hyperspectrale
Mots-clés : Théorie des matrices aléatoires, imagerie hyperspectrale, modèle linéaire, bruit CES, estimation d'ordre de modèle, démélange spectral, estimation robuste, optimisation sous contraintes.

Résumé : La finesse de la résolution spectrale et spatiale des images hyperspectrales en font des données de très grande dimension. C'est également le cas d'autres types de données, où leur taille tend à augmenter pour de plus en plus d'applications. La complexité des données provenant de l'hétérogénéité spectrale et spatiale, de la non gaussianité du bruit et des processus physiques sous-jacents, renforcent la richesse des informations présentes sur une image hyperspectrale. Exploiter ces informations demande alors des outils statistiques adaptés aux grandes données mais aussi à leur nature non gaussienne. Des méthodes reposant sur la théorie des matrices aléatoires, théorie adaptée aux données de grande dimension, et reposant sur la robustesse, adaptée aux données non gaussiennes, sont ainsi proposées dans cette thèse, pour des applications à l'imagerie hyperspectrale.

Cette thèse propose d'améliorer deux aspects du traitement des images hyperspectrales : l'estimation du nombre d'endmembers ou de l'ordre du modèle et le problème du démélange spectral. En ce qui concerne l'estimation du nombre d'endmembers, trois nouveaux algorithmes adaptés au modèle choisi sont proposés, le dernier présentant de meilleures performances que les deux autres, en raison de sa plus grande robustesse. Une application au domaine de la finance est également proposée. Pour le démélange spectral, trois méthodes sont proposées, qui tiennent comptent des différentes particularités possibles des images hyperspectrales.

Cette thèse a permis de montrer que la théorie des matrices aléatoires présente un grand intérêt pour le traitement des images hyperspectrales. Les méthodes développées peuvent également s'appliquer à d'autres domaines nécessitant le traitement de données de grandes dimensions.

Title: Random Matrix Theory for Hyperspectral Imaging
Keywords: Random matrix theory, hyperspectral imaging, linear model, CES noise, model order selection, unmixing, robust estimation, constrained optimization

Abstract: Hyperspectral imaging generates large data due to the spectral and spatial high resolution, as it is the case for more and more other kinds of applications. For hyperspectral imaging, the data complexity comes from the spectral and spatial heterogeneity, the non-gaussianity of the noise and other physical processes. Nevertheless, this complexity enhances the wealth of collected informations, that need to be processed with adapted methods. Random matrix theory and robust processes are here suggested for hyperspectral imaging application: the random matrix theory is adapted to large data and the robustness enables to better take into account the non-gaussianity of the data.

This thesis aims to enhance the model order selection on a hyperspectral image and the unmixing problem. As the model order selection is concerned, three new algorithms are developped, and the last one, more robust, gives better performances. One financial application is also presented. As for the unmixing problem, three methods that take into account the peculierities of hyperspectral imaging are suggested.

The random matrix theory is of great interest for hyperspectral image processing, as demonstrated in this thesis. Differents methods developped here can be applied to other field of signal processing requiering the processing of large data.

université PARIS-SACLAY
DÉPARTEMENT
Sciences et Technologies de l'Information et de la Communication
DGA