Algorithmes d'Estimation et de Détection en Contexte Hétérogène Rang Faible

Arnaud Breloy

université PARIS-SACLAY
ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

SATIE

NNT : 2015SACLN021

Laboratoire SATIE, ENS Cachan

THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE PARIS-SACLAY

présentée par

Arnaud Breloy

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE PARIS-SACLAY

Domaine : Traitement du signal

Sujet de la thèse :

Algorithmes d'Estimation et de Détection en Contexte Hétérogène Rang Faible

Thèse présentée et soutenue à Cachan le 23 Novembre 2015, devant le jury composé de :

Besson Olivier	Professeur	Rapporteur
Chevalier Pascal	Professeur	Rapporteur
Comon Pierre	Directeur de Recherche	Examinateur
Chong Chin Yuan	Ingénieur de Recherche	Examinateur
Ginolhac Guillaume	Professeur	Examinateur (Directeur de Thèse)
Pascal Frédéric	Professeur	Examinateur (Encadrant)
Forster Philippe	Professeur	Examinateur (Encadrant)

Laboratoire SONDRA, Supélec

SONDRA

Résumé

Une des finalités du traitement d'antenne est la détection et la localisation de cibles en milieu bruité. Dans la plupart des cas pratiques, comme par exemple pour les traitements adaptatifs RADAR, il faut estimer dans un premier temps les propriétés statistiques du bruit, plus précisément sa matrice de covariance. Dans ce contexte, on formule généralement l'hypothèse de bruit gaussien. Il est toutefois connu que le bruit en RADAR est de nature impulsive et que l'hypothèse gaussienne est parfois mal adaptée. C'est pourquoi, depuis quelques années, le bruit, et en particulier le fouillis de sol, est modélisé par des processus couvrant un panel plus large de distributions, notamment les Spherically Invariant Random Vectors (SIRVs). Dans ce nouveau cadre théorique, la Sample Covariance Matrix (SCM) estimant classiquement la matrice de covariance du bruit entraîne des pertes de performances importantes des détecteurs/estimateurs. Dans ce contexte non-gaussien, d'autres estimateurs (e.g. les M-estimateurs), mieux adaptés à ces statistiques de bruits impulsifs, ont été développés.

Parallèlement, il est connu que le bruit RADAR se décompose sous la forme d'une somme d'un fouillis de rang faible (la réponse de l'environnement) et d'un bruit blanc (le bruit thermique). La matrice de covariance totale du bruit a donc une structure de type rang faible plus identité. Cette information peut être utilisée dans le processus d'estimation afin de réduire le nombre de données nécessaires. De plus, il aussi est possible de construire des traitements adaptatifs basés sur un estimateur du projecteur orthogonal au sous espace fouillis, à la place d'un estimateur de la matrice de covariance. Les traitements adaptatifs basés sur cette approximation nécessitent aussi moins de données secondaires pour atteindre des performances satisfaisantes. On estime classiquement ce projecteur à partir de la décomposition en valeurs singulières d'un estimateur de la matrice de covariance.

Néanmoins l'état de l'art ne présente pas d'estimateurs à la fois robustes aux distributions impulsives, et rendant compte de la structure rang faible des données. C'est pourquoi nos travaux se focalisent sur le développement de nouveaux estimateurs (de covariance et de sous espace fouillis) directement adaptés au contexte considéré. Les contributions de cette thèse s'orientent donc autour de trois axes :

• Nous présenterons le modèle de sources impulsives ayant une matrice de covariance de rang faible noyées dans un bruit blanc gaussien. Ce modèle, fortement justifié dans de nombreuses applications, a cependant peu été étudié pour la problématique d'estimation de matrice de covariance. Le maximum de vraisemblance de la matrice de covariance pour ce contexte n'ayant pas une forme analytique directe, nous développerons différents algorithmes pour l'atteindre efficacement.

• Nous développerons de plus nouveaux estimateurs directs de projecteur sur le sous espace fouillis, ne nécessitant pas un estimé de la matrice de covariance intermédiaire, adaptés au contexte considéré.

• Nous étudierons les performances des estimateurs proposés sur une application de Space Time Adaptative Processing (STAP) pour radar aéroporté, au travers de simulations et de données réelles.

Mots clés : estimation, estimation robuste, matrice de covariance, rang faible, sous espace, algorithmes d'optimisation, détection adaptative, traitement du signal, radar.

Abstract

One purpose of array processing is the detection and location of a target in a noisy environment. In most cases (as RADAR or active SONAR), statistical properties of the noise, especially its covariance matrix, have to be estimated using i.i.d. samples. Within this context, several hypotheses are usually made : Gaussian distribution, training data containing only noise, perfect hardware. Nevertheless, it is well known that a Gaussian distribution doesn't provide a good empirical fit to RADAR clutter data. That is why noise is now modeled by elliptical process, mainly Spherically Invariant Random Vectors (SIRV). In this new context, the use of the SCM (Sample Covariance Matrix), a classical estimate of the covariance matrix, leads to a loss of performances of detectors/estimators. More efficient estimators have been developed, such as the Fixed Point Estimator and M-estimators.

If the noise is modeled as a low-rank clutter plus white Gaussian noise, the total covariance matrix is structured as a low rank plus identity matrix. This information can be used in the estimation process to reduce the number of samples required to reach acceptable performance. Moreover, it is possible to estimate the basis vectors of the clutter-plus-noise orthogonal subspace rather than the total covariance matrix of the clutter, which requires less data and is more robust to outliers. The orthogonal projection to the clutter plus noise subspace is usually calculated from an estimatd of the covariance matrix.

Nevertheless, the state of art does not provide estimators that are both robust to various distributions and low rank structured. In this Thesis, we therefore develop new estimators that are fitting the considered context, to fill this gap. The contributions are following three axes :

- We present a precise statistical model : low rank heterogeneous sources embedded in a white Gaussian noise.We express the maximum likelihood estimator for this context.Since this estimator has no closed form, we develop several algorithms to reach it effitiently.
- For the considered context, we develop direct new clutter subspace estimators that are not requiring an intermediate Covariance Matrix estimate.
- We study the performances of the proposed methods on a Space Time Adaptive Processing for airborne radar application. Tests are performed on both synthetic and real data.

Keywords : estimation, adaptive detection, robust estimation, covariance matrix, low rank, subspace, optimization methods, signal processing, radar.

Table des matières

Introduction **17**

1 Modélisation de bruit hétérogène et estimation de la matrice de covariance : état de l'art **22**
1.1 Définitions et modèle des données 22
1.1.1 Modèle général 22
1.1.2 Définitions générales 23
1.2 Distributions de signaux 25
1.2.1 La distribution gaussienne 25
1.2.2 Les distributions complexes elliptiques symétriques (CES) 25
1.2.3 Le cas particulier des gaussiennes composées (CG) 28
1.2.4 Quelques exemples de distributions complexes elliptiques symétriques 29
1.3 Estimation de la matrice de covariance 31
1.3.1 Propriétés attendues des estimateurs 31
1.3.2 La Sample Covariance Matrix 32
1.3.3 La Normalized Sample Covariance Matrix 32
1.3.4 Le maximum de vraisemblance des distributions complexes elliptiques symétriques 33
1.3.5 Les M-estimateurs 35
1.3.6 Les Estimateurs Robustes Régularisés 37
1.4 Introduction de la problématique considérée 39
1.4.1 Estimation de matrice structurées 39
1.4.2 Matrices structurées rang faible 40
1.4.3 Estimation de covariance à structure rang faible : le cas gaussien 41
1.4.4 Estimation robuste de la matrice de covariance sous contrainte de structure rang faible : un problème ouvert 42
1.4.5 Estimation de la matrice de covariance en contexte hétérogène rang faible : problématique considérée dans cette thèse 42
1.5 Synthèse du chapitre 1 43

2 Estimation de la matrice de covariance en contexte hétérogène rang faible **45**
2.1 Motivations 45
2.2 Modèle 47
2.3 Maximum de vraisemblance de la matrice de covariance du fouillis CG rang faible . . . 49

2.4 Premier algorithme : 2-Step approché . . . 50
2.4.1 Relaxation au travers de variables indépendantes $d_r^k = \tau_k c_r$ . . . 50
2.4.2 Description de l'algorithme et propriétés . . . 50
2.4.3 Étape 1 : Estimation des textures et valeurs propres via régularisation des EMV $\{\hat{d}_r^k\}$ . . . 50
2.4.4 Étape 2 : Estimation du sous-espace fouillis pour textures et valeurs propres fixées 52
2.4.5 Dernière étape : Estimation de facteur d'échelle . . . 53
2.5 Deuxième algorithme : 2-Step exact sous hypothèse de fort rapport fouillis à bruit . . . 54
2.5.1 Seconde relaxation : hypothèse de fort rapport fouillis à bruit . . . 54
2.5.2 Description et propriétés de l'algorithme . . . 54
2.5.3 Étape 1 : Estimation des textures et valeurs propres grâce à la relaxation fort rapport fouillis à bruit . . . 55
2.5.4 Étape 2 : Estimation du sous-espace fouillis pour textures et valeurs propres fixées 57
2.6 Algorithmes Majorization-Minimization . . . 59
2.6.1 Motivations . . . 59
2.6.2 Principe général des algorithmes MM par blocs . . . 59
2.6.3 Algorithme MLE-MM1 - "direct block-MM" . . . 60
2.6.4 Algorithme MLE-MM2 - "Eigenspace block-MM" . . . 61
2.7 Simulations . . . 65
2.7.1 Paramètres . . . 65
2.7.2 Estimateurs considérés . . . 65
2.7.3 Résultats . . . 65
2.8 Synthèse du Chapitre 2 . . . 68

A Preuves du chapitre 2 **73**
A.1 Preuve du Théorème 2.3.1 . . . 73
A.2 Preuve du Théorème 2.4.1 . . . 76
A.3 Preuve du Théorème 2.5.1 . . . 76

B Article : Développement des Algorithmes MM1 et MM2 **78**

3 Estimation de projecteur sur le sous-espace fouillis en contexte hétérogène rang faible **92**
3.1 Motivations . . . 92
3.1.1 L'approximation rang faible et ses motivations . . . 92
3.2 Relaxation sur l'orthogonalité entre sous-espaces : l'heuristique LR-FPE . . . 93
3.3 Relaxation sur les valeurs propres : estimateur AEMV . . . 95
3.3.1 Densité de probabilité de textures connue . . . 95
3.3.2 Densité de probabilité de textures inconnue . . . 95
3.3.3 Interprétations de AEMV . . . 97
3.4 AEMV sous hypothèse de données contaminées . . . 98
3.4.1 Problème de robustesse à la contamination : un bref état de l'art . . . 98
3.4.2 Estimateur AEMV modifié . . . 99
3.5 Simulations . . . 101

3.5.1 Paramètres . . . 102
3.5.2 Résultats . . . 102
3.6 Synthèse du chapitre 3 . . . 105

C Preuves du chapitre 3 **109**
C.1 Preuve du théorème 3.3.1 . . . 109

4 Application au radar STAP **112**
4.1 Présentation du système . . . 112
4.1.1 Présentation du radar . . . 113
4.1.2 Modèle des signaux . . . 114
4.2 Application basée sur l'estimation de la matrice de covariance : détection . . . 118
4.2.1 Problème considéré . . . 118
4.2.2 Résultats de Simulations . . . 118
4.2.3 Résultats sur données réelles . . . 120
4.3 Application basée sur l'estimation du sous-espace fouillis : filtrage rang faible . . . 125
4.3.1 Problème considéré . . . 125
4.3.2 Résultats de simulations . . . 127
4.3.3 Résultats sur données réelles . . . 129
4.4 Synthèse du chapitre 4 . . . 130

Conclusion et perspectives **137**

Table des figures

2.1 Évolution de la vraisemblance négative (valeur opposée de la vraisemblance) en fonction des itérations pour les différents algorithmes "EMV". $M = 60$, $R = 15$, $K = 100$, $\nu = 1$. Haut : CNR= 10, bas : CNR= 30dB. 69
2.2 NMSE sur $\mathbf{\Sigma}_{tot}$ des différents estimateurs. $M = 60$, $R = 15$, $\nu = 1$. De haut en bas : CNR= $[3, 10, 30]$dB. 70
2.3 NMSE sur $\mathbf{\Sigma}_{tot}$ des différents estimateurs. $M = 60$, $R = 15$, $\nu = 0.1$. De haut en bas : CNR= $[3, 10, 30]$dB. 71

3.1 NMSE sur $\mathbf{\Pi}_c$ des estimateurs EMV obtenus par les différents algorithmes dérivés au chapitre 2. $M = 60$, $R = 15$, $\nu = 0.1$, CNR= 10. 102
3.2 NMSE sur $\mathbf{\Pi}_c$ des différents estimateurs considérés. $M = 60$, $R = 15$, $\nu = 1$. De haut en bas : CNR= $[3, 10, 30]$dB. 106
3.3 NMSE sur $\mathbf{\Pi}_c$ des différents estimateurs considérés. $M = 60$, $R = 15$, $\nu = 0.1$. De haut en bas : CNR= $[3, 10, 30]$dB. 107

4.1 Principe et intérêt du filtrage spatio-temporel. 113
4.2 Géométrie du STAP. 113
4.3 Data Cube STAP et arrangement de données. 114
4.4 Gauche : probabilité de fausse alarme en fonction du seuil. Droite : probabilité de détection en fonction du SNR pour PFA= 10^{-3}. $K = NM + 1$, $\nu = 1$, CNR= 30dB. 120
4.5 Gauche : probabilité de fausse alarme en fonction du seuil. Droite : probabilité de détection en fonction du SNR pour PFA= 10^{-3}. $K = NM + 1$, $\nu = 0.1$, CNR= 30dB. 120
4.6 Gauche : probabilité de fausse alarme en fonction du seuil. Droite : probabilité de détection en fonction du SNR pour PFA= 10^{-3}. $K = 3R$, $\nu = 1$, CNR= 30dB. 121
4.7 Gauche : probabilité de fausse alarme en fonction du seuil. Droite : probabilité de détection en fonction du SNR pour PFA= 10^{-3}. $K = 3R$, $\nu = 0.1$, CNR= 30dB. 121
4.8 Gauche : probabilité de fausse alarme en fonction du seuil. Droite : probabilité de détection en fonction du SNR pour PFA= 10^{-3}. $K = 3R$, $\nu = 1$, CNR= 10dB. 121
4.9 Gauche : probabilité de fausse alarme en fonction du seuil. Droite : probabilité de détection en fonction du SNR pour PFA= 10^{-3}. $K = 3R$, $\nu = 0.1$, CNR= 10dB. 122
4.10 Probabilité de fausse alarme en fonction du seuil pour l'estimateur EMV calculé avec différents rangs R autour de la vraie valeur $R = 15$. $K = NM + 1$, $\nu = 1$, CNR= 30dB. 122
4.11 Sortie des détecteurs, de gauche à droite : Λ_{SCM}, Λ_{RCML}, Λ_{SFPE} and Λ_{EMV}. $K = 100$. 124

4.12 Sortie des détecteurs, de gauche à droite : Λ_{SCM}, Λ_{RCML}, Λ_{SFPE} and Λ_{EMV}. $K = 300$. 125
4.13 Sortie des détecteurs, de gauche à droite : Λ_{SCM}, Λ_{RCML}, Λ_{SFPE} and Λ_{EMV}. $K = 100$. Une donnée corrompue par la cible. 126
4.14 SINR-Loss moyen des filtres rang faible construits à partir de différents estimateurs de projecteurs, en fonction de K. De haut en bas : CNR = 10dB, CNR = 20dB, CNR = 30dB. Fouillis modérément hétérogène $\nu = 1$. 131
4.15 SINR-Loss moyen des filtres rang faible construits à partir de différents estimateurs de projecteurs, en fonction de K. De haut en bas : CNR = 10dB, CNR = 20dB, CNR = 30dB. Fouillis fortement hétérogène $\nu = 0.1$. 132
4.16 SINR-Loss moyen des filtres rang faible construits à partir de différents estimateurs de projecteurs, en fonction du pourcentage de données corrompues $K = 100$ avec une cible de puissance ONR= 10dB et CNR= 20dB. Haut : $\nu = 1$, bas $\nu = 0.1$. 133
4.17 SINR-Loss des filtres rang faible construits à partir de différents estimateurs de projecteurs, en fonction de de la puissance de l'outlier pour $K = 100$, $\nu = 0.1$ et CNR= 20dB. 134
4.18 Sortie des filtres rang faible adaptatifs sur les données STAP CELAR. $K = 400$ données. 134
4.19 Sortie des filtres rang faible adaptatifs sur les données STAP CELAR. $K = 400$ données, 2 données contaminées. 135

Acronymes

2S	2-Step
AEMV	Estimateur du Maximum de Vraisemblance Approché
CES	Complex Elliptically Symmetric
CG	Compound Gaussian
CNR	Clutter to Noise Ratio
EMV	Estimateur du Maximum de Vraisemblance
FPE	Fixed Point Estimator
LR-FPE	Low Rank Fixed Point Estimator
MM	Majorization-Minimization
MUSIC	MUltiple SIgnal Classification
MVCES	Maximum de Vraisemblance des distributions CES
NMSE	Normalized Mean Square Error
NSCM	Normalized Sample Covariance Matrix
R-AEMV	Estimateur du Maximum de Vraisemblance Approché Robuste
RCML	Rank Constrained Maximum Likelihood
SAR	Synthetic Aperture Radar
SCM	Sample Covariance Matrix
SFPE	Shrinkage Fixed Point Estimator
SINR	Signal to Interference plus Noise Ratio
SINR-Loss	Loss in Signal to Interference plus Noise Ratio
SIRV	Spherically Invariant Random Vector
STAP	Space-Time Adaptive Process
SVD	Singular Value Decomposition

Symboles généraux

$\mathbb{R}$	Ensemble des nombres réels.
$\mathbb{C}$	Ensemble des nombres complexes.
$\mathbb{S}_M^+$	Ensemble des matrices hermitiennes semi-définies positives.
a	représente un scalaire.
$\mathbf{a}$	représente un vecteur.
$\mathbf{A}$	représente une matrice.
$\mathbf{I}_M$	est la matrice identité de taille $M \times M$.
$[\![1, R]\!]$	Ensemble des entiers de 1 à R.
$\{a_r\}$	est l'ensemble des éléments a_r. Par convention et si non précisé, l'indice maximum est dénoté par la même lettre en majuscule, soit $r \in [\![1, R]\!]$.
^{T}	représente l'opérateur transposée.
^{H}	représente l'opérateur hermitien.
$\delta_{i,j}$	représente le symbole de Kronecker : $\delta_{i,j} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{sinon} \end{cases}$
$\otimes$	représente le produit de Kronecker.
vec	est l'opérateur qui transforme une matrice $m \times n$ en un vecteur de taille mn, en concaténant ces n colonnes en une seule colonne.
diag	est l'opérateur qui transforme un vecteur de taille m en une matrice diagonale $m \times m$ dont les éléments diagonaux sont ceux du vecteur.
$\mathbb{U}(.)$	est l'opérateur de concaténation de vecteurs, transformant r vecteurs colonnes de taille m en une matrice $m \times r$.
Tr	représente l'opérateur trace.
$\|\|.\|\|$	indique une norme.
$\sim$	signifie "distribué selon".
$\mathcal{CN}$	distribution gaussienne complexe.
$\mathcal{CG}$	distribution gaussienne composée complexe.
$\mathcal{CE}$	distribution elliptique symétrique complexe.
$(A \mid B)$	signifie "A sachant B".
$\mathbb{P}(a)$	probabilité de l'évènement a.
$\mathbb{E}[.]$	désigne l'espérance mathématique.
$\overset{d}{=}$	signifie "a la même distribution que".
$\xrightarrow{d}$	représente la convergence en distribution.
$\xrightarrow{\mathbb{P}}$	représente la convergence en probabilité.

Introduction

Une des finalités globales du Traitement du Signal est d'extraire de l'information d'un ensemble de mesures. Ces mesures, les signaux, peuvent provenir de multiples sources, comme par exemple, un récepteur de télécommunications, un système d'imagerie, une station météorologique, le cours de la bourse... Selon l'information recherchée au travers de ces signaux, on parle de différentes problématiques et de leurs traitements respectifs. Pour ne citer qu'eux : la détection vise à déterminer si un signal est présent ou non, l'estimation consiste à déterminer les paramètres d'intérêt d'un signal, la séparation de sources permet de démélanger une somme de signaux différents...

La complexité des systèmes actuels fait que l'extraction de l'information recherchée ne peut plus se résumer à la simple inversion d'un modèle mathématique, aussi sophistiqué soit-il. En effet, les observations, de plus en plus nombreuses, fines et dépendantes de nombreux paramètres, font que l'observateur se heurte à des problèmes de fluctuations de mesures, d'incertitudes de modèle, ou encore, de parties du signal inconnues... Afin de prendre en compte ces incertitudes, désormais inhérentes aux signaux observés, on utilise généralement un modèle de signal probabiliste. Parallèlement, les mesures obtenues étant généralement une série d'échantillons, on considère donc les signaux reçus comme étant des vecteurs aléatoires, ce qui explique pourquoi une grande partie des techniques de Traitement du Signal moderne relèvent du domaine de l'analyse statistique multivariée. De manière très générale, le problème considéré est alors d'étudier la statistique des signaux observés, et c'est au travers de cette étude que l'on peut réaliser les traitements voulus.

Dans cette thèse, nous nous intéresserons plus particulièrement à l'estimation de la matrice de covariance. Celle-ci joue, en effet, un rôle récurrent et prépondérant dans les traitements actuels, qui nécessitent la connaissance de la statistique d'ordre deux du milieu observé afin d'atteindre des performances "optimales" (au sens de divers critères). En pratique, la matrice de covariance du signal n'est pas connue et doit être estimée à partir de plusieurs réalisations indépendantes du même vecteur aléatoire. Remplacer la vraie matrice de covariance par son estimée dans les traitements donne lieu aux traitements dits adaptatifs, tels que le Filtrage ou la Détection. Remarquons aussi qu'en dehors du cas adaptatif, l'estimation de la matrice de covariance peut représenter une part inhérente au traitement, c'est à dire que ce paramètre contient lui même l'information que l'on cherche. C'est le cas dans des applications comme l'estimation de direction d'arrivée ou l'Analyse en Composantes Principales. Quoi qu'il en soit, les performances de tous ces traitements reposent directement sur la précision de l'estimation de la matrice de covariance.

La qualité d'estimation de la matrice de covariance est bien entendu dépendante de la véracité du modèle que l'on associe aux données obtenues. Classiquement, c'est l'hypothèse de la statistique gaussienne qui est utilisée. L'omniprésence de cette hypothèse s'est justifiée tout d'abord d'un point de vue pratique : la loi normale a longtemps été une bonne approximation à de nombreuses observations. D'un point de vue plus fondamental, le Théorème de la limite centrale vient de plus renforcer la légitimité de ce modèle. Considérer les variables comme gaussiennes conduit alors à la Sample Covariance Matrix (SCM), Maximum de Vraisemblance pour ce cas, comme estimateur de la matrice de covariance. Cet estimateur classique permet en effet d'atteindre des résultats satisfaisants dans bon nombre de cas. C'est pourquoi l'étude et l'utilisation de la SCM a prédominé dans la littérature du Traitement du Signal et que celle-ci reste un outil majeur encore aujourd'hui.

Cependant, à l'heure de la haute résolution et de la diversification des données récoltées, la loi normale peut s'avérer aujourd'hui être une mauvaise approximation de la physique sous-jacente aux observations. De nombreux travaux et campagnes de mesures ont montré que l'hypothèse gaussienne plus vérifiée, et viennent appuyer le fait que l'on ne peut disposer d'un modèle statistique unique, adapté à toutes les situations. La SCM est connue pour être un estimateur peu adapté aux contextes de distributions à queue lourde, ainsi qu'a la présence de données aberrantes (outliers). C'est pourquoi les traitements construits autour de cet estimateur peuvent conduire à de mauvaises performances dans des contextes non gaussiens.

Pour répondre à cette problématique, le cadre de travail de l'estimation robuste a attiré considérablement d'attention ces dernières années. Ces travaux se concentrent notamment sur le modèle des distributions elliptiques symétriques : une famille de distribution englobant la plupart des distributions usuelles (normale, t-distribution, K-distribution, Weibull...), pouvant modéliser des distributions à queue lourde. Dans ce cadre, les M-estimateurs (Maximums de Vraisemblance généralisés) offrent une solution robuste sur l'ensemble de ces distributions, mais aussi robuste à de potentiels outliers. Néanmoins, les M-estimateurs nécessitent un grand nombre d'échantillons, supérieur à la taille des données pour pouvoir être calculés. Ceci pose actuellement problème car de nombreuses applications utilisent des données de grande dimension et font face à un nombre d'échantillons limité.

Dans certains cas, des considérations en amont sur le système étudié (par exemple, sa géométrie) permettent de dégager un a priori sur la structure de la matrice de covariance (Toeplitz, Persymétrie...). Prendre en compte cette structure permet alors de réduire les degrés de liberté lié au problème de son estimation, donc de réduire théoriquement le nombre de données nécessaires à une estimation précise. C'est pourquoi de nombreuses méthodes d'estimation de paramètres ou de régularisation ont été développées pour répondre à la problématique d'estimation de covariance structurée. Initialement développées pour le contexte gaussien, ces méthodes s'étendent peu à peu au cadre plus complexe des M-estimateurs afin d'obtenir des estimateurs alliant robustesse et performance avec peu de données. Cependant, ces travaux ne sont pas terminés et c'est dans ce cadre que se placent les contributions de cette thèse.

Dans ce travail, nous nous intéresserons particulièrement à des signaux étant contenus dans un sous-espace de dimension inférieure aux données. Cette hypothèse est en effet vérifiée dans de nombreuses applications où le nombre de sources observées est restreint (estimation de direction d'arrivée, imagerie

hyperspectrale...), ainsi que pour les mesures de fouillis radar (la réponse de l'environnement au signal émis). La matrice de covariance de ce type de sources, potentiellement hétérogènes, présente alors une structure de rang faible. C'est pourquoi cette Thèse considèrera le problème d'estimation robuste de matrices de covariances satisfaisant cette structure, une problématique encore non couverte par l'état de l'art. Dans ce contexte, nous verrons notamment que l'estimation du sous-espace contenant les signaux joue un rôle clé. Ce paramètre étant de plus utilisé dans certaines applications, qui ne nécessitent pas une estimation de la matrice de covariance totale, une partie des travaux présentés considèrera donc spécifiquement la problématique de l'estimation ce sous-espace.

Ce mémoire est construit en quatre chapitres et s'organise comme suit. Le chapitre 1 présente dans un premier temps les distributions elliptiques symétriques (modèles de bruits non gaussiens) ainsi que les estimateurs robustes de la matrice de covariance. Il décrit ensuite un état de l'art (non exhaustif) de l'estimation structurée de la matrice de covariance. Cette introduction permet alors d'introduire la problématique considérée : celle de l'estimation de la matrice de covariance en contexte hétérogène rang faible. Les contributions de cette thèse se concentrent ensuite autour de 3 axes, chacun exploré par un chapitre :

- Chapitre 2, nous présenterons tout d'abord un modèle statistique précis : celui de sources hétérogènes ayant une matrice de covariance rang faible noyées dans un bruit blanc gaussien. Ce modèle est, par exemple, fortement justifié pour des applications de type radar. Il a cependant été peu étudié pour la problématique d'estimation de matrice de covariance. Nous développerons donc l'expression du maximum de vraisemblance de la matrice de covariance pour ce contexte. Cette expression n'ayant pas une forme analytique simple, différents algorithmes seront proposés et analysés pour construire efficacement cet estimateur. Nous proposerons dans un premier temps deux algorithmes basés sur la maximisation d'une fonction vraisemblance relaxée. Enfin, nous proposerons un méthode de résolution exacte via des algorithmes de type Majorization-Minimisation. L'ensemble des estimateurs issus des algorithmes développés sera ensuite comparé à l'état de l'art au travers de simulations de validation.
- Chapitre 3, nous nous focaliserons sur l'estimation du projecteur sur le sous-espace engendré par les vecteurs propres dominants de la matrice de covariance. En effet pour le modèle considéré, il est possible de décomposer le signal sur deux sous-espaces orthogonaux à l'aide de la décomposition en valeurs singulières de cette matrice de covariance. Les projecteurs orthogonaux sur chacun de ces sous-espaces peuvent alors être construits, permettant de développer des méthodes dites de "rang faible" (ou "à sous-espaces") telles que l'annulation d'interférence. Il a été montré que ces méthodes permettent d'obtenir des performances équivalentes à celles des traitements classiques tout en réduisant significativement le nombre d'échantillons nécessaires. Nous développerons en conséquence de nouveaux estimateurs de projecteur adaptés au contexte considéré. Ces estimateurs seront développés en considérant diverses relaxation et modification du modèle initialement présenté chapitre 2. Ces considérations permettront, d'une part, de dériver des algorithmes simplifiés, atteignant tout de même des performances similaires au maximum de vraisemblance en terme d'estimation de sous-espace. D'autre part, nous considérerons aussi le problème d'estimation robuste de ce sous-espace en présence de données corrompues par des outliers. Certaines modifications du modèle envisagées permettront, en effet, de proposer des estimateurs adaptés (offrant un bon compromis performances-robustesse) à cette problématique. L'ensemble des esti-

mateurs développés sera ensuite comparé à l'état de l'art au travers de simulations de validation.

– Le dernier axe, présenté chapitre 4, consistera en l'étude des performances des estimateurs proposés et de l'état de l'art sur une application de Space Time Adaptative Processing (STAP) pour radar aéroporté, au travers de simulations et de données réelles. Les données reçues dans cette application suivent en effet le modèle considéré au long de cette thèse car il est connu que la matrice de covariance du fouillis est de rang faible (ce rang peut être évalué grâce à la formule de Brennan). Nous illustrerons les performances des estimateurs de matrice de covariance développés chapitre 2 sur une problématique de détection adaptative. Les performances des estimateurs sur le sous-espace fouillis développés chapitre 3 seront étudiées au travers d'une application de filtrage adaptatif rang faible. Afin de compléter cette étude, nous illustrerons de plus la robustesse des méthodes développées à des erreurs de modèle (mauvaise estimation du rang, présence d'outliers...).

Enfin, nous conclurons par un synthèse générale de ces travaux et dégagerons les diverses perspectives de recherche qu'ils suggèrent.

Chapitre 1

Modélisation de bruit hétérogène et estimation de la matrice de covariance : état de l'art

Le but de ce chapitre est d'introduire les concepts et outils qui serviront la problématique étudiée dans cette thèse. Dans un premier temps, nous rappellerons quelques définitions importantes, ainsi que le modèle des distributions CES, permettant de couvrir de nombreuses distributions multivariées usuelles. Nous nous proposons ensuite de récapituler les principales méthodes existantes, pour estimer les statistiques du second ordre des données reçues. Nous introduirons notamment les estimateurs robustes, naturellement adaptés au contexte des distributions CES. Puis, nous dresserons un bref état de l'art (non exhaustif) des méthodes d'estimation de matrices de covariance structurées. Cet état de l'art sera l'occasion de mettre en avant la problématique considérée, qui se situe à l'intersection des deux thèmes abordés que sont l'estimation robuste et l'estimation structurée.

1.1 Définitions et modèle des données

1.1.1 Modèle général

Comme évoqué précédemment, dans les applications de traitement de signal on s'intéresse très souvent à la distribution des données reçues. De manière très générale, les données sont modélisées par des vecteurs complexes $\mathbf{z}$ de taille M, $\mathbf{z} \in \mathbb{C}^M$:

$$\mathbf{z} = \mathbf{s} + \mathbf{n} \tag{1.1}$$

avec $\mathbf{s}$ le signal cible (ici considéré comme déterministe) et $\mathbf{n}$ le bruit additif (aléatoire) résultant de fluctuations de mesures ou servant à rendre compte des incertitudes de modèle. Selon l'application, diverses hypothèses sont faites sur ces deux quantités. Nous pouvons citer quelques exemples :

– Détection : Le problème de détection est de dire si oui ou non, un signal d'intérêt $\mathbf{s}$ est présent dans la mesure $\mathbf{z}$. Le problème est généralement exprimé sous la forme d'un test d'hypothèse

binaire :

$$\begin{cases} H_0 : \mathbf{z} = \mathbf{n} \\ H_1 : \mathbf{z} = \mathbf{s} + \mathbf{n} \end{cases} \tag{1.2}$$

- Estimation de paramètres du signal : dans ce type d'application, on dispose d'un modèle sur le signal $\mathbf{s}$, généralement dépendant d'un vecteur de paramètres $\boldsymbol{\theta}$ et noté $\mathbf{s}(\boldsymbol{\theta})$. Le problème est alors de retrouver ces paramètres au travers de mesures de $\mathbf{s}$ corrompue par le bruit $\mathbf{n}$.
- Estimation de paramètres du bruit : dans cette application, c'est le bruit $\mathbf{n}$ qui dépend des paramètres $\boldsymbol{\theta}$ et est noté $\mathbf{n}(\boldsymbol{\theta})$. Le problème est alors de retrouver ces paramètres au travers d'observations du bruit $\mathbf{n}$. Dans ce cas de figure, on considère le plus souvent qu'aucun signal ne vient perturber la mesure de ce bruit, soit $\mathbf{s} = \mathbf{0}$.

Avant d'aller plus loin, la section suivante définira quelques termes généraux dont nous aurons besoin pour poursuivre. Nous en profiterons pour poser dès à présent quelques hypothèses sur les types de signaux que nous considérerons au cours de cette thèse.

1.1.2 Définitions générales

Définition 1.1.2.1 *Moyenne*
La moyenne $\boldsymbol{\mu}$ de la variable aléatoire considérée $\mathbf{z}$ est définie comme son espérance mathématique :

$$\boldsymbol{\mu} = \mathbb{E}(\mathbf{z}) \tag{1.3}$$

elle est aussi appelée moment d'ordre 1, ou statistique d'ordre 1.

Dans de nombreuses applications la moyenne de la variable aléatoire observée est tout simplement considérée comme nulle : $\boldsymbol{\mu} = \mathbf{0}$. C'est d'ailleurs l'hypothèse que nous ferons tout au long de cette thèse car elle est vérifiée dans de nombreux cas pratiques, notamment pour les signaux ondulatoires (ondes sonores, électromagnétiques ...). Cette hypothèse est aussi équivalente à celle de la moyenne connue, qui amène alors à considérer directement la variable recentrée $\mathbf{z}' = \mathbf{z} - \boldsymbol{\mu}$.

Cependant, notons tout de même que les extensions de résultats au cas de la moyenne non nulle ne sont pas toujours évidents et nécessitent un travail particulier [39].

Nous avons évoqué précédemment, les statistiques du second ordre des données reçues. De manière plus concrète, il s'agit de la matrice de covariance et de la matrice de pseudo-covariance. Avec ces deux paramètres, on est en mesure de décrire complètement les statistiques du second ordre d'un vecteur aléatoire complexe. Notons $\mathbb{S}_+^M$ l'ensemble des matrices hermitiennes semi-définies positives de taille $M \times M$.

Définition 1.1.2.2 *Matrice de Covariance*
La Matrice de Covariance $cov(\mathbf{z}) \in \mathbb{S}_+^M$ du vecteur aléatoire centré $\mathbf{z} \in \mathbb{C}^M$ est définie par :

$$cov(\mathbf{z}) = \mathbb{E}\left[\mathbf{z}\mathbf{z}^H\right] \tag{1.4}$$

Définition 1.1.2.3 *Matrice de Pseudo-Covariance*
La Matrice de Pseudo-Covariance $pcov(\mathbf{z})$ du vecteur aléatoire centré $\mathbf{z} \in \mathbb{C}^M$ est définie par :

$$pcov(\mathbf{z}) = \mathbb{E}\left[\mathbf{z}\mathbf{z}^T\right] \tag{1.5}$$

Une notion importante est celle de la propriété de circularité des variables :

Définition 1.1.2.4 *Symétrie circulaire*
Un vecteur $\mathbf{z} \in \mathbb{C}^M$ *est dit circulaire si*

$$\mathbf{z} \overset{d}{=} e^{j\theta}\mathbf{z} \ \ \forall \theta \in \mathbb{R} \tag{1.6}$$

Définition 1.1.2.5 *Circularité au second-ordre*
Un vecteur $\mathbf{z} \in \mathbb{C}^M$ *est dit circulaire du second ordre si pcov*$(\mathbf{z}) = \mathbf{0}$

La circularité du second-ordre est une hypothèse très souvent exploitée en traitement de signal. En effet, le bruit additif est sans trop d'erreur, communément modélisé par une distribution circulaire. De plus, de nombreux signaux complexes synthétisés en communication sans fil ou en traitement d'antenne ont des propriétés de symétrie circulaire.

Dans ce document, les signaux seront toujours considérés comme circulaires. Pour se rapprocher des termes habituellement utilisés, nous utiliserons d'ailleurs simplement le terme de "circularité" lorsqu'il s'agit de circularité du second-ordre.

Afin de ne rien négliger, notons toutefois qu'il existe de nombreux exemples de signaux non-circulaires dans la littérature [103]. Prendre en compte cette non-circularité peut permettre d'améliorer considérablement les performances dans certaines applications [1], [21], [6]. Par ailleurs, des tests de circularité ont été établis pour vérifier si les signaux observés satisfont cette hypothèse : [5] chapitre 2, [78], [81]

1.2 Distributions de signaux

Hormis le modèle de signal considéré, c'est bien la distribution de ce signal qui joue un rôle clé puisque c'est elle qui conditionnera les méthodes à employer pour mener à bien le traitement désiré. Le choix de la distribution associée au bruit est donc crucial.

Classiquement, on utilise la distribution gaussienne, ou loi normale. Néanmoins, comme évoqué précédemment, cette modélisation peut s'avérer être une mauvaise approximation de la statistique des signaux observés. C'est pourquoi les distributions CES (famille très générale englobant la loi normale) ont attiré beaucoup d'intérêt dans la communauté de Traitement du Signal.

1.2.1 La distribution gaussienne

Définition 1.2.1.1 *Loi gaussienne complexe multivariée (loi normale)*
Le vecteur $\mathbf{z} \in \mathbb{C}^M$ a une distribution gaussienne (ou normale) si sa densité de probabilité f s'écrit :

$$f(\mathbf{z}) = \frac{1}{\pi^M |\boldsymbol{\Sigma}|} \exp\left(- (\mathbf{z} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \boldsymbol{\mu}) \right) \tag{1.7}$$

où $\boldsymbol{\mu} \in \mathbb{C}^M$ est la moyenne statistique et $\boldsymbol{\Sigma} \in \mathbb{S}_+^M$ est la Matrice de Covariance. On notera alors :

$$\mathbf{z} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{1.8}$$

Comme le nom du paramètre l'indique, on vérifie bien la propriété suivante :

Théorème 1.2.1 *Matrice de Covariance d'un vecteur gaussien*
Soit un vecteur centré $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$, alors :

$$cov(\mathbf{z}) = \boldsymbol{\Sigma} \tag{1.9}$$

1.2.2 Les distributions complexes elliptiques symétriques (CES)

Les distributions elliptiques symétriques (CES) ont été introduites par Kelker [64] et étudiées en profondeur par des auteurs tels que Cambanis [20] et Fang [36]. Deux très bonnes synthèses sont données par les thèses de Frahm [38] et Mahot [71]. L'article [84] propose aussi une excellente synthèse et illustrent l'intérêt de considérer ce modèle dans des applications de traitement d'antenne (radar, sonar, localisation de source...).

La famille des distributions CES est très appréciée par les statisticiens, pour sa grande flexibilité et le fait qu'elle permette de représenter une multitude de distributions particulières, tout en ne dépendant que de peu de paramètres. Ces distributions permettent, en effet, de conserver une formalisation de la densité de probabilité similaire à - ou "généralisant" - la loi normale. Plus concrètement, les densités de probabilités considérées auront une forme de type :

$$\begin{array}{rcl} \text{Densité de probabilité gaussienne} & \rightarrow & \text{Densité de probabilité elliptique symétrique} \\ f(\mathbf{z}) \propto |\boldsymbol{\Sigma}|^{-1} \exp\left(-(\mathbf{z}-\boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{z}-\boldsymbol{\mu})\right) & \rightarrow & f(\mathbf{z}) \propto |\boldsymbol{\Sigma}|^{-1} g\left((\mathbf{z}-\boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{z}-\boldsymbol{\mu})\right) \end{array}$$

avec g une fonction appelée générateur de densité. S'affranchir de l'exponentielle, inhérente à la gaussienne, au profit d'une fonction générique g permet alors de modéliser des distributions à queues plus ou moins lourdes. Cette généralisation nécessite cependant plus de précisions pour être utilisable en pratique, ce qui sera l'objet des théorèmes de représentation détaillés dans cette section.

On rappelle que l'on considèrera uniquement des variables aléatoires circulaires (cf. définition 1.1.2.5). Les variables CES ne sont pas nécessairement circulaires, mais prendre en compte cette propriété conduit à une formalisation moins compacte et ne sert pas directement le propos de cette thèse. Nous renvoyons donc le lecteur aux références [84] [71] pour de plus amples précisions à ce sujet.

Définition 1.2.2.1 *Fonction caractéristique d'un vecteur aléatoire suivant une distribution CES*
Un vecteur aléatoire $\mathbf{z} \in \mathbb{C}^M$ à une distribution CES si sa fonction caractéristique est de la forme :

$$\Phi_{\mathbf{z}}(\mathbf{c}) = \exp\left(j\Re\left(\mathbf{c}^H\boldsymbol{\mu}\right)\right)\Phi\left(\mathbf{c}^H\boldsymbol{\Sigma}\mathbf{c}\right) \tag{1.10}$$

pour une fonction $\Phi : \mathbb{R}^+ \to \mathbb{R}$ appelée générateur caractéristique, une matrice $\boldsymbol{\Sigma} \in \mathbb{S}_+^M$ appelée matrice de dispersion et un vecteur $\boldsymbol{\mu} \in \mathbb{C}^M$ appelé centre de symétrie. On notera :

$$\mathbf{z} \sim \mathcal{CE}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \Phi\right) \tag{1.11}$$

Ce type de distribution peut être relié à une représentation plus "concrète" au travers du théorème suivant [20] :

Théorème 1.2.2 *Théorème de Représentation Stochastique*
Un vecteur aléatoire $\mathbf{z} \sim \mathcal{CE}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \Phi\right)$, avec $\boldsymbol{\Sigma}$ de rang $R \leq M$ peut toujours être représenté par :

$$\mathbf{z} \overset{d}{=} \boldsymbol{\mu} + \mathcal{R}\mathbf{A}\mathcal{U}^{(R)} \tag{1.12}$$

où $\mathcal{U}^{(R)} \in \mathbb{C}^N$ est un vecteur aléatoire et de distribution uniforme sur l'hypersphère unité $\mathcal{S}_{\mathbb{C}}^{R-1}$, avec :

$$\mathcal{S}_{\mathbb{C}}^{R-1} = \left\{\mathbf{x} \in \mathbb{C}^N \ : \ ||\mathbf{x}|| = 1\right\} \tag{1.13}$$

$\mathcal{R}$ est une variable aléatoire positive indépendante de $\mathcal{U}^{(R)}$, et $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^H$ est une factorisation de $\boldsymbol{\Sigma}$ telle que $\mathbf{A} \in \mathbb{C}^{M\times R}$.

Tout d'abord, remarquons que les distributions CES ne sont pas définies de manière unique. En effet, considérons les couple $(\boldsymbol{\Sigma}, \Phi)$ et $(\boldsymbol{\Sigma}', \Phi')$ tels que $\boldsymbol{\Sigma}' = \gamma^2\boldsymbol{\Sigma}$ et $\Phi'(t) = \Phi(t/\gamma^2)$ pour $\gamma \in \mathbb{R}^+$ quelconque : il y a un problème d'identifiabilité puisque ces deux couples définissent une même distribution. En termes de représentation stochastique, l'ambiguïté d'échelle se visualise aisément en remarquant qu'une constante γ peut être absorbée soit par la variable $\mathcal{R}$, soit par le factorisation $\mathbf{A}$. Afin de lever cette ambiguïté, il convient donc de fixer une convention préalable. Les conventions les plus répandues étant les normalisations $\mathbb{E}(\mathcal{R}^2) = 1$ où $\mathrm{Tr}(\boldsymbol{\Sigma}) = M$ (parfois $|\boldsymbol{\Sigma}| = 1$).

Le théorème de représentation stochastique permet une compréhension plus intuitive des distributions CES : une réalisation peut se voir comme paramétrée par une direction et un module, tout deux aléatoires et indépendants. Cette représentation offre de plus un moyen simple de générer des vecteurs

suivant une distribution CES. La distribution uniforme sur l'hypersphère peut s'obtenir depuis un vecteur gaussien $\mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_R)$ au travers le la relation $\mathcal{U}^{(R)} \overset{d}{=} \mathbf{y}/||\mathbf{y}||_2$. La densité de probabilité de la variable aléatoire $\mathcal{R}$, notée $f_{\mathcal{R}}$, est appelée fonction génératrice de distribution. C'est cette fonction qui caractérise une famille particulière de distributions CES. Elle peut être reliée directement au générateur caractéristique Φ [84]. Comme nous allons le voir, $f_{\mathcal{R}}$ peut aussi être directement reliée à la densité de probabilité de la variable $\mathbf{z}$, si celle ci existe. En effet, de la définition au travers de la Fonction Caractéristique , il ne ressort pas directement que la variable $\mathbf{z}$ ait une densité de probabilité. Cette dernière existe si $\mathbf{\Sigma}$ est de rang plein et que la variable $\mathcal{R}$ est absolument continue. Dans ce cas, on a la définition suivante :

Définition 1.2.2.2 *Densité de probabilité d'une variable CES*
Soit $\mathbf{z}$ une variable CES, sa densité de probabilité (si elle existe) est de la forme :

$$f_{\mathbf{z}}(\mathbf{z}) = c_{M,g}|\mathbf{\Sigma}|^{-1} g\left((\mathbf{z}-\boldsymbol{\mu})^H \mathbf{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu})\right) \tag{1.14}$$

avec g, la fonction appelée générateur de densité, satisfaisant :

$$\delta_{M,g} = \int_0^\infty t^{M-1} g(t) dt < \infty \tag{1.15}$$

qui assure que f soit intégrable. $c_{M,g}$ est une constante de normalisation, de sorte que f définisse bien une densité de probabilité.

$$c_{M,g} = 2\left(s_m \delta_{M,g}\right)^{-1} \tag{1.16}$$

où $s_m = 2\pi^M/\Gamma(M)$ est l'aire de l'hypersphère $\mathcal{S}_{\mathbb{C}}^{M-1}$. On notera :

$$\mathbf{z} \sim \mathcal{CE}(\boldsymbol{\mu}, \mathbf{\Sigma}, g) \tag{1.17}$$

Cette définition au travers de la densité de probabilité se relie directement au théorème de représentation stochastique par la variable $\mathcal{Q} = \mathcal{R}^2$ de densité de probabilité $f_{\mathcal{Q}}$:

$$f_{\mathcal{Q}}(t) = t^{M-1} g(t) \delta_{M,g}^{-1} \tag{1.18}$$

On remarque que $f_{\mathbf{z}}$ ne dépend de $\mathbf{z}$ qu'au travers de la forme quadratique $(\mathbf{z}-\boldsymbol{\mu})^H \mathbf{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu})$. Par conséquent, les lignes de niveau de $f_{\mathbf{z}}$ sont des ellipsoïdes, ce qui explique le nom de distributions "elliptiques".

Pour conclure cette série de définitions, la matrice de covariance d'un vecteur CES fait l'objet du théorème suivant [84] :

Théorème 1.2.3 *Matrice de covariance d'un vecteur CES*
Soit un vecteur aléatoire centré $\mathbf{z} \sim \mathcal{CE}(\mathbf{0}, \mathbf{\Sigma}, g)$, avec $\mathbf{\Sigma}$ de rang $R < M$. Si $\mathbb{E}(\mathcal{R}^2) < \infty$, alors la matrice de covariance de $\mathbf{z}$ existe et

$$cov(\mathbf{z}) = \sigma_g \mathbf{\Sigma} \tag{1.19}$$

avec

$$\sigma_g = \frac{\mathbb{E}(\mathcal{R}^2)}{R} \tag{1.20}$$

Notons donc que la matrice de covariance peut ne pas exister pour certaines distributions (c'est le cas pour la loi de Cauchy par exemple). Cependant, si elle existe, elle est proportionnelle à la matrice de dispersion $\mathbf{\Sigma}$ qui, elle, est toujours définie.

1.2.3 Le cas particulier des gaussiennes composées (CG)

La famille des gaussiennes composées (notée CG pour Compound Gaussian) forme une sous-classe importante des distributions CES. En effet, beaucoup de distributions CES classiques font partie de cette famille, mais pas toutes [88, 128]. Ces distributions, aussi appelée SIRVs [127], pour Spherically Invariant Random Vectors, ont beaucoup été étudiées par la communauté de Traitement du Signal, notamment en Radar [113, 51, 28, 32, 45]. L'idée principale est d'améliorer la modélisation du milieu en considérant que celui ci est localement gaussien, mais qu'il présente une variabilité spatiale de puissance, due par exemple, aux hautes résolutions. L'intérêt de cette modélisation est renforcée par une bonne correspondance avec des distributions empiriques de données réelles, comme en témoignent les nombreuses études : [58, 121, 123, 25, 26, 97, 44, 85]. Les CG suivent le théorème de représentation suivant [127] :

Définition 1.2.3.1 *Représentation d'un vecteur aléatoire CG*
Un vecteur aléatoire CG (ou SIRV) de matrice de dispersion $\mathbf{\Sigma}$ et de moyenne $\boldsymbol{\mu}$ est représenté par :

$$\mathbf{z} \overset{d}{=} \boldsymbol{\mu} + \sqrt{\tau}\mathbf{x} \tag{1.21}$$

avec $\mathbf{x}$ un vecteur aléatoire gaussien $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma})$ et τ une variable aléatoire positive et indépendante de $\mathbf{x}$, appelée texture, de densité de probabilité f_τ. On notera :

$$\mathbf{z} \sim \mathcal{CG}(\boldsymbol{\mu}, \mathbf{\Sigma}, f_\tau) \tag{1.22}$$

Ce Théorème de représentation est intéressant puisque conditionnellement à la texture, le vecteur aléatoire est simplement gaussien :

$$(\mathbf{z}|\tau) \sim \mathcal{CN}(\boldsymbol{\mu}, \tau\mathbf{\Sigma}) \tag{1.23}$$

Au travers de cette représentation, il est aussi possible de faire directement le lien avec les distributions CES. Soit $\mathbf{\Sigma} = \mathbf{A}\mathbf{A}^H$, une factorisation de $\mathbf{\Sigma}$ telle que $\mathbf{A} \in \mathbb{C}^{M\times R}$, un vecteur aléatoire CG à la représentation stochastique suivante :

$$\mathbf{z} \overset{d}{=} \boldsymbol{\mu} + \sqrt{\tau}\mathbf{A}\mathbf{y} \tag{1.24}$$

avec $\mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_R)$. On rappelle alors l'identité $\mathcal{U}^{(R)} \overset{d}{=} \mathbf{y}/||\mathbf{y}||_2$, conduisant alors à :

$$\mathbf{z} \overset{d}{=} \boldsymbol{\mu} + \sqrt{\tau}\mathbf{x} \overset{d}{=} \boldsymbol{\mu} + \mathcal{R}\mathbf{A}\mathcal{U}^{(R)} \tag{1.25}$$

avec $\mathcal{R} \overset{d}{=} ||\mathbf{y}||_2\sqrt{\tau}$. Un vecteur CG suit donc bien le théorème de représentation des distributions CES. Les CG sont donc une sous-famille des CES. Il en existe d'autres, comme par exemple, les gaussiennes généralisées multivariées (voir par exemple [88]).

De plus, les CG "héritent" du problème d'identifiabilité soulevé pour les distributions CES. C'est pourquoi il convient de fixer au préalable une convention ($\mathbb{E}(\tau) = 1$ où $\text{Tr}(\mathbf{\Sigma}) = M$) pour assurer une définition unique.

La densité de probabilité d'un vecteur CG existe toujours et se définit comme suit :

Définition 1.2.3.2 *Densité de probabilité d'un vecteur CG*

$$f_{\mathbf{z}}(\mathbf{z}) = \pi^{-M}|\mathbf{\Sigma}|^{-1}\int_0^\infty \tau^{-M}\exp\left(\frac{-(\mathbf{z}-\boldsymbol{\mu})^H\mathbf{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu})}{\tau}\right) f_\tau(\tau)d\tau \tag{1.26}$$

on notera :

$$\mathbf{z} \sim \mathcal{CG}(\boldsymbol{\mu}, \mathbf{\Sigma}, f_\tau) \tag{1.27}$$

Ici encore, le lien avec les distributions CES peut se faire en notant $g(t) \propto \int_0^\infty \tau^{-M}\exp(-t/\tau)f_\tau(\tau)\mathrm{d}\tau$: la densité de probabilité est bien conforme à la définition 1.2.2.2.

Enfin, la matrice de covariance de $\mathbf{z}$, si elle existe, est proportionnelle à la matrice de dispersion et fait l'objet du théorème suivant :

Théorème 1.2.4 *Matrice de covariance d'un vecteur CG*
Soit un vecteur aléatoire centré $\mathbf{z} \sim \mathcal{CG}(\mathbf{0}, \mathbf{\Sigma}, f_\tau)$*. Si* $\mathbb{E}(\tau) < \infty$*, alors la matrice de covariance de* $\mathbf{z}$ *existe et*

$$cov(\mathbf{z}) = \mathbb{E}(\tau)\mathbf{\Sigma} \tag{1.28}$$

1.2.4 Quelques exemples de distributions complexes elliptiques symétriques

Cette section donne quelques exemples de distributions CES. Bien sûr, nous ne détaillerons pas toutes les lois pouvant être couvertes par cette formalisation. On notera cependant que de nombreuses distributions usuelles sont incluses dans cette famille, par exemple : la loi normale, la loi de Laplace, la loi de Student t, la loi de Weibull, la loi de Cauchy... L'expression de ces lois sous forme de distributions CES ou de CG (ainsi que d'autres exemples) peut se trouver dans la synthèse [84].

Gausienne

La loi normale $\mathbf{z} \sim \mathcal{CN}(\boldsymbol{\mu}, \mathbf{\Sigma})$ est une distribution CES : $\mathbf{z} \sim \mathcal{CE}(\boldsymbol{\mu}, \mathbf{\Sigma}, g)$ avec $g(t) = \exp(-t)$. Elle correspond aussi à la distribution CG $\mathbf{z} \sim \mathcal{CG}(\boldsymbol{\mu}, \mathbf{\Sigma}, f_\tau)$ avec $f_\tau(\tau) = \delta(\tau - 1)$.

K-distribution

La K-distribution est une loi CG $\mathbf{z} \sim \mathcal{CG}(\boldsymbol{\mu}, \mathbf{\Sigma}, f_\tau)$ avec :

$$f_\tau(\tau, \nu) = \frac{\nu^\nu}{\Gamma(\nu)}\tau^{\nu-1}e^{-\nu\tau} \tag{1.29}$$

La texture suit donc une loi Gamma $\tau \sim \mathrm{Gam}(\nu, 1/\nu)$ de paramètre de forme $\nu > 0$.

On remarque que, de par la définition donnée, $\mathbb{E}(\tau) = 1$. Il n'y a donc pas d'ambiguïté d'échelle et la matrice de covariance $\mathrm{cov}(\mathbf{z}) = \mathbf{\Sigma}$.

En termes de "comportement", le cas limite $\nu \to \infty$ correspond à la loi normale. Plus ν se rapproche de 0, plus la K-distribution considérée est à queue lourde, donc plus le bruit est impulsif.

La K-distribution est utilisée, par exemple, pour modéliser le fouillis de radar de mer [27, 122].

Distributions gaussiennes généralisées

Les Distributions gaussiennes généralisées [88, 128] notées MGGD (pour Multivariate Generalized Gaussian Distribution), suivent la densité de probabilité $\mathbf{z} \sim \mathcal{CE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ avec :

$$g(t) = \exp(-t^s/b) \tag{1.30}$$

où s est le paramètre d'exposant, et b le paramètre d'échelle. La variable $\mathbf{z}$ suit alors le théorème de représentation 1.2.2 où la variable $\mathcal{Q} = \mathcal{R}^2$ est distribuée comme $\mathcal{Q} \overset{d}{=} G^{1/s}$, avec $G \sim \text{Gam}(\frac{m}{s}, b)$.

Pour lever l'ambiguïté d'échelle sur le couple (g,$\boldsymbol{\Sigma}$), on peut forcer $b = \left[m\Gamma\left(\frac{m}{s}\right)/\Gamma\left(\frac{m+1}{s}\right)\right]$, qui impose $\text{cov}(\mathbf{z}) = \boldsymbol{\Sigma}$.

En termes de comportement, la loi normale s'obtient avec le paramètre $s = 1$. Les gaussiennes généralisées permettent de modéliser des distributions à queues plus lourdes (pour $s < 1$) ou plus légères (pour $s > 1$).

1.3 Estimation de la matrice de covariance

Nous introduisons dans cette section le problème majeur considéré dans cette thèse : l'estimation de la matrice de covariance. Ce problème est fondamental et apparaît dans de nombreuses applications. En effet, la plupart du temps, la matrice de covariance du signal est inconnue et doit donc être estimée en vue d'effectuer les traitements voulus.

Classiquement, on dispose de K échantillons $\mathbf{z}_k, k \in [\![1, K]\!]$, appelées aussi données secondaires, que l'on assume indépendantes et identiquement distribuées ($i.i.d.$) et ayant une statistique d'ordre 2, $\mathbf{\Sigma}$ (covariance ou dispersement), inconnue. Le problème est alors d'estimer la vraie matrice $\mathbf{\Sigma}$ au travers de ces observations. L'estimateur, noté $\hat{\mathbf{\Sigma}}$, devant idéalement être le plus "précis" possible (selon un critère à définir).

Avant tout, effectuons un rappel des précédentes hypothèses (évoquées au cours des définitions) :
- Le moment d'ordre un de la distribution observée sera considéré comme nul (ou connu).
- Les signaux considérés sont circulaires à l'ordre 2, c'est à dire, ayant une matrice de pseudo-covariance nulle.
- Nous nous limiterons aux distributions CES dont la densité de probabilité existe.

1.3.1 Propriétés attendues des estimateurs

Définition 1.3.1.1 *Biais*
Le biais d'un estimateur $B(\hat{\mathbf{\Sigma}})$ d'un estimateur $\hat{\mathbf{\Sigma}}$ de $\mathbf{\Sigma}$ est défini par :

$$B(\hat{\mathbf{\Sigma}}) = \mathbb{E}\left[\hat{\mathbf{\Sigma}}\right] - \mathbf{\Sigma} \tag{1.31}$$

De plus, si $B(\hat{\mathbf{\Sigma}}) = \mathbf{0}$ l'estimateur est dit "non biaisé" ou "sans biais".

Définition 1.3.1.2 *Consistance*
Un estimateur $\hat{\mathbf{\Sigma}}$ de $\mathbf{\Sigma}$ est dit consistant s'il converge en probabilité vers $\mathbf{\Sigma}$ quand le nombre d'échantillon K tend vers l'infini :

$$\forall \epsilon > 0 \ , \ \mathbb{P}\left(||\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}|| \geq \epsilon\right) \xrightarrow[K \longrightarrow \infty]{\mathbb{P}} 0 \, , \tag{1.32}$$

où $||.||$ est une norme matricielle quelconque.

Définition 1.3.1.3 *Gaussianité asymptotique*
Un estimateur $\hat{\mathbf{\Sigma}}$ de $\mathbf{\Sigma}$ est asymptotiquement gaussien si

$$\sqrt{K}\left(vec\left(\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\right)\right) \xrightarrow[K \longrightarrow \infty]{d} \mathcal{CN}\left(\mathbf{0}, \mathbf{C}\right) \tag{1.33}$$

où $\mathbf{C} \in \mathbb{S}^+_{m^2}$.

1.3.2 La Sample Covariance Matrix

Sous l'hypothèse de bruit gaussien $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma})$, la vraisemblance des données secondaires $\{\mathbf{z}_k\}_{k\in[\![1,K]\!]}$ est :

$$f(\{\mathbf{z}_k\}|\mathbf{\Sigma}) \propto |\Sigma|^{-K} \prod_{k=1}^{K} \exp\left(-\mathbf{z}_k^H \mathbf{\Sigma}^{-1} \mathbf{z}_k\right) \tag{1.34}$$

donnant la log-vraisemblance suivante :

$$\ln f(\{\mathbf{z}_k\}|\mathbf{\Sigma}) = -K \ln |\mathbf{\Sigma}| - \sum_{k=1}^{K} \mathbf{z}_k^H \mathbf{\Sigma}^{-1} \mathbf{z}_k + \gamma = -K \ln |\mathbf{\Sigma}| - \mathrm{Tr}\left(\mathbf{\Sigma}^{-1} \mathbf{S}_K\right) + \gamma \tag{1.35}$$

où $\mathbf{S}_K = \sum_{k=1}^{K} \mathbf{z}_k \mathbf{z}_k^H$, et γ est une constante indépendante de $\{\mathbf{z}_k\}$ et $\mathbf{\Sigma}$. Les formules de dérivées matricielles suivantes : $\partial \ln |\mathbf{\Sigma}| / \partial \mathbf{\Sigma} = \mathbf{\Sigma}^{-1}$ et $\partial \mathbf{z}_k^H \mathbf{\Sigma}^{-1} \mathbf{z}_k / \partial \mathbf{\Sigma} = -\mathbf{z}_k \mathbf{z}_k^H \mathbf{\Sigma}^{-2}$ ([45] et [63] p.520), permettent alors d'aboutir au maximum de vraisemblance de la matrice de covariance dans le cas gaussien :

$$\frac{\partial \ln f(\{\mathbf{z}_k\}|\mathbf{\Sigma})}{\partial \mathbf{\Sigma}} = \mathbf{0} \Leftrightarrow \mathbf{0} = -K\mathbf{\Sigma}^{-1} + \mathbf{S}_k \mathbf{\Sigma}^{-2} \Leftrightarrow \mathbf{\Sigma} = \hat{\mathbf{\Sigma}}_{SCM} \tag{1.36}$$

avec $\hat{\mathbf{\Sigma}}_{SCM}$, la Sample Covariance Matrix (SCM) :

$$\hat{\mathbf{\Sigma}}_{SCM} = \frac{1}{K} \sum_{k=1}^{K} \mathbf{z}_k \mathbf{z}_k^H \tag{1.37}$$

Cet estimateur est facile à mettre en œuvre et a longtemps été utilisé dans la plupart des applications. En effet, dans le cas d'échantillons gaussiens, les performances de la SCM sont parfaitement connues, celle-ci est non-biaisée, consistante et suit une loi de Wishart [63] (gaussienne à distance finie). Cet estimateur est de plus efficace en termes d'erreur quadratique moyenne : il atteint la borne de Cramér-Rao.

Si la SCM est utilisée avec des échantillons issus d'une distribution CES (non nécessairement gaussiens), sous condition que la matrice de covariance existe, la SCM conserve toujours ses propriétés de non biais, consistance ainsi que la gaussianité asymptotique, ceci grâce au théorème de la limite centrale.

Cependant les performances de la SCM peuvent être fortement dégradées à distance finie quand la distribution des échantillons est à queue lourde ou en présence des données aberrantes, même à faible pourcentage (comme illustré dans [71]). C'est pourquoi l'on se tourne désormais vers des estimateurs plus robustes à ces conditions.

1.3.3 La Normalized Sample Covariance Matrix

La Normalized Sample Covariance (NSCM) est un estimateur de la matrice de covariance introduit initialement dans [29, 30] pour le contexte des modèles CG.

La NSCM est définie comme la SCM des données normalisées :

$$\hat{\mathbf{\Sigma}}_{NSCM} = \frac{M}{K} \sum_{k=1}^{K} \frac{\mathbf{z}_k \mathbf{z}_k^H}{\mathbf{z}_k^H \mathbf{z}_k} \tag{1.38}$$

Tout d'abord, on remarque que cet estimateur est bien adapté au contexte des données suivant une distribution CES $\mathbf{z} \sim \mathcal{CG}(\mathbf{0}, \mathbf{\Sigma}, g)$. En effet, la NSCM fait naturellement disparaitre le module aléatoire $\mathcal{R}$ des variables $\mathbf{z} \overset{d}{=} \mathcal{R}\mathbf{A}\mathcal{U}$. En effet :

$$\hat{\mathbf{\Sigma}}_{NSCM} = \frac{M}{K}\sum_{k=1}^{K}\frac{\mathbf{z}_k\mathbf{z}_k^H}{\mathbf{z}_k^H\mathbf{z}_k} = \frac{M}{K}\sum_{k=1}^{K}\frac{\mathbf{A}\mathcal{U}_k\mathcal{U}_k^H\mathbf{A}^H}{\mathbf{A}\mathcal{U}_k^H\mathcal{U}_k\mathbf{A}^H} \tag{1.39}$$

La NSCM est donc indépendante de la loi de $\mathcal{R}$ (ou du générateur g) et se comportera identiquement au cas gaussien quelle que soit la distribution initiale des échantillons.

De part sa robustesse aux différentes distributions la NSCM semble intéressante. De plus, les performances asymptotiques de la NSCM sont connues : ses moment d'ordre un et deux ont été calculés dans [11]. Néanmoins, cet estimateur n'est pas sans défaut puisque s'avère être biaisé, ce qui peut conduire à de mauvaises performances des traitements s'appuyant dessus [46, 48, 50]

1.3.4 Le maximum de vraisemblance des distributions complexes elliptiques symétriques

Considérons un vecteur aléatoire CES $\mathbf{z} \sim \mathcal{CG}(\mathbf{0}, \mathbf{\Sigma}, g)$, la vraisemblance d'un jeu de données secondaires $\{\mathbf{z}_k\}$ est donnée par :

$$f(\{\mathbf{z}_k\}|\mathbf{\Sigma}) \propto |\Sigma|^{-K}\prod_{k=1}^{K} g\left(\mathbf{z}_k^H\mathbf{\Sigma}^{-1}\mathbf{z}_k\right) , \tag{1.40}$$

donnant la log-vraisemblance :

$$\ln f(\{\mathbf{z}_k\}|\mathbf{\Sigma}) = -K\ln|\mathbf{\Sigma}| - \sum_{k=1}^{K}\ln g\left(\mathbf{z}_k^H\mathbf{\Sigma}^{-1}\mathbf{z}_k\right) + \gamma , \tag{1.41}$$

où γ est une constante indépendante de $\{\mathbf{z}_k\}$ et $\mathbf{\Sigma}$. Les formules de dérivées matricielles rappelées section 1.3.2 (ainsi qu'une simple composition) permettent d'obtenir l'expression du maximum de vraisemblance de la matrice de covariance :

$$\frac{\partial \ln f(\{\mathbf{z}_k\}|\mathbf{\Sigma})}{\partial \mathbf{\Sigma}} = \mathbf{0} \Leftrightarrow \mathbf{0} = -K\mathbf{\Sigma}^{-1} + \frac{g'\left(\mathbf{z}_k^H\hat{\mathbf{\Sigma}}_{MVCE}^{-1}\mathbf{z}_k\right)}{g\left(\mathbf{z}_k^H\hat{\mathbf{\Sigma}}_{MVCE}^{-1}\mathbf{z}_k\right)}\mathbf{z}_k\mathbf{z}_k^H\mathbf{\Sigma}^{-2} \Leftrightarrow \mathbf{\Sigma} = \hat{\mathbf{\Sigma}}_{MVCE} \tag{1.42}$$

où $g'(t) = \partial g(t)/\partial t$, et avec :

$$\hat{\mathbf{\Sigma}}_{MVCE} = \frac{1}{K}\sum_{k=1}^{K}\psi\left(\mathbf{z}_k^H\hat{\mathbf{\Sigma}}_{MVCE}^{-1}\mathbf{z}_k\right)\mathbf{z}_k\mathbf{z}_k^H \tag{1.43}$$

en notant $\psi(t) = g'(t)/g(t)$.

Dans le cas gaussien, $\psi(t) = 1$, on retrouve donc bien la SCM comme unique maximum de vraisemblance de la matrice de covariance. Hormis ce cas particulier, la fonction ψ n'est généralement pas constante et le maximum de vraisemblance se présente sous forme implicite, puisqu'il dépend lui même de son inverse. Obtenir cette solution nécessite donc un algorithme itératif. L'existence et l'unicité de

cette solution, ainsi que la convergence de l'algorithme itératif associé, fait l'objet de conditions bien spécifiques, dont la formulation la plus générale est présentée dans [84] et ré-exprimée ci-dessous.

Tout d'abord, introduisons les notations suivantes. Aux vecteurs complexes $\mathbf{z}_k = \mathbf{x}_k + j\mathbf{y}_k$ (avec $j = \sqrt{-1}$) on associe les paires :

$$\mathbf{v}_k = \left(\mathbf{x}_k^T, \mathbf{y}_k^T\right)^T \ , \mathbf{v}_{M+k} = \left(-\mathbf{y}_k^T, \mathbf{x}_k^T\right)^T$$

pour $k \in [\![1, K]\!]$. On note de plus :

$$\phi(t) = t\psi(t) \qquad \text{et} \qquad \mathcal{K} = \sup\{\phi(t) \ , \ t \geq 0\} \tag{1.44}$$

Soient les conditions suivantes :

- **M1** : ψ est non négative, continue et non croissante.
- **M2** : ϕ est strictement croissante, où ϕ est non décroissante et strictement croissante pour $\psi(t) < M + \epsilon$ pour $\epsilon > 0$.
- **M3** : Soit $\mathbb{P}_{2M}$ la distribution empirique des échantillons $\{\mathbf{v}_1, \ldots, \mathbf{v}_{2M}\}$. Alors pour tous les sous-espaces linéaires $V \in \mathbb{R}^{2M}$, avec $\dim(V) \leq 2M-1$, on vérifie $\mathbb{P}_{2M}(V) < 1-(2M - \dim(V))/2\mathcal{K}$

On remarque que les conditions M1 et M2 imposent une forme particulière sur la fonctionnelle ψ. La condition M3 impose aussi $\mathcal{K} > M$ car $0 \leq \mathbb{P}_{2M}(\{\mathbf{0}\}) < 1 - m/K$. Cependant M3 est aussi une condition sur la "diversité spatiale" de l'ensemble des données secondaires. Si $\mathcal{K} = \infty$, M3 implique que ces données ne sont pas contenues dans un sous-espace de dimension inférieure à M (dans le cas où la matrice de covariance est de rang plein, la condition impose directement que le nombre de données $K > M$).

Ces conditions énoncées, on a le Théorème suivant :

Théorème 1.3.1 *Existence, unicité et calcul de $\hat{\mathbf{\Sigma}}_{MVCE}$*
Si les conditions M1, M2 et M3 sont satisfaites, alors $\hat{\mathbf{\Sigma}}_{MVCE}$ existe et correspond à l'unique solution de (1.43). De plus les itérations

$$\mathbf{\Sigma}_{(n+1)} = \frac{1}{K}\sum_{k=1}^{K} \psi(\mathbf{z}_k^H \mathbf{\Sigma}_{(n)}^{-1} \mathbf{z}_k)\mathbf{z}_k\mathbf{z}_k^H \tag{1.45}$$

convergent vers $\hat{\mathbf{\Sigma}}_{MVCE}$ quelle que soit l'initialisation $\mathbf{\Sigma}_{(0)} \in \mathbb{S}_+^M$ inversible.

Si la fonctionnelle g correspond à l'exacte distribution des données, toutes les propriétés habituelles du maximum de vraisemblance sont conservées (biais nul, consistance, efficacité et gaussianité asymptotique).

Cependant, la distribution exacte des données (donc g) est bien souvent inconnue. C'est pourquoi il est utile de replacer le maximum de vraisemblance des distributions CES dans le cadre plus large des M-estimateurs.

1.3.5 Les M-estimateurs

Les M-estimateurs ont été proposés comme estimateurs robustes de la matrice de covariance dans [74]. Beaucoup étudiés dans la littérature statistique pour le cas réel [55, 65, 115, 116, 117, 118, 75], leur extension au cas complexe a été notamment considéré dans [80, 89, 79, 84]. Ces estimateurs ont été un sujet de recherche très actif au cours des dernières années pour des applications de Traitement du Signal comme l'estimation de direction d'arrivée, la détection radar [32, 45, 71] ou hyperspectrale [40, 41, 39], l'analyse en composantes indépendantes [79], ainsi que de nombreuses autres applications telles que l'optimisation de protfolio [38].

Soit $\{\mathbf{z}_k\}$, K réalisations $i.i.d.$ d'un vecteur aléatoire CES $\mathbf{z} \sim \mathcal{CE}(\mathbf{0}, \mathbf{\Sigma}, g)$. On considère le M-estimateur $\hat{\mathbf{\Sigma}}_M$, solution de l'équation :

$$\hat{\mathbf{\Sigma}}_M = \frac{1}{K}\sum_{k=1}^{K} \varphi(\mathbf{z}_k^H \hat{\mathbf{\Sigma}}_M^{-1} \mathbf{z}_k)\mathbf{z}_k\mathbf{z}_k^H \tag{1.46}$$

où φ est une fonction à valeur réelle donnée.

Comme vu section précédente, l'EMV correspond au cas particulier $\varphi(t) = g'(t)/g(t)$. Néanmoins, il est important de noter que la définition de $\hat{\mathbf{\Sigma}}_M$ ne force pas de lien entre la fonction φ et le générateur de densité g. L'estimateur existe en effet tant que φ et les données vérifient les conditions du théorème 1.3.1. Les M-estimateurs se voient donc comme une généralisation du maximum de vraisemblance. Plutôt que de résoudre un cas particulier, l'idée est d'obtenir, au travers d'un choix judicieux de la fonction φ, un estimateur "robuste" sur l'ensemble des distributions CES.

Considérons l'équation suivante :

$$\mathbf{\Sigma}_M = \mathbb{E}\left[\varphi(\mathbf{z}^H \mathbf{\Sigma}_M^{-1} \mathbf{z})\mathbf{z}\mathbf{z}^H\right] \tag{1.47}$$

Si les conditions du théorème 1.3.1 sont respectées, alors [74] montre que :

- $\mathbf{\Sigma}_M$ existe de manière unique
- et vérifie :

$$\mathbf{\Sigma}_M \propto \mathbf{\Sigma} \tag{1.48}$$

  Le facteur de proportionnalité entre ces deux matrices σ_φ dépend de φ et du générateur de densité g et peut s'obtenir numériquement [84].
- $\hat{\mathbf{\Sigma}}_M$ est un estimateur consistant de $\mathbf{\Sigma}$ à un facteur d'échelle près. On peut l'exprimer à l'aide la normalisation suivante :

$$\frac{\hat{\mathbf{\Sigma}}_M}{\mathrm{Tr}(\hat{\mathbf{\Sigma}}_M)} \xrightarrow[K\longrightarrow\infty]{p} \frac{\mathbf{\Sigma}}{\mathrm{Tr}(\mathbf{\Sigma})} \tag{1.49}$$

- les M-estimateurs sont asymptotiquement gaussiens. Les paramètres de leur distribution asymptotiques sont étudiés dans [72] et [84].

On remarque donc que les M-estimateurs permettent d'estimer principalement la matrice de dispersion ($i.e.$ la matrice covariance à un facteur d'échelle près). Ceci n'est pas nécessairement un problème pour certaines applications dont le résultat n'est pas dépendant de l'échelle. Par exemple : le détecteur ANMF [67, 68] ou l'analyse en composante principale.

Nous détaillons ci-dessous deux exemples de M-estimateurs parmi les plus usuels.

M-estimateur de Huber

L'estimateur de Huber [56], noté $\hat{\mathbf{\Sigma}}_H$, est défini par Eq.(1.46) pour la fonction de pénalisation suivante :

$$\varphi(t) = \begin{cases} 1/b \text{ si } t \leq c^2 \\ c^2/(tb) \text{ si } t > c^2 \end{cases} \tag{1.50}$$

où c est un paramètre servant à régler la robustesse de l'estimateur. Le cas limite $c \to \infty$ fait tendre l'estimateur de Huber vers la SCM, tandis que le cas $c = 0$ conduit à l'estimateur de Tyler définit ci-après. Entre les deux, l'estimateur traite les données de manière "hybride" : il pénalise les données ayant une norme $\mathbf{z}_k^H \hat{\mathbf{\Sigma}}_H^{-1} \mathbf{z}_k$ supérieure à c^2. le facteur de normalisation b est fixé de telle sorte que l'estimateur résultant soit consistant avec l'echelle de la la vraie matrice de covariance, ou qu'il satisfasse une normalisation voulue.

Estimateur de Tyler

L'estimateur de Tyler [114], aussi appelé estimateur du point fixe [89], noté $\hat{\mathbf{\Sigma}}_{FPE}$, est défini par Eq.(1.46) pour la fonction de pénalisation suivante :

$$\varphi(t) = M/t \tag{1.51}$$

Conduisant à la définition :

$$\hat{\mathbf{\Sigma}}_{FPE} = \frac{M}{K} \sum_{k=1}^{K} \frac{\mathbf{z}_k \mathbf{z}_k^H}{\mathbf{z}_k^H \hat{\mathbf{\Sigma}}_{FPE}^{-1} \mathbf{z}_k} \tag{1.52}$$

On peut donc le voir comme un cas limite de l'estimateur de Huber dans le cas $c = 0$. Néanmoins, la fonction φ ne respecte pas la condition M2 du théorème 1.3.1. Il a été montré dans [114, 89, 84] que l'estimateur existe et est unique à un facteur d'échelle près, pour $K > M$ si les échantillons $\mathbf{z}_k$ sont linéairement indépendants et que $\mathbf{z}_k \neq \mathbf{0}\ \forall k$. De plus, les itérations :

$$\mathbf{\Sigma}_{(n+1)} = \frac{M}{K} \sum_{k=1}^{K} \frac{\mathbf{z}_k \mathbf{z}_k^H}{\mathbf{z}_k^H \mathbf{\Sigma}_{(n)}^{-1} \mathbf{z}_k} \tag{1.53}$$

convergent vers $\hat{\mathbf{\Sigma}}_{FPE}$ quelle que soit l'initialisation.

Cet estimateur correspond au maximum de vraisemblance dans le cas de données CG où les textures sont considérées déterministes inconnues [32, 45]. Ce résultat a été étendu à toutes les distributions CES [82].

Cet estimateur a soulevé beaucoup d'intérêt et a fait l'objet de nombreuses études dans le cadres des distributions CES (comme en témoigne l'ensemble des références de cette section). En effet, comme la NSCM, cet estimateur a un comportement indépendant du générateur de densité g puisque son expression fait naturellement disparaître les facteurs de puissances aléatoires. Il aura donc les mêmes performances quelle que soit la distribution CES sous-jacente aux données. Cependant, il ne souffre pas des défauts majeurs de la NSCM : le Point Fixe est non biaisé et consistant pour l'ensemble des distributions CES. Comme les M-estimateurs, sa distribution est asymptotiquement gaussienne de paramètres connus [91, 84, 72].

1.3.6 Les Estimateurs Robustes Régularisés

Les M-estimateurs sont connus pour leur robustesse à différentes distributions ainsi qu'à leur résistance aux outliers. Néanmoins, ceux ci nécessitent un nombre d'échantillons $K > M$ afin d'être définis. En effet, les M-estimateurs sont définis au travers de leur inverse (inexistante si $K < M$) dans l'équation de point fixe (1.43). Selon le système considéré, satisfaire cette condition n'est pas toujours possible en pratique. C'est pourquoi de récents travaux (principalement axés sur l'estimateur du point fixe) ont proposé de régulariser ces estimateurs afin qu'ils soient adaptés aux problèmes de "grandes dimensions" ($K < M$).

Ces travaux peuvent se voir comme une extension de [70], qui propose de régulariser la SCM de la manière suivante :

$$\hat{\mathbf{\Sigma}}_{SSCM} = \frac{(1-\beta)}{K}\sum_{k=1}^{K}\mathbf{z}_k\mathbf{z}_k^H + \beta\mathbf{I}_M \tag{1.54}$$

afin d'obtenir un estimateur inversible et bien conditionné, même dans les cas où $K < M$.

Initialement [3] propose de régulariser l'algorithme du Point Fixe en ajoutant un diagonal-loading et en forçant une normalisation à chaque étape :

$$\begin{cases} \mathbf{\Sigma}_{(n+1)} = (1-\beta)\dfrac{M}{K}\displaystyle\sum_{k=1}^{K}\frac{\mathbf{z}_k\mathbf{z}_k^H}{\mathbf{z}_k^H\mathbf{\Sigma}_{(n)}^{-1}\mathbf{z}_k} + \beta\mathbf{I}_M \\ \mathbf{\Sigma}_{(n+1)} = \mathbf{\Sigma}_{(n+1)}/\mathrm{Tr}(\mathbf{\Sigma}_{(n+1)} \end{cases} \tag{1.55}$$

Il a été prouvé que cet algorithme converge vers une unique solution dans [22]. Cependant, cet algorithme n'est pas associé à la maximisation d'une fonctionnelle, ce qui rend son interprétation difficile.

Par la suite [90, 111, 83] ont considéré la log-vraisemblance pénalisée suivante :

$$\mathcal{L}_\beta\left(\{\mathbf{z}_k\}|\mathbf{\Sigma}\right) = -M(1-\beta)\sum_{k=1}^{K}\mathbf{z}_k^H\mathbf{\Sigma}^{-1}\mathbf{z}_k - K\ln|\mathbf{\Sigma}| - K\beta\mathrm{Tr}(\mathbf{\Sigma}^{-1}) \tag{1.56}$$

Soit la vraisemblance conduisant à l'estimateur du Point Fixe plus une fonctionnelle $\mathcal{P}(\mathbf{\Sigma}) = \beta\mathrm{Tr}(\mathbf{\Sigma}^{-1})$ pénalisant les matrices mal conditionnées. Il a été établit le théorème suivant [90, 111]

Théorème 1.3.2 *Existence, unicité et calcul du Point Fixe régularisé*
La log-vraisemblance pénalisée $\mathcal{L}_\beta$ admet pour unique maximum la solution $\hat{\mathbf{\Sigma}}_{SFPE}(\beta)$, définie comme :

$$\hat{\mathbf{\Sigma}}_{SFPE}(\beta) = (1-\beta)\frac{M}{K}\sum_{k=1}^{K}\frac{\mathbf{z}_k\mathbf{z}_k^H}{\mathbf{z}_k^H\hat{\mathbf{\Sigma}}_{SFPE}^{-1}(\beta)\mathbf{z}_k} + \beta\mathbf{I}_M \tag{1.57}$$

pour tout $\beta \in [max(0, 1-K/M)\ ,\ 1]$. Cette solution satisfait naturellement la contrainte :

$$Tr(\mathbf{\Sigma}^{-1}) = M \tag{1.58}$$

De plus, les itérations

$$\mathbf{\Sigma}_{(n+1)} = (1-\beta)\frac{M}{K}\sum_{k=1}^{K}\frac{\mathbf{z}_k\mathbf{z}_k^H}{\mathbf{z}_k^H\mathbf{\Sigma}_{(n)}^{-1}\mathbf{z}_k} + \beta\mathbf{I}_M \tag{1.59}$$

convergent vers $\hat{\mathbf{\Sigma}}_{SFPE}(\beta)$ pour toute initialisation.

Ce résultat a été généralisé à l'ensemble des M-estimateurs dans [83]

L'apport de la régularisation est d'offrir une possibilité de calculer un estimateur robuste, même si $K < M$. De plus, la régularisation permet de contraindre un bon conditionnement de l'estimateur. Cependant, cette opération conduit induit un biais non nul [60]. Le réglage du paramètre de régularisation β est donc une affaire de compromis, et est surtout dépendant de l'application considérée. Plusieurs techniques de choix de β adaptatif optimal (selon le contexte) ont été envisagées dans la littérature. Citons notamment les méthodes Oracle proposées dans [22, 83] visant à minimiser l'erreur moyenne sur la covariance. Les références [2, 13] proposent un alternative au travers de l'étude de l'"Expected Likelihood" des distributions CES. D'autres solutions ont été envisagées dans le cadre de la théorie des grandes matrices aléatoires, par exemple pour minimiser l'erreur quadratique moyenne [33] ou pour obtenir un estimateur optimal en termes de performances de détection [61].

1.4 Introduction de la problématique considérée

1.4.1 Estimation de matrice structurées

Le problème du nombre limité d'échantillons n'est pas seulement lié aux questions d'existence d'estimateurs. Il se pose en effet de manière récurrente en Traitement du Signal. Prenons un exemple concret : dans le cas gaussien, le filtre adaptatif construit avec la SCM requiert $K \simeq 2M$ pour atteindre des performances satisfaisantes [1]. Si l'on considère un système radar utilisant 4 antennes et 64 impulsions, la taille des données est $M = 256$. Cela signifie que 512 échantillons sont théoriquement nécessaires pour effectuer un filtrage convenable. Bien évidemment, ce nombre conséquent d'échantillons n'est pas toujours disponible en pratique.

Afin de palier au problème, il est possible d'utiliser un a priori sur la matrice de covariance. La plupart des applications sont en effet associées à des structures particulières de matrice de covariance. Exploiter la connaissance de cette structure permet alors de réduire les degrés de liberté du problème d'estimation, ce qui se traduit par un traitement nécessitant moins de données secondaires.

Englobant les structures matricielles les plus utilisées, on peut citer les catégories suivantes :

- Les matrices appartenant à un groupe linéaire : $\mathcal{S} = \{\mathbf{\Sigma} = \sum_{i=0}^{I} a_i \mathbf{B}_i \ , \ a_i \in \mathbb{R}\}$ avec une base de matrices $\{\mathbf{B}_i\}$ connue. Ce modèle décrit notamment les structures Toeplitz, circulantes, matrices bandes et les sommes de matrices de rang 1. La connaissance de la base $\{\mathbf{B}_i\}$ provient d'hypothèses sur le bruit ou sur la configuration du système ($e.g.$ la géométrie de l'antenne [37]). Le problème d'estimation de la matrice de covariance pour ces structures revient alors à estimer les coefficients a_i.
- Les matrices appartenant à un groupe symétrique [104] : $\mathcal{F}_{\mathcal{H}} = \{\mathbf{\Sigma} \mid \mathbf{H}_h \mathbf{\Sigma} \mathbf{H}_h^H = \mathbf{\Sigma}, \ \forall \, \mathbf{H}_h \in \mathcal{H}\}$ avec $\mathcal{H} = \{\mathbf{H}_h\}$ un groupe multiplicatif de matrices orthogonales. Ce modèle décrit notamment les structures persymétriques [31, 87] et les matrices circulantes.
- Les matrices structurées au travers de leur spectre, c'est à dire dont les valeurs propres reflètent une structure particulière. Cette catégorie recouvre notamment les matrices de type : hermitienne plus identité (valeurs propres minorées), rang faible plus identité (valeurs propres égales à un après un certain indice) ainsi que les matrices ayant un conditionnement contraint.

D'un point de vue technique, différentes méthodes peuvent êtres envisagées pour obtenir un estimateur satisfaisant une structure donnée, ou s'en approchant. On peut citer notamment les approches suivantes [2] :

- Projeter un estimateur sur l'ensemble considéré [104, 108], le procédé est aussi appelé rectification [37].
- Maximiser la vraisemblance directement selon les paramètres associés à la structure [86, 19, 92].
- Maximiser directement la vraisemblance sous contraintes de structure sur la matrice de covariance

1. Nous reviendrons plus en détail sur ce propos par la suite.

2. Les méthodologies évoquées étant très générales, les références associées relèvent de l'illustration et ne prétendent aucunement être exhaustives.

[9, 62].
- Ajouter une pénalisation à la vraisemblance.
- Optimiser une fonctionnelle modifiée assurant à la fois la structure et bonnes propriétés statistiques de l'estimateur [119, 57].

Initialement développées pour le contexte gaussien, ces méthodes s'étendent peu à peu au cadre plus complexe des estimateurs robustes, comme en témoignent les travaux [87, 124, 125, 2, 108, 107, 110]. Néanmoins, ces travaux de généralisation ne sont pas terminés : la section suivante présentera le cas particulier des matrices à structure rang faible. Considéré plus récemment dans [62] pour le cas gaussien, l'estimation de la matrice de covariance en contexte de bruit hétérogène à structure rang faible n'a, à notre connaissance, pas été considérée en dehors des travaux présentés dans cette thèse.

1.4.2 Matrices structurées rang faible

Définition 1.4.2.1 *Décomposition en valeurs singulières*
Soit $\mathbf{A}$ *une matrice hermitienne semi-définie positive, le théorème spectral énonce qu'elle admet l'unique décomposition suivante :*

$$\mathbf{A} = \sum_{m=1}^{M} c_m \mathbf{v}_m \mathbf{v}_m^H = \mathbf{V}_m \mathbf{C}_M \mathbf{V}_m^H \tag{1.60}$$

où $\{\mathbf{v}_m\}$ *est la base des vecteurs propres de* $\mathbf{A}$*, concaténés en* $\mathbf{V}_M$ *et vérifiant :*

$$\mathbf{V}_m^H \mathbf{V}_m = \mathbf{I}_M \tag{1.61}$$

et où $c_m \in \mathbb{R}^+ \forall m \in [\![1, M]\!]$ *sont les valeurs propres correspondantes, avec* $\mathbf{C}_M = diag(\{c_m\})$*. On adopte la convention* $c_1 \geq c_2 \geq \ldots \geq c_M \geq 0$*.*

Définition 1.4.2.2 *Rang d'une matrice*
Le rang R *d'une matrice est l'indice de sa dernière valeur propre non nulle. Une matrice* $\mathbf{A}$ *est dite de rang faible si* $R < M$ *et peut donc se décomposer*

$$\mathbf{A} = \sum_{r=1}^{R} c_r \mathbf{v}_R \mathbf{v}_R^H \tag{1.62}$$

Inversement, si $R = M$*,* $\mathbf{A}$ *est dite de rang plein, ou inversible.*

Dans de nombreuses applications, le signal d'intérêt (parfois une perturbation) réside dans un sous-espace de dimension inférieure à la taille des données : sa matrice de covariance est donc de rang faible. Cette hypothèse se vérifie, par exemple, dans les problématiques d'estimation de direction d'arrivée [101]. C'est également le cas dans certaines applications où sont présents des brouilleurs ou interférences [120]. On notera aussi que, pour les problèmes de démélange d'imagerie hyperspectrale, le signal est modélisé comme la somme d'un nombre fini de contributions "sources" (les endmembers). Pour l'ensemble de ces applications peut donc raisonnablement considérer que le signal est composé d'une source (non nécessairement gaussienne) ayant une matrice de covariance rang faible $\mathbf{\Sigma}_c$:

$$\mathbf{\Sigma}_c = \sum_{r=1}^{R} c_r \mathbf{v}_r \mathbf{v}_r^H \tag{1.63}$$

ainsi que du classique bruit blanc gaussien additif. La matrice de covariance totale du signal est donc de la forme rang faible plus identité :

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_c + \sigma^2 \mathbf{I}_M \tag{1.64}$$

Ce type de structure, très présent dans la littérature du Traitement du Signal et est aussi appelé modèle "spiked" [59].

Comme évoqué précédemment, la connaissance de cette structure a priori peut être exploitée pour limiter le nombre de données nécessaires à l'estimation de la matrice de covariance. Certaines études montrent aussi qu'ignorer cette structure et simplement effectuer des traitements classiques peut nuire aux performances de ceux-ci [76]. En définitive, une structure qui correspond à une réalité pour de nombreux systèmes et s'avère exploitable pour réduire le nombre de données nécessaires, mérite que l'on s'y intéresse.

1.4.3 Estimation de covariance à structure rang faible : le cas gaussien

Maximiser la vraisemblance sous contrainte de rang faible dans le cas gaussien revient à résoudre le problème :

$$\begin{aligned} \max_{\boldsymbol{\Sigma}} \quad & -\frac{1}{K}\sum_{k=1}^{K} \mathbf{z}_k^H \boldsymbol{\Sigma}^{-1} \mathbf{z}_k - \ln|\boldsymbol{\Sigma}| \\ \text{s.c.} \quad & \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_c + \sigma^2 \mathbf{I} \\ & \boldsymbol{\Sigma}_c \succcurlyeq \mathbf{0} \\ & \text{rank}(\boldsymbol{\Sigma}_c) = R \end{aligned} \tag{1.65}$$

Exprimée de cette façon, la contrainte de rang, non convexe, ne fait pas apparaître un problème directement solvable. Néanmoins le maximum global existe et peut être obtenu simplement [62]. Il convient de paramétrer la matrice $\boldsymbol{\Sigma}$ par sa SVD $\boldsymbol{\Sigma} = \mathbf{V}\mathbf{D}\mathbf{V}^H$. Notons la SCM $\hat{\boldsymbol{\Sigma}}_{SCM} = \sum_{k=1}^{K} \mathbf{z}_k \mathbf{z}_k^H / K$ et sa SVD $\hat{\boldsymbol{\Sigma}}_{SCM} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H$. La vraisemblance s'écrit alors

$$f(\boldsymbol{\Sigma}) = -\text{Tr}\left(\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H \mathbf{V}\mathbf{D}^{-1}\mathbf{V}^H\right) - \ln|\mathbf{D}| \tag{1.66}$$

Comme $\text{Tr}\left(\boldsymbol{\Lambda}\mathbf{U}^H\mathbf{V}\mathbf{D}^{-1}\mathbf{V}^H\mathbf{U}\right) \geq \text{Tr}\left(\boldsymbol{\Lambda}\mathbf{D}^{-1}\right)$ avec égalité vérifiée pour $\mathbf{V} = \mathbf{U}$, les vecteurs propres qui maximisent la vraisemblance, indépendamment de $\mathbf{D}$, sont ceux de la SCM. Le problème se simplifie alors puisqu'à vecteurs propres fixés, seules les valeurs propres sont désormais à optimiser. Soit $\boldsymbol{\Lambda} = \text{diag}(\{\lambda_m\})$, le seuillage suivant :

$$\lambda_m^{RCML} = \begin{cases} \max(\lambda_m, \sigma^2) & \text{pour } m \in [\![1, R]\!] \\ \sigma^2 & \text{pour } m \in [\![R+1, M]\!] \end{cases}$$

définit l'unique solution du maximum de vraisemblance sous contrainte de structure rang faible (appelé RCML) :

$$\hat{\boldsymbol{\Sigma}}_{RCML} = \mathbf{U}\ \text{diag}(\{\lambda_m^{RCML}\})\ \mathbf{U}^H \tag{1.67}$$

qui correspond donc à la SCM ayant ses valeurs propres seuillées.

1.4.4 Estimation robuste de la matrice de covariance sous contrainte de structure rang faible : un problème ouvert

La formulation robuste, similaire aux M-estimateurs (cf. section 1.3.5), du problème précédent s'exprime :

$$\begin{aligned} \max_{\boldsymbol{\Sigma}} \quad & \sum_{k=1}^{K} \ln g\left(\mathbf{z}_k^H \boldsymbol{\Sigma}^{-1} \mathbf{z}_k\right) - \ln |\boldsymbol{\Sigma}| \\ \text{s.c.} \quad & \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_c + \sigma^2 \mathbf{I} \\ & \boldsymbol{\Sigma}_c \succcurlyeq \mathbf{0} \\ & \text{rank}(\boldsymbol{\Sigma}_c) = R \end{aligned} \tag{1.68}$$

Comme on l'a précédemment vu section 1.3.5, le maximum de f s'exprime en fonction de son inverse : il n'est plus possible de faire disparaître les vecteurs propres à travers la SVD comme précédemment. Le problème est d'autant plus complexe que la fonctionnelle considérée est non convexe, tout comme la contrainte de rang. Les conditions d'existence et d'éventuelle unicité et de la solution restent donc des questions ouvertes. Notons que de récents travaux [110] proposent un algorithme permettant d'approcher une solution au moyen d'un algorithme de type Majorization-Minimization [69].

1.4.5 Estimation de la matrice de covariance en contexte hétérogène rang faible : problématique considérée dans cette thèse

Il est nécessaire d'insister sur le fait que l'approche précédemment évoquée ne reflète pas exactement le modèle réaliste que nous voulons considérer dans ces travaux. En effet, la vraisemblance contrainte de (1.68) correspond à un bruit CES ayant une matrice de covariance ayant une structure rang faible plus identité. Résoudre ce problème peut potentiellement conduire à un estimateur robuste aux diverses distributions.

Cependant un modèle plus réaliste correspond à des sources hétérogènes (CES ou CG) ayant une covariance $\boldsymbol{\Sigma}_c$ de rang faible, plus un bruit blanc gaussien indépendant :

$$\mathbf{z} = \mathbf{c} + \mathbf{b} \tag{1.69}$$

où $\mathbf{c} \sim \mathcal{CE}(\mathbf{0}, \boldsymbol{\Sigma}_c, g)$ et $\mathbf{b} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$. La matrice de covariance totale du vecteur $\mathbf{z}$ observé $\boldsymbol{\Sigma}_{tot} = \sigma_g \boldsymbol{\Sigma}_c + \mathbf{I}_M$ est donc bien de structure rang faible.

Notons d'ores et déjà que la somme d'un CES et d'un bruit blanc gaussien ne peut pas s'exprimer simplement comme un CES et nécessite une étude à part entière. C'est pourquoi nous présenterons et considérerons la vraisemblance associée à ce modèle spécifique au cours des chapitres suivants. Pour tirer parti du théorème de représentation 1.2.3.1, nous nous limiterons néanmoins à la famille des CG pour modéliser les sources hétérogènes (modèle détaillé et justifié dans le chapitre suivant).

1.5 Synthèse du chapitre 1

Dans cette partie, nous avons d'abord présenté les distributions CES, famille générale couvrant bon nombre de distributions. Ce modèle très générique permet de s'affranchir de l'hypothèse gaussienne (donc modéliser des distributions plus ou moins hétérogènes) tout en gardant une formalisation de la distribution entièrement paramétrée par les statistiques d'ordre un et deux.

Nous avons ensuite évoqué les méthodes d'estimation les plus usuelles de la matrice de covariance (ou de sa forme) : la SCM, la NSCM, et le maximum de vraisemblance des distributions CES (MVCES). Enfin nous avons présenté les M-estimateurs : ces estimateurs se voient comme une généralisations des MVCES et sont robustes aux différentes distributions de signaux (ainsi qu'à de potentielles corruptions). Les M-estimateurs nécessitent cependant un nombre de données $K > M$ pour être calculés. Nous avons donc aussi évoqué les estimateurs robustes régularisés comme potentielles solutions à ce problème.

Une autre solution pour palier le problème du nombre limité de données est de se tourner vers l'estimation de matrices structurées. En effet, exploiter une connaissance a priori sur la structure de la matrice de covariance permet de réduire significativement le nombre de données nécessaires à son estimation. Nous avons donc dressé un bref état de l'art de l'estimation de CM structurée. Initialement adaptée au contexte gaussien, ce cadre s'étend peu à peu à l'ensemble des estimateurs robustes.

Nous avons cependant mis en évidence un manque à combler en ce qui concerne l'estimation de matrice à structure rang faible. Cette structure spécifique correspond à une réalité pour de nombreux systèmes, c'est pourquoi il semble intéressant de la considérer en particulier. Des algorithmes adaptés à ce contexte existent dans le cas gaussien mais ils peuvent être mis en défaut si le bruit est hétérogène. La motivation de cette thèse est donc de proposer des estimateurs et traitements adaptés au contexte de distributions CES de rang faible. Ceci afin de profiter simultanément de deux propriétés que sont la robustesse aux différentes distributions ou aux erreurs de modélisation, et la réduction du nombre de données secondaires nécessaires à l'estimation.

Il est important de noter que cette thèse se concentrera sur des problématiques où le rang R est considéré comme connu. L'infirmation a priori sur ce rang peut, dans certaines cas, être directement obtenue par la connaissance du système (c'est le cas pour le STAP [120]). Dans d'autres cas, on suppose le nombre de sources connues avant d'effectuer le traitement (comme pour l'algorithme MUSIC [101]). Autrement, on supposera que ce rang a été estimé au préalable suffisamment précisément. Pour ne rien négliger, on renverra le lecteur vers ces références concernant l'estimation de rang [7, 10, 109, 93, 42].

Chapitre 2

Estimation de la matrice de covariance en contexte hétérogène rang faible

Dans ce chapitre, nous présentons le modèle considéré au cours de cette thèse ainsi que les algorithmes d'estimations développés pour ce contexte. Nous rappelons d'abord rapidement les motivations qui conduisent à considérer ce modèle en particulier. Le cadre théorique est ensuite formellement détaillé au travers de l'expression de la vraisemblance (ainsi que diverses relaxations). Nous dérivons ensuite l'expression de l'estimateur du maximum de vraisemblance de la matrice de covariance de ce modèle. Cet estimateur n'est pas directement tractable et apparaît comme un point fixe d'une fonctionnelle non explicite. C'est pourquoi nous avons développé différents algorithmes afin de le calculer, ou de s'en approcher. Les performances des estimateurs proposés sont finalement illustrées au travers de simulations de validation.

En vue d'alléger ce chapitre, les annexes comprennent les calculs et preuves des théorèmes (partie A). L'annexe B comprend l'article dérivant les algorithmes développées en collaboration avec l'équipe du Professeur Daniel P. Palomar, au Department of Electronic and Computer Engineering de Hong Kong University of Science and Technology (HKUST).

2.1 Motivations

Le modèle que nous considérons (ainsi que les traitements développés) à été initialement motivé par des applications radar. En effet, en radar, le bruit reçu se compose :

- de la réponse de l'environnement au signal émis : une interférence que nous appellerons fouillis par la suite. Dû aux hautes résolutions des systèmes actuels, ce fouillis ne peut plus être modélisé avec précision par un processus gaussien. Pour modéliser ce bruit hétérogène, les distributions CG offrent une alternative en adéquation avec des données réelles [58, 121, 123, 25, 26, 97, 44, 85]. De plus, le fouillis est bien souvent contenu dans un sous-espace de dimension inférieure à la taille des données : sa matrice de covariance est donc de rang faible. Dans certains cas, le rang du fouillis peut être connu par des considérations sur la géométrie du système [18, 120].
- du bruit thermique, modélisé par un bruit blanc additif gaussien indépendant du fouillis et correspondant au bruit de l'électronique du système ($e.g.$ bruit de capteurs...).

En définitive, la somme d'un vecteur CG rang faible plus un bruit blanc gaussien indépendant, correspond à un modèle réaliste dans ce contexte [95, 96, 8, 94, 50].

Ce type de distribution n'est pas couverte par la classe des CES et c'est pourquoi son étude à part entière s'avère intéressante. De plus, le modèle associé à ces données offre la possibilité de développer d'autres traitements spécifiques, comme nous le verrons au chapitre 3.

Ajoutons que le radar n'est pas la seule application où les signaux d'intérêt ont une structure rang faible. C'est pourquoi l'extension des travaux présentés à d'autres domaines est une perspective fortement suggérée.

2.2 Modèle

On suppose que K échantillons $\{\mathbf{z}_k\} \in \mathbb{C}^M$ sont disponibles. Ces échantillons $\mathbf{z}_k \in \mathbb{C}^M$ sont des réalisations $i.i.d.$ d'un fouillis CG $\mathbf{c}_k$ plus un bruit blanc gaussien indépendant $\mathbf{n}_k$:

$$\mathbf{z}_k = \mathbf{n}_k + \mathbf{c}_k \tag{2.1}$$

tous deux circulaires au second ordre et de moyenne nulle.

Le bruit blanc gaussien est donc distribué selon :

$$\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_M) \, , \tag{2.2}$$

où nous supposons la puissance σ^2 connue. Cette hypothèse est faite afin de présenter un cadre théorique tractable et plus léger. Si σ^2 est inconnu en pratique, un estimateur consistant peut simplement remplacer la vraie valeur dans les résultats dérivés. Sous cette hypothèse, on peut donc supposer la puissance du bruit blanc unitaire $\sigma^2 = 1$. En effet, il suffit alors de considérer les données remises à l'échelle $\mathbf{z}'_k = \mathbf{z}_k/\sigma$ qui ne modifie pas le rapport de puissance fouillis à bruit. Le fouillis $\mathbf{c}_k$ est modélisé comme un vecteur CG [1]. D'après le théorème de représentation 1.2.3.1 section 1.2.3, il peut donc se représenter comme un vecteur gaussien de moyenne nulle et de matrice de covariance $\mathbf{\Sigma}_c$, multiplié par la racine carrée d'une variable aléatoire positive indépendante (la texture) τ_k. Nous ne supposons aucune information sur la densité de probabilité des textures, et chaque texture τ_k sera considérée comme un paramètre déterministe inconnu. Chaque $\mathbf{c}_k$ est donc distribué, conditionnellement à τ_k, selon :

$$(\mathbf{c}_k | \tau_k) \sim \mathcal{CN}(\mathbf{0}, \tau_k \mathbf{\Sigma}_c) \ \, , \tag{2.3}$$

La covariance $\mathbf{\Sigma}_c$ est de rang $R < M$, supposé connu. Soit sa SVD donnée par ses vecteurs propres $\mathbf{v}_r$ et ses valeurs propres correspondantes c_r pour $r \in [\![1, R]\!]$:

$$\mathbf{\Sigma}_c = \sum_{r=1}^{R} c_r \mathbf{v}_r \mathbf{v}_r^H = \mathbf{V}_c \mathbf{C}_c \mathbf{V}_c^H \, , \tag{2.4}$$

où l'on note $\mathbf{C}_c = \text{diag}(\{c_r\})$ la matrice diagonale $R \times R$ des valeurs propres, et $\mathbf{V}_c$ la matrice $M \times R$ concaténant les vecteurs propres. Ceux-ci formant une base orthonormée, ils vérifient $\mathbf{V}_c^H \mathbf{V}_c = \mathbf{I}_R$. On définit de plus le projecteur orthogonal sur le sous-espace fouillis :

$$\mathbf{\Pi}_c = \sum_{r=1}^{R} \mathbf{v}_r \mathbf{v}_r^H = \mathbf{V}_c \mathbf{V}_c^H \ \, , \tag{2.5}$$

ainsi que son projecteur complémentaire $\mathbf{\Pi}_c^\perp = \mathbf{I}_M - \mathbf{\Pi}_c$. La complétion de la base $\mathbf{V}_c$ est un ensemble de $M - R$ vecteurs $\{\mathbf{v}_r\}_{r \in [\![R+1, M]\!]}$, concaténés en une matrice $\mathbf{V}_c^\perp$ vérifiant $\mathbf{\Pi}_c^\perp = \mathbf{V}_c^\perp \mathbf{V}_c^{\perp H}$. Par définition, la concaténation $\mathbf{V} = \left[\mathbf{V}_c \mathbf{V}_c^\perp\right]$ vérifie alors $\mathbf{V}^H \mathbf{V} = \mathbf{V}\mathbf{V}^H = \mathbf{I}_M$.

1. Nous ne considérons pas la classe plus large des CES car seul le théorème de représentation des CG permet d'aboutir à un modèle tractable en Eq.(2.6). Cependant, une robustesse des estimateurs développés peut être attendue sur l'ensemble des CES, par analogie avec les M-estimateurs.

Étant donné le modèle considéré, chaque échantillon $\mathbf{z}_k$ est distribué conditionnellement à τ_k comme (somme de deux variables aléatoires gaussiennes) :

$$(\mathbf{z}_k|\tau_k) \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_k) \ , \tag{2.6}$$

avec

$$\mathbf{\Sigma}_k = \mathbf{I}_M + \tau_k \mathbf{\Sigma}_c \ . \tag{2.7}$$

La vraisemblance (conditionnelle aux τ_k) est alors donnée par :

$$f(\{\mathbf{z}_k\}|\{\mathbf{\Sigma}_c\}, \{\tau_k\}) = \prod_{k=1}^{K} \frac{e^{-\mathbf{z}_k^H \mathbf{\Sigma}_k^{-1} \mathbf{z}_k}}{\pi^M |\mathbf{\Sigma}_k|} \ . \tag{2.8}$$

Et la log-vraisemblance est (sans les termes constants) :

$$\log f(\{\mathbf{z}_k\}|\{\mathbf{\Sigma}_c\}, \{\tau_k\}) = -\sum_{k=1}^{K} \mathbf{z}_k^H \mathbf{\Sigma}_k^{-1} \mathbf{z}_k - \sum_{k=1}^{K} \ln |\mathbf{\Sigma}_k| \ . \tag{2.9}$$

En utilisant la SVD de $\mathbf{\Sigma}_k$, on obtient :

$$\mathbf{z}_k^H \mathbf{\Sigma}_k^{-1} \mathbf{z}_k = \mathbf{z}_k^H \mathbf{V}_c (\tau_k \mathbf{C}_c + \mathbf{I}_R)^{-1} \mathbf{V}_c^H \mathbf{z}_k + \mathbf{z}_k^H \mathbf{V}_c^{\perp} \mathbf{V}_c^{\perp H} \mathbf{z}_k \ , \tag{2.10}$$

ainsi que :

$$\sum_{k=1}^{K} \ln |\mathbf{\Sigma}_k| = \sum_{k=1}^{K} \sum_{r=1}^{R} \log(1 + \tau_k c_r) \ . \tag{2.11}$$

Afin de contraindre intrinsèquement l'orthogonalité entre les bases $\mathbf{V}_c$ et $\mathbf{V}_c^{\perp}$ dans le processus d'estimation [2], on utilise le changement de variable $\mathbf{V}_c^{\perp} \mathbf{V}_c^{\perp H} = \mathbf{I}_M - \mathbf{V}_c \mathbf{V}_c^H = \mathbf{I}_M - \mathbf{\Pi}_c$, conduisant à :

$$\log(f\left(\{\mathbf{z}_k\}\right) = -\sum_{k=1}^{K} \mathbf{z}_k^H \mathbf{V}_c (\tau_k \mathbf{C}_c + \mathbf{I}_R)^{-1} \mathbf{V}_c^H \mathbf{z}_k - \sum_{k=1}^{K} \mathbf{z}_k^H (\mathbf{I}_M - \mathbf{\Pi}_c) \mathbf{z}_k - \sum_{k=1}^{K} \sum_{r=1}^{R} \log(1 + \tau_k c_r) \tag{2.12}$$

L'inversion de la matrice du premier terme donne :

$$\mathbf{V}_c (\tau_k \mathbf{C}_c + \mathbf{I}_R)^{-1} \mathbf{V}_c^H = \sum_{r=1}^{R} \frac{1}{\tau_k c_r + 1} \mathbf{v}_r \mathbf{v}_r^H \ . \tag{2.13}$$

et le développement $\mathbf{z}_k^H (\mathbf{I}_M - \mathbf{\Pi}_c) \mathbf{z}_k = \mathbf{z}_k^H \mathbf{z}_k - \sum_{r=1}^{R} \mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k$, conduisent (ignorant les termes constants) à la log-vraisemblance suivante :

$$\log f = \sum_{k=1}^{K} \sum_{r=1}^{R} \left[\frac{\tau_k c_r}{1 + \tau_k c_r} \mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k - \log(1 + \tau_k c_r) \right] \tag{2.14}$$

2. Une relaxation de cette contrainte sera envisagée dans le chapitre suivant.

2.3 Maximum de vraisemblance de la matrice de covariance du fouillis CG rang faible

Considérant le modèle précédemment décrit, l'estimateur du maximum de vraisemblance (EMV) de la matrice de covariance $\mathbf{\Sigma}_c$, paramétrée par sa SVD $\{c_r, \mathbf{v}_r\}$ pour $r \in [\![1, R]\!]$, fait l'objet du théorème suivant :

Théorème 2.3.1 *L'EMV des paramètres $\{\mathbf{v}_r\}$ pour $r \in [\![1, R]\!]$, est la solution du problème suivant :*

$$\begin{aligned} \underset{\{\mathbf{v}_r\}}{max} \quad & \sum_{r=1}^{R} \mathbf{v}_r^H \hat{\mathbf{M}}_r \mathbf{v}_r \\ s.c. \quad & \mathbf{v}_r^H \mathbf{v}_r = 1 \ , \ r \in [\![1, R]\!] \\ & \mathbf{v}_r^H \mathbf{v}_j = 0 \ , \ r, j \in [\![1, R]\!] \ , \ r \neq j \end{aligned}$$

où les matrices $\hat{\mathbf{M}}_r$ sont définies :

$$\hat{\mathbf{M}}_r = \sum_{k=1}^{K} \frac{\hat{c}_r \hat{\tau}_k}{1 + \hat{c}_r \hat{\tau}_k} \mathbf{z}_k \mathbf{z}_k^H \ , \tag{2.15}$$

avec $\hat{c}_r(\{\hat{\mathbf{v}}_r\})$ et $\hat{\tau}_k(\{\hat{\mathbf{v}}_r\})$ les EMV de c_r et τ_k respectivement (dont la dépendance en $\{\hat{\mathbf{v}}_r\}$ est omise pour alléger la notation de $\hat{\mathbf{M}}_r$).

Preuve 2.3.1 *$c.f.$ Annexe A.*

Néanmoins, cette formulation ne permet pas directement d'obtenir la solution puisque le MLE des paramètres c_r et τ_k ne sont pas tractables ($c.f.$ preuve A.1.3 en annexe A). De plus, l'EMV de ces paramètres dépend de l'EMV des paramètres $\{\mathbf{v}_r\}$, impliquant que l'estimateur recherché soit un point fixe d'une fonction non explicite.

Précisions qu'aucune considération préalable ne permet de garantir l'unicité de la solution de l'EMV. En effet, la vraisemblance (2.14) n'est pas convexe selon les paramètres $\{\tau_k\}$ et $\{c_r\}$. De plus, les contraintes unitaires sur $\{\mathbf{v}_r\}$ ne sont pas convexes. Les précédents résultats concernant l'unicité d'estimateurs robustes [89, 125, 124] ne s'appliquent pas ici car la distribution considérée n'est pas un CES de rang plein.

C'est pourquoi les sections suivantes proposent des algorithmes itératifs pour calculer un maximum de vraisemblance local (ou de s'en approcher au mieux).

2.4 Premier algorithme : 2-Step approché

2.4.1 Relaxation au travers de variables indépendantes $d_r^k = \tau_k c_r$

De prime abord, la vraisemblance initiale du modèle s'avère difficile à manipuler. Afin de palier les problèmes d'identifiabilité et de tractablité, nous avons donc envisagé dans un premier temps une relaxation du modèle en introduisant les variables $d_r^k = \tau_k c_r$. Ces facteurs représentent la puissance du fouillis pour chaque réalisation k et direction $\mathbf{v}_r$. La vraisemblance relaxée apparaît alors comme :

$$\log f = \sum_{k=1}^{K} \sum_{r=1}^{R} \left[\frac{d_r^k}{1 + d_r^k} \mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k - \log(1 + d_r^k) \right] \tag{2.16}$$

où les paramètres $\{d_r^k\}$ sont considérés comme déterministes inconnus et positifs. Cette relaxation peut apparaitre plus "générale" puisqu'elle implique plus de degrés de libertés ($R \times K$ au lieu de $R + K$). Cependant, ce modèle ne prend pas en compte le lien inhérent entre les d_r^k's, ce qui peut conduire à une mauvaise estimation de ces paramètres. Ce problème sera discuté dans les sections suivantes.

2.4.2 Description de l'algorithme et propriétés

Ce premier algorithme visant à approcher l'EMV du modèle considéré est basé sur une approche 2-Step. Le principe général est de maximiser une vraisemblance relaxée (2.16) de manière alternée selon les paramètres d'intérêt $\{d_r^k\}$ et $\{\mathbf{v}_r\}$. La relaxation considérée permet en effet d'obtenir une formule tractable de l'EMV des paramètres relaxés $\{d_r^k\}$. De plus, afin de recouvrer les vrais paramètres d'intérêt, nous introduisons une méthode de régularisation "ad-hoc", conduisant à des estimateurs $\{\tau_k\}$ et $\{c_r\}$ (qui ne sont donc pas des EMV). Nous vérifions a posteriori qu'alterner les étapes de maximisation-régularisation augmente bien la vraisemblance. Néanmoins, cet algorithme ne garantit pas d'atteindre un EMV car les paramètres régularisés $\{\tau_k\}$ et $\{c_r\}$ ne sont pas des EMV.

Les deux étapes de l'algorithme sont décrites ci dessous et l'algorithme complet, résumant ces étapes est détaillé dans l'encadré 1 "MLE-2SD" (où D fait référence aux paramètres relaxés d_r^k).

2.4.3 Étape 1 : Estimation des textures et valeurs propres via régularisation des EMV $\{\hat{d}_r^k\}$

Cette étape considère les paramètres $\{\mathbf{v}_r\}$, pour $r \in [\![1, R]\!]$, connus et fixés dans la vraisemblance (2.16). On a alors le Théorème suivant :

Théorème 2.4.1 *L'EMV sous contrainte de positivité des paramètres d_r^k, noté $\hat{d}_r^k$ pour $r \in [\![1, R]\!]$ et $k \in [\![1, K]\!]$ est :*

$$\hat{d}_r^k = \begin{cases} \mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k - 1 & \textit{si } \mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k > 1 \\ 0 & \textit{sinon} \end{cases}, \tag{2.17}$$

Preuve 2.4.1 *c.f. Annexe A.*

Comme vu précédemment, un bruit CG correspond à un cas particulier du modèle relaxé avec $d_r^k = \tau_k c_r$, où τ_k est la texture du k^{ieme} échantillon et c_r est la r^{ieme} valeur propre de la matrice de

Algorithm 1 MLE-2SD

Initialiser $\{\mathbf{v}_r^{(0)}\}$ avec les R vecteurs propres les plus forts de la SCM
repeat
  Estimer $\{d_r^k\}$ conditionnellement à $\{\mathbf{v}_r\}$ avec le Théorème 2.4.1
  Calculer $\tilde{\tau}_k$ et $\tilde{c}_r$ avec (2.23) et (2.24)
  Régulariser les $\{d_r^k\} \leftarrow \{\tilde{d}_r^k\}$ avec (2.25)
  Calculer le jeu de matrices $\{\mathbf{M}_r\}$ avec (2.15)
  Utiliser l'Algorithme 15 de [73] pour maximiser

$$\sum_{r=1}^{R} \mathbf{v}_r^H \mathbf{M}_r^{(n)} \mathbf{v}_r$$

  sous contrainte unitaire, la solution met à jour les variables $\{\mathbf{v}_r\}$
until Convergence
$\hat{\mathbf{\Pi}}_c = \sum_{r=1}^{R} \mathbf{v}_r \mathbf{v}_r^H$
$\hat{\mathbb{E}}(\tau) = \text{mean}(\{\tilde{\tau}_k\})$
$\hat{\mathbf{\Sigma}}_c = \sum_{r=1}^{R} \tilde{c}_r \mathbf{v}_r \mathbf{v}_r^H$
$\hat{\mathbf{\Sigma}}^{MLE} = \hat{\mathbb{E}}(\tau)\hat{\mathbf{\Sigma}}_c + \mathbf{I}_M$

covariance du fouillis. Nous proposons maintenant une méthode pour estimer les paramètres τ_k et c_r à l'aide de l'EMV des d_r^k obtenu théorème 2.4.1. En effet, les paramètres d_r^k étant corrélés, il est possible d'utiliser leur structure intrinsèque afin d'améliorer le processus d'estimation. De plus, ce processus s'avère nécessaire pour obtenir des estimés des paramètres d'intérêt initiaux $\{\tau_k\}$ et $\{c_r\}$. Tout d'abord, fixons une condition d'itentifiabilité. En effet, pour tout facteur d'échelle $\gamma \neq 0$ les couples $\{\tau_k, c_r\}$ et $\{\gamma\tau_k, c_r/\gamma\}$ peuvent décrire la même solution $d_r^k = \tau_k c_r$. En accord avec les précédent travaux sur l'estimation de matrice de covariance, nous utiliserons la normalisation classique $\text{Tr}(\mathbf{\Sigma}_c) = R$, imposant alors la contrainte $\sum_{r=1}^{R} c_r = R$. Considérons la matrice des vrais paramètres $[\mathbf{D}]_{k,r} = d_r^k$:

$$\mathbf{D} = \begin{pmatrix} d_1^1 & \cdots & d_1^K \\ \vdots & \ddots & \vdots \\ d_R^1 & \cdots & d_R^K \end{pmatrix} , \tag{2.18}$$

Nous visons à retrouver les paramètres τ_k et c_r à partir de $\mathbf{D}$. Puisque $d_r^k = \tau_k c_r$ le problème est de trouver deux vecteurs $\boldsymbol{\tau}$ (les estimateurs de textures concaténés) et $\mathbf{c}$ (les estimateurs de valeurs propres concaténées) qui vérifient :

$$\mathbf{D} = \mathbf{c}\boldsymbol{\tau}^T = \begin{pmatrix} c_1 \\ \vdots \\ c_R \end{pmatrix} \begin{pmatrix} \tau_1 & \cdots & \tau_K \end{pmatrix} . \tag{2.19}$$

La contrainte d'identifiabilité impose impose $\sum_{r=1}^{R} c_r = R$, donc chaque colonne de $\mathbf{D}$ vérifie :

$$\sum_{r=1}^{R} d_r^k = \sum_{r=1}^{R} c_r \tau_k = \tau_k R , \tag{2.20}$$

conduisant à :

$$\tau_k = \sum_{r=1}^{R} d_r^k / R . \tag{2.21}$$

on peut alors obtenir les c_r's au travers des ratios suivants :

$$c_r = \frac{\sum_{k=1}^{K} d_r^k}{\sum_{k=1}^{K} \tau_k} \tag{2.22}$$

De manière identique, nous proposons de dériver des estimateurs régularisés de d_r^k, notés $\tilde{d}_r^k$, en appliquant le même procédé sur la matrice des EMV $[\hat{\mathbf{D}}]_{k,r} = \hat{d}_r^k$, avec $\hat{d}_r^k$ donné en Théorème 2.4.1. La régularisation est alors obtenue avec :

$$\tilde{\tau}_k = \sum_{r=1}^{R} \hat{d}_r^k / R \; . \tag{2.23}$$

$$\tilde{c}_r = \frac{\sum_{k=1}^{K} \hat{d}_r^k}{\sum_{k=1}^{K} \tau_k} \tag{2.24}$$

d'où

$$\tilde{d}_r^k = \tilde{\tau}_k \tilde{c}_r = \frac{\sum_{r=1}^{R} \hat{d}_r^k \sum_{k=1}^{K} \hat{d}_r^k}{\sum_{k=1}^{K} \sum_{r=1}^{R} \hat{d}_r^k} \; . \tag{2.25}$$

Pour conclure, l'EMV exact des paramètres $\{ \{\tau_k\} \, , \, \{c_r\} \}$ n'est pas tractable. Cependant, l'EMV des paramètres relaxés $d_r^k = \tau_k c_r$ l'est. De plus, il est possible de dériver des estimés de $\{\tau_k, c_r\}$ (qui ne sont pas des EMV) à l'aide de l'ensemble $\{\hat{d}_r^k\}$ au moyen de considérations sur leur structure intrinsèque. Les paramètres obtenus $\{ \{\tilde{\tau}_k\} \, , \, \{\tilde{c}_r\} \}$ peuvent êtres utilisés pour obtenir des estimateurs régularisés $\{\tilde{d}_r^k\}$, en ce sens qu'ils satisfont la structure imposée par le modèle non relaxé.

2.4.4 Étape 2 : Estimation du sous-espace fouillis pour textures et valeurs propres fixées

On suppose désormais que les textures $\{\tau_k\}$ et valeurs propres $\{c_r\}$ sont connues et fixées par la précédente étape. La vraisemblance (2.16) s'exprime :

$$\log f = \sum_{r=1}^{R} \mathbf{v}_r^H \mathbf{M}_r \mathbf{v}_r + \gamma \tag{2.26}$$

où γ est une constante ne dépendant pas de $\{\mathbf{v}_r\}$, et où les matrices $\mathbf{M}_r$ sont définies :

$$\mathbf{M}_r = \sum_{k=1}^{K} \left(\frac{d_r^k}{1 + d_r^k} \right) \mathbf{z}_k \mathbf{z}_k^H \tag{2.27}$$

Cette vraisemblance doit être maximisée par rapport aux paramètres $\{\mathbf{v}_r\}$ sous contrainte d'orthonormalité $\mathbf{V}_c^H \mathbf{V}_c = \mathbf{I}_R$. L'EMV de la base du sous-espace fouillis est donc la solution du problème d'optimisation sous contrainte suivant :

$$\begin{aligned} \max_{\{\mathbf{v}_r\}} \quad & f_0(\{\mathbf{v}_r\}) = \sum_{r=1}^{R} \mathbf{v}_r^H \mathbf{M}_r \mathbf{v}_r \\ \text{s.c.} \quad & \mathbf{v}_r^H \mathbf{v}_r = 1 \; , \; r \in [\![1, R]\!] \\ & \mathbf{v}_r^H \mathbf{v}_j = 0 \; , \; r, j \in [\![1, R]\!] \; , \; r \neq j \end{aligned} \tag{A}$$

Intuitivement, ce problème s'apparente à celui de la SVD [3]. Néanmoins, les matrices $\mathbf{M}_r$ n'étant pas identiques, la solution n'est pas directement contenue dans leurs vecteurs propres. A notre connaissance, il n'existe pas de solution directe à ce problème, c'est pourquoi des algorithmes pour atteindre un maximum local de la fonction f_0 doivent êtres utilisés.

Nous avons utilisé des algorithmes d'optimisation sous contrainte unitaire $\mathbf{V}_c^H \mathbf{V}_c = \mathbf{I}_R$: notamment, l'algorithme 15 de [73] appelé "Descente de Gradient sur la variété de Stiefel". Ce type d'algorithme assure d'atteindre un maximum local formant une base orthonormée, donc un EMV valide pour cette étape. Notons aussi que d'autres algorithmes d'optimisation adaptés à ce contexte existent [35, 4]. Enfin, une autre possibilité est d'approcher la solution au moyen de relaxations. Ces types de solutions seront envisagées et développées dans le chapitre suivant.

2.4.5 Dernière étape : Estimation de facteur d'échelle

Cette étape n'est pas incluse dans les itérations associée à l'algorithme 2-Step. Néanmoins, elle s'avère nécessaire si l'application envisagée requiert un estimateur de matrice de covariance totale. En effet, la matrice de covariance totale du processus CG plus bruit blanc est :

$$\mathbf{\Sigma}_{tot} = \mathbb{E}(\tau)\mathbf{\Sigma}_c + \mathbf{I}_M \tag{2.28}$$

c'est pourquoi le rapport fouillis à bruit doit être estimé afin de résoudre une ambiguïté d'échelle.

Les textures étant considérées comme déterministes inconnues la matrice de covariance totale peut être estimée via :

$$\hat{\mathbf{\Sigma}} = \hat{\mathbb{E}}(\tau)\hat{\mathbf{\Sigma}}_c + \mathbf{I}_M \tag{2.29}$$

avec $\hat{\mathbb{E}}(\tau)$ la moyenne empirique des textures estimées :

$$\hat{\mathbb{E}}(\tau) = \frac{1}{K}\sum_{k=1}^{K} \hat{\tau}_k \tag{2.30}$$

3. Une interprétation plus approfondie, ainsi que diverses modifications et relaxations du problème seront présentés dans le chapitre suivant.

2.5 Deuxième algorithme : 2-Step exact sous hypothèse de fort rapport fouillis à bruit

2.5.1 Seconde relaxation : hypothèse de fort rapport fouillis à bruit

La seconde relaxation du modèle consiste à supposer qu'un fort rapport de puissance fouillis à bruit est présent. Cette approche est réaliste pour la plupart des systèmes, où la puissance du signal d'intérêt (les valeurs propres de $\mathbf{\Sigma}_c$) est bien supérieure à la puissance du bruit blanc σ^2 (ici fixée arbitrairement à 1). Suivant cette hypothèse, il est alors possible de négliger le bruit blanc devant le fouillis sur le sous-espace $\mathbf{\Pi}_c$. Rappelons que conditionnellement à τ_k, $(\mathbf{z}_k|\tau_k) \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_k)$, où la vraie matrice de covariance $\mathbf{\Sigma}_c$ s'exprime :

$$\mathbf{\Sigma}_k = \left[\begin{array}{cc} \mathbf{V}_c & \mathbf{V}_c^{\perp} \end{array} \right] \begin{pmatrix} \tau_k \mathbf{C}_c + \mathbf{I}_R & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{M-R} \end{pmatrix} \left[\begin{array}{cc} \mathbf{V}_c & \mathbf{V}_c^{\perp} \end{array} \right]^H \tag{2.31}$$

L'hypothèse fort rapport fouillis à bruit peut se formuler

$$1 \ll \tau_k c_r \ , \ \forall\, k,\ r \tag{2.32}$$

conduisant alors à :

$$\mathbf{\Sigma}_k \approx \left[\begin{array}{cc} \mathbf{V}_c & \mathbf{V}_c^{\perp} \end{array} \right] \begin{pmatrix} \tau_k \mathbf{C}_c & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{M-R} \end{pmatrix} \left[\begin{array}{cc} \mathbf{V}_c & \mathbf{V}_c^{\perp} \end{array} \right]^H \tag{2.33}$$

On considère alors la vraisemblance relaxée :

$$f(\{\mathbf{z}_k\}) = -\sum_{k=1}^{K} \mathbf{z}_k^H \mathbf{\Sigma}_k^{-1} \mathbf{z}_k - \sum_{k=1}^{K} \ln |\mathbf{\Sigma}_k| \tag{2.34}$$

avec $\mathbf{\Sigma}_k$ défini par (2.33).

Cette relaxation peut aussi s'interpréter comme une approche robuste par rapport au modèle considéré. On suppose alors que la somme du fouillis et du bruit blanc se comporte comme un vecteur CG de loi inconnue. C'est ce type de raisonnement qui est couramment utilisé, par exemple pour les systèmes radar [28, 32, 45, 89, 84, 2, 13, 53] où les signaux reçus sont communément supposés CG (ou CES) de rang plein et où le bruit blanc est ignoré.

2.5.2 Description et propriétés de l'algorithme

Ce second algorithme vise à atteindre l'EMV du modèle relaxé décrit précédemment en section 2.5.1 par une méthode "2-Step". Le principe est donc de maximiser une vraisemblance relaxée de manière alternée selon les paramètres d'intérêt $\{\{\tau_k\}, \mathbf{C}_c\}$ et $\mathbf{V}_c = \mathbb{U}(\{\mathbf{v}_r\})$ (où $\mathbb{U}$ désigne l'opérateur de concaténation). La relaxation considérée permet en effet d'obtenir une formule tractable de l'EMV des paramètres relaxés $\{\{\tau_k\}, \mathbf{C}_c\}$. Alterner les étapes de maximisation assure donc une convergence vers un maximum local de la vraisemblance relaxée. On vérifiera a posteriori qu'effectuer cette hypothèse de fort rapport fouillis à bruit est acceptable. De plus, comme évoqué section 2.3, il n'est pas possible de garantir l'unicité de la solution et l'on se contentera donc d'un EMV local.

Les deux étapes de l'algorithme sont décrites ci dessous et l'algorithme complet, résumant ces étapes est détaillé dans l'encadré 2 "MLE-2SR" (où R fait référence à "robuste").

Algorithm 2 MLE-2S-R

Initialiser $\mathbf{V}_c$ et $\mathbf{U}_c$ avec les R vecteurs propres associés aux plus grandes valeurs propres de la SCM
repeat
  Calculer $\mathbf{M}_c$ avec les itérations (2.37) pour $\mathbf{U}_c$ fixé
  $\hat{\mathbf{\Sigma}}_c = \mathbf{U}_c\mathbf{M}_c\mathbf{U}_c^H$
  $\hat{\mathbf{\Sigma}}_c \xrightarrow[\text{SVD}]{} \mathbf{C}_c = \text{diag}(\{\hat{c}_r\})\ ,\ \mathbf{V}_c = \mathbb{U}(\{\mathbf{v}_r\})$
  $\hat{\tau}_k = \mathbf{z}_k\mathbf{U}_c\mathbf{M}_c^{-1}\mathbf{U}_c^H\mathbf{z}_k/R \quad \forall k$
  $\mathbf{M}_r = \sum_{k=1}^K \frac{\hat{c}_r\hat{\tau}_k}{1+\hat{c}_r\hat{\tau}_k}\mathbf{z}_k\mathbf{z}_k^H \quad \forall r$
  mettre à jour $\mathbf{U}_c$ comme solution du Problème A à l'aide de l'algorithme 15 de [73]
until Convergence
$\hat{\mathbf{\Pi}}_c^{MLE} = \mathbf{V}_c\mathbf{V}_c^H$
$\hat{\mathbb{E}}(\tau) = \text{mean}(\{\tau_k\})$
$\hat{\mathbf{\Sigma}}^{MLE} = \hat{\mathbb{E}}(\tau)\hat{\mathbf{\Sigma}}_c + (\mathbf{I}_M - \hat{\mathbf{\Pi}}_c^{MLE})$

2.5.3 Étape 1 : Estimation des textures et valeurs propres grâce à la relaxation fort rapport fouillis à bruit

Considérons maintenant la relaxation fort rapport fouillis à bruit du modèle, nous introduisons l'intermédiaire de notation suivant : soit $\mathbf{U}_c$ une base arbitraire du sous-espace fouillis [4]. Identiquement à $\mathbf{V}_c$, cette base engendre le sous-espace fouillis $\mathbf{U}_c\mathbf{U}_c^H = \mathbf{V}_c\mathbf{V}_c^H = \mathbf{\Pi}_c$. Les bases $\mathbf{U}_c$ et $\mathbf{V}_c$ sont donc égales à une rotation près. On définit la base complémentaire $\mathbf{U}_c^\perp$ telle que la concaténation $\mathbf{U} = \left[\ \mathbf{U}_c\ \mathbf{U}_c^\perp\ \right]$ vérifie $\mathbf{U}\mathbf{U}^H = \mathbf{U}^H\mathbf{U} = \mathbf{I}_M$. La matrice de covariance associée à l'échantillon $(\mathbf{z}_k|\tau_k) \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_k)$ s'exprime alors aussi dans cette base :

$$\mathbf{\Sigma}_k \approx \left[\ \mathbf{U}_c\ \ \mathbf{U}_c^\perp\ \right] \begin{pmatrix} \tau_k\mathbf{M}_c & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{M-R} \end{pmatrix} \left[\ \mathbf{U}_c\ \ \mathbf{U}_c^\perp\ \right]^H , \tag{2.35}$$

où la matrice "noyau" $\mathbf{M}_c$ n'est donc pas nécessairement diagonale.

L'EMV des paramètres d'intérêt s'obtient alors à l'aide du théorème suivant :

Théorème 2.5.1 *Pour une base $\mathbf{U}_c$ fixée, l'EMV du noyau $\mathbf{M}_c$ est défini comme l'unique solution de point fixe :*

$$\hat{\mathbf{M}}_c = \frac{R}{K}\sum_{k=1}^K \frac{\mathbf{U}_c^H\mathbf{z}_k\mathbf{z}_k^H\mathbf{U}_c}{\mathbf{z}_k^H\mathbf{U}_c\hat{\mathbf{M}}_c^{-1}\mathbf{U}_c^H\mathbf{z}_k} \tag{2.36}$$

obtenu via les itérations :

$$\mathbf{M}_{(n+1)} = \frac{R}{K}\sum_{k=1}^K \frac{\mathbf{U}_c^H\mathbf{z}_k\mathbf{z}_k^H\mathbf{U}_c}{\mathbf{z}_k^H\mathbf{U}_c\mathbf{M}_{(n)}^{-1}\mathbf{U}_c^H\mathbf{z}_k} \tag{2.37}$$

4. Rappelons que pour cette étape, ce paramètre est supposé connu et fixé.

qui convergent vers la solution $\hat{\mathbf{M}}_c$ si $K > R$ et qu'aucun $\mathbf{U}_c^H\mathbf{z}_k$ ne se soit un vecteur nul. De plus, l'EMV des textures est obtenu par :

$$\hat{\tau}_k = \frac{\mathbf{z}_k^H\mathbf{U}_c\hat{\mathbf{M}}_c^{-1}\mathbf{U}_c^H\mathbf{z}_k}{R} \tag{2.38}$$

Preuve 2.5.1 *$c.f.$ Annexe A.*

Pour une base du sous-espace fouillis donnée $\mathbf{U}_c$ on a alors :

$$\hat{\mathbf{\Sigma}}_c = \mathbf{U}_c\hat{\mathbf{M}}_c\mathbf{U}_c^H = \hat{\mathbf{V}}_c\hat{\mathbf{C}}_c\hat{\mathbf{V}}_c^H \tag{2.39}$$

où le paramètre $\hat{\mathbf{C}}_c$ est obtenu par la SVD de la matrice $\hat{\mathbf{\Sigma}}_c$.

Remarque : Généralisation aux M-estimateurs

Il est intéressant de replacer l'estimation de la matrice noyau et des textures dans le contexte plus large des M-estimateurs. En pratique la densité de probabilité de la texture f_τ est inconnue, c'est pourquoi nous considérons les textures comme des paramètres déterministes inconnus. Néanmoins si f_τ est connue, l'EMV prend la forme ($c.f.$ Section 1.3.4) :

$$\hat{\mathbf{M}}_c = \frac{1}{K}\sum_{k=1}^{K}\psi(\mathbf{z}_k^H\mathbf{U}_c\hat{\mathbf{M}}_c^{-1}\mathbf{U}_c^H\mathbf{z}_k^H)\mathbf{U}_c^H\mathbf{z}_k^H\mathbf{z}_k^H\mathbf{U}_c \tag{2.40}$$

avec $\psi(t) = -h_r'(t)/h_r(t)$ et $h_r(t) = \int_0^{+\infty}\exp(-t/\tau)\tau^R f_\tau(\tau)\mathrm{d}\tau$. Les conditions d'existence de cette solutions sont énoncées Théorème 1.3.1. Cet EMV peut être obtenu via les itérations suivantes :

$$\mathbf{M}_{(n+1)} = \frac{1}{K}\sum_{k=1}^{K}\psi(\mathbf{z}_k^H\mathbf{U}_c\mathbf{M}_{(n)}^{-1}\mathbf{U}_c^H\mathbf{z}_k^H)\mathbf{U}_c^H\mathbf{z}_k\mathbf{z}_k^H\mathbf{U}_c \tag{2.41}$$

Et l'EMV des textures est obtenu par :

$$\hat{\tau}_k = \psi(\mathbf{z}_k^H\mathbf{U}_c\hat{\mathbf{M}}_c^{-1}\mathbf{U}_c^H\mathbf{z}_k^H)^{-1} \tag{2.42}$$

Bien sûr, f_τ n'est pas connue en pratique. Dans ce cas, il est possible de recourir aux M-estimateurs [75, 56, 84] et d'utiliser une fonctionnelle ψ n'étant pas nécessairement reliée à la densité de probabilité f_τ ($c.f.$ section 1.3.5). Par exemple, l'estimateur du Point Fixe correspond au cas spécial $\psi(t) = R/t$. Le choix de cette fonctionnelle peut être lié à un a priori sur la vraie densité de probabilité, soit à un compromis performance-robustesse. L'étape d'estimation proposée se généralise donc simplement aux M-estimateurs en remplaçant les itérations Eq.(2.37) par Eq.(2.41) et la formule d'estimation des textures Eq.(2.38) par Eq.(2.42).

2.5.4 Étape 2 : Estimation du sous-espace fouillis pour textures et valeurs propres fixées

On suppose désormais les textures $\{\tau_k\}$ et valeurs propres $\mathbf{C}_c$ connues et fixées. La vraisemblance s'exprime :

$$f(\{\mathbf{z}_k\}|\mathbf{V}_c) = -\sum_{k=1}^{K} \mathbf{z}_k^H \mathbf{\Sigma}_k^{-1} \mathbf{z}_k + \gamma \tag{2.43}$$

où γ est une constante ne dépendant pas de $\{\mathbf{v}_r\}$, en utilisant (2.33) on obtient donc :

$$\log f = -\sum_{k=1}^{K} \frac{1}{\tau_k} \mathbf{z}_k^H \mathbf{V}_c \mathbf{C}_c^{-1} \mathbf{V}_c^H \mathbf{z}_k - \sum_{k=1}^{K} \mathbf{z}_k^H \mathbf{V}_c^{\perp} \mathbf{V}_c^{\perp H} \mathbf{z}_k + \gamma \, . \tag{2.44}$$

Comme effectué dans dans la section 2.2, afin de contraindre intrinsèquement l'orthogonalité entre les bases $\mathbf{V}_c$ et $\mathbf{V}_c^{\perp}$ dans le processus d'estimation [5], on utilise le changement de variable $\mathbf{V}_c^{\perp}\mathbf{V}_c^{\perp H} = \mathbf{I}_m - \mathbf{V}_c\mathbf{V}_c^H$, conduisant à :

$$f(\{\mathbf{z}_k\}|\mathbf{V}_c) = -\sum_{k=1}^{K} \frac{1}{\tau_k} \mathbf{z}_k^H \mathbf{V}_c \mathbf{C}_c^{-1} \mathbf{V}_c^H \mathbf{z}_k - \sum_{k=1}^{K} \mathbf{z}_k^H \mathbf{I}_m \mathbf{z}_k + \sum_{k=1}^{K} \mathbf{z}_k^H \mathbf{V}_c \mathbf{V}_c^H \mathbf{z}_k + \gamma \tag{2.45}$$

qui peut être réécrit de la façon suivante :

$$f(\{\mathbf{z}_k\}|\mathbf{V}_c) = \sum_{k=1}^{K} \sum_{r=1}^{R} \left(1 - \frac{1}{c_r \tau_k}\right) \mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k + \gamma \tag{2.46}$$

où γ absorbe la constante $\sum_{k=1}^{K} ||\mathbf{z}_k||^2$. On définit alors les matrices $\mathbf{M}_r$:

$$\mathbf{M}_r = \sum_{k=1}^{K} \left(1 - \frac{1}{c_r \tau_k}\right) \mathbf{z}_k \mathbf{z}_k^H \tag{2.47}$$

Maximiser la vraisemblance sous contraintes unitaire $\mathbf{V}_c^H \mathbf{V}_c = \mathbf{I}_R$ revient à résoudre le problème :

$$\begin{aligned} \max_{\{\mathbf{v}_r\}} \quad & f_0(\mathbf{V}) = f_0(\{\mathbf{v}_r\}) = \sum_{r=1}^{R} \mathbf{v}_r^H \mathbf{M}_r \mathbf{v}_r \\ \text{s.c.} \quad & \mathbf{v}_r^H \mathbf{v}_r = 1 \ , \ r \in [\![1, R]\!] \\ & \mathbf{v}_r^H \mathbf{v}_j = 0 \ , \ r, j \in [\![1, R]\!] \ , \ r \neq j \end{aligned} \tag{A}$$

Remarquons de plus que les facteurs α_r^k définissant les matrices $\mathbf{M}_r = \sum_{k=1}^{K} \alpha_r^k \mathbf{z}_k \mathbf{z}_k^H$ peuvent être réinterprétés grâce à l'hypothèse de fort rapport fouillis à bruit :

$$\alpha_r^k = \left(1 - \frac{1}{c_r \tau_k}\right) \simeq \left(\frac{\tau_k c_r}{1 + \tau_k c_r}\right) \tag{2.48}$$

Le problème est donc identique (même formulation et types de matrices) à celui présenté section 2.4.4. Nous tournons donc vers la même solution : l'appel de l'algorithme 15 de [73] permettant d'atteindre un maximum local de la vraisemblance sur l'espace des $\{\mathbf{v}_r\}$ satisfaisant les contraintes.

5. Rappelons qu'une relaxation de cette contrainte sera envisagée dans le chapitre suivant

Estimation de facteur d'échelle

Pour reconstruire la matrice de covariance totale, il convient d'estimer le rapport fouillis à bruit. Cette opération est strictement identique à celle proposée section 2.4.5.

2.6 Algorithmes Majorization-Minimization

Ces travaux ont été réalisés avec l'équipe du Professeur Daniel P. Palomar au Department of Electronic and Computer Engineering de Hong Kong University of Science and Technology (HKUST).

2.6.1 Motivations

Suite aux précédents développements, la motivation à envisager de nouveaux algorithmes pour atteindre l'EMV est double.

Dans un premier temps, on remarque que les deux algorithmes de type 2-Step ne permettent pas d'obtenir une solution tractable de l'EMV exact des paramètres. En effet, le premier algorithme dérivé (2SD) repose sur une heuristique dont on observe simplement qu'elle augmente la vraisemblance. Le second algorithme développé (2SR) correspond bien à une maximisation alternée de la vraisemblance mais repose sur une hypothèse supplémentaire de fort rapport fouillis à bruit. Ces heuristiques, ou relaxations sur la vraisemblance, conduisent donc seulement à des EMV approchés. Bien que ces approximations se soient avérées satisfaisantes dans un premier temps ($c.f.$ Section 2.7.3) nous désirions atteindre un EMV exact.

De plus, pour résoudre le problème (A) posé en deuxième étape des précédents algorithmes, nous avons initialement recouru à des méthodes d'optimisation sur la variété de Stiefel [73]. Ces algorithmes peuvent s'avérer difficiles à implémenter et lourds en termes de calculs, c'est pourquoi envisager d'autres options peut s'avérer intéressant d'un point de vue pratique.

Le cadre théorique des algorithmes Majorization-Minimization (MM) [69] par blocs s'avère être un outil de résolution approprié, permettant de proposer une solution à ces deux problématiques.

2.6.2 Principe général des algorithmes MM par blocs

Conceptuellement, les algorithmes MM procèdent de la même manière que les algorithmes 2-Step présentés dans les sections précédentes : les ensembles de paramètres ("blocs") sont mis à jours itérativement en fixant les autres blocs. Cependant, chaque étape ne va pas chercher à maximiser directement la vraisemblance, mais une une borne supérieure de celle-ci (la fonction de substitution). Formellement exprimé, si l'on considère le problème d'optimisation suivant :

$$\begin{aligned} \min_{\mathbf{x}} \quad & f\left(\mathbf{x}\right) \\ \text{s.c.} \quad & \mathbf{x} \in \mathcal{X}, \end{aligned} \tag{2.49}$$

où les variables $\mathbf{x}$ peuvent êtres partitionnées en m blocs comme $\mathbf{x} = \left(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)}\right)$, avec chaque bloc $\mathbf{x}^{(i)} \in \mathcal{X}_i$ de dimension $n_{(i)}$, vérifiant $\mathcal{X} = \prod_{i=1}^{m} \mathcal{X}_i$. À l'itération $(t+1)$, le i^{ieme} bloc $\mathbf{x}^{(i)}$ est mis à jour en résolvant le problème de minimisation de la fonction de substitution g_i, du bloc i au point $\mathbf{x}_t$, notée $g_i(\mathbf{x}^{(i)}|\mathbf{x}_t)$:

$$\begin{aligned} \min_{\mathbf{x}^{(i)}} \quad & g_i\left(\mathbf{x}^{(i)}|\mathbf{x}_t\right) \\ \text{s.c.} \quad & \mathbf{x}^{(i)} \in \mathcal{X}_i \end{aligned} \tag{2.50}$$

avec $i = (t \bmod m) + 1$ (où mod est le modulo), les blocs sont donc mis à jour de façon cyclique, et où la fonction de substitution $g_i\left(\mathbf{x}^{(i)}|\mathbf{x}_t\right)$ vérifie :

$$\begin{aligned} f\left(\mathbf{x}_t\right) &= g_i\left(\mathbf{x}_t^{(i)}|\mathbf{x}_t\right), \\ f\left(\mathbf{x}_t^{(1)}, \ldots, \mathbf{x}^{(i)}, \ldots, \mathbf{x}_t^{(m)}\right) &\leq g_i\left(\mathbf{x}^{(i)}|\mathbf{x}_t\right) \ \forall \mathbf{x}^{(i)} \in \mathcal{X}_i, \\ f'\left(\mathbf{x}_t; \mathbf{d}_i^0\right) &= g_i'\left(\mathbf{x}_t^{(i)}; \mathbf{d}_i|\mathbf{x}_t\right) \\ & \quad \forall \mathbf{x}_t^{(i)} + \mathbf{d}_i \in \mathcal{X}_i, \\ & \quad \mathbf{d}_i^0 \triangleq (\mathbf{0}; \ldots; \mathbf{d}_i; \ldots; \mathbf{0}), \end{aligned}$$

où $f'(\mathbf{x}; \mathbf{d})$ est la dérivée au point $\mathbf{x}$ dans la direction $\mathbf{d}$. Sous certaines conditions (Quasi-Convexité des fonction de substitution et unicité de leur minimum), ces itérations convergent vers un maximum local de la fonction considérée [98].

Afin de ne pas alourdir ce chapitre, la dérivation complète des algorithmes liés à notre problématique est reportée en annexe C, contenant l'article issu de ces travaux. Nous présenterons donc ici simplement leur principe général et leurs propriétés.

2.6.3 Algorithme MLE-MM1 - "direct block-MM"

Cet algorithme, résumé dans l'encadré 3 "MLE-M1" vise à atteindre un maximum local de la vraisemblance. Les paramètres considérés dans les blocs sont les textures $\{\tau_k\}$ et la matrice $\mathbf{W} \in M \times R$ paramétrant la matrice de covariance du fouillis rang faible comme :

$$\boldsymbol{\Sigma}_c = \mathbf{W}\mathbf{W}^H \tag{2.51}$$

Mise à jour de $\{\tau_k\}$ pour $\mathbf{W}$ fixe

Soit le paramètre $\mathbf{W}$ fixé. Définissons la SVD $\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$, avec $\boldsymbol{\Lambda} = \mathrm{diag}\left(\lambda_1, \ldots, \lambda_R, 0, \ldots, 0\right)$ et $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_M]$. Maximiser la vraisemblance selon τ_k revient à chercher le minimum de la fonction :

$$L\left(\tau_k|\mathbf{W}\right) = \sum_{m=1}^{R} \log\left(\tau_k\lambda_m + 1\right) + \sum_{m=1}^{R} s_{km}\left(\tau_k\lambda_m + 1\right)^{-1} + \text{const.}, \tag{2.52}$$

avec $s_{km} = \left\|\mathbf{z}_k^H\mathbf{u}_m\right\|^2$. Cette fonction $L\left(\tau_k|\mathbf{W}\right)$ n'est pas convexe et n'admet pas de minimum tractable. Une fonction de substitution est donc dérivée dans la proposition 1 de l'annexe C. Cette fonction $L\left(\tau_k|\mathbf{W}\right)$ peut être majorée au point τ_k^t par la fonction quasi-convexe $L\left(\tau_k|\tau_k^t, \mathbf{W}\right)$:

$$L\left(\tau_k|\tau_k^t, \mathbf{W}\right) = -\beta_k \log\tau_k + \left(\sum_{m=1}^{R}\alpha_{km}\right)\log\left(\frac{\left(\sum_{m=1}^{R}\frac{\alpha_{km}\lambda_m}{1+\lambda_m\tau_k^t}\right)}{\sum_{m=1}^{R}\alpha_{km}}\tau_k + \frac{\sum_{m=1}^{R}\frac{\alpha_{km}}{1+\lambda_m\tau_k^t}}{\sum_{m=1}^{R}\alpha_{km}}\right) + \text{const.}, \tag{2.53}$$

avec

$$\alpha_{km} = s_{km}\frac{\tau_k^t\lambda_m}{\tau_k^t\lambda_m + 1} + 1 \quad \text{et} \quad \beta_k = \sum_{m=1}^{R} s_{km}\frac{\tau_k^t\lambda_m}{\tau_k^t\lambda_m + 1}. \tag{2.54}$$

et vérifie bien l'égalité en $\tau_k = \tau_k^t$.

L'unique minimum de $L\left(\tau_k|\tau_k^t, \mathbf{W}\right)$, définissant la mise à jour des variables $\{\tau_k\}$ est obtenu par (proposition 2 annexe C) :

$$\tau_k^{t+1} = \frac{1}{R} \cdot \frac{\left(\sum_{m=1}^{R} s_{km} \frac{\tau_k^t \lambda_m}{\tau_k^t \lambda_m + 1}\right) \cdot \left(\sum_{m=1}^{R} \frac{\alpha_{km}}{1+\lambda_m \tau_k^t}\right)}{\sum_{m=1}^{R} \frac{\alpha_{km}\lambda_m}{1+\lambda_m \tau_k^t}}. \tag{2.55}$$

Mise à jour de $\mathbf{W}$ pour $\{\tau_k\}$ fixe

Pour variables $\{\tau_k\}$ fixées, maximiser la vraisemblance selon $\mathbf{W}$ revient à minimiser la fonction :

$$L\left(\mathbf{W}|\tau_k\right) = \sum_{k=1}^{K} \log|\mathbf{\Sigma}_k| + \sum_{k=1}^{K} \mathbf{z}_k^H \mathbf{\Sigma}_k^{-1} \mathbf{z}_k, \tag{2.56}$$

avec $\mathbf{\Sigma}_k = \tau_k \mathbf{W}\mathbf{W}^H + \mathbf{I}$. Comme pour l'étape précédente, il n'est pas possible de trouver un minimum de manière tractable. $L\left(\mathbf{W}|\tau_k\right)$ peut être majorée par la fonction quadratique convexe (proposition 3 Annexe C) :

$$L\left(\mathbf{W}|\tau_k, \mathbf{W}^t\right) = \mathrm{Tr}\left(\mathbf{W}\mathbf{H}\mathbf{W}^H\right) - \mathrm{Tr}\left(\mathbf{L}\mathbf{W}^H\right) - \mathrm{Tr}\left(\mathbf{L}^H\mathbf{W}\right), \tag{2.57}$$

où l'égalité est bien vérifiée en $\mathbf{W} = \mathbf{W}^t$, et où

$$\mathbf{H} = \sum_{k=1}^{K} \left(\left(\tilde{\mathbf{\Sigma}}^t + \tau_k^{-1}\mathbf{I}\right)^{-1} + \left(\tau_k^{-1}\mathbf{I} + \tilde{\mathbf{\Sigma}}^t\right)^{-1} \left(\mathbf{W}^{\mathbf{t}}\right)^H \mathbf{z}_k \mathbf{z}_k^H \mathbf{W}^t \left(\tau_k^{-1}\mathbf{I} + \tilde{\mathbf{\Sigma}}^t\right)^{-1} \right), \tag{2.58}$$

avec $\tilde{\mathbf{\Sigma}}^t = (\mathbf{W}^t)^H \mathbf{W}^t$ et

$$\mathbf{L} = \sum_{k=1}^{K} \mathbf{z}_k \mathbf{z}_k^H \mathbf{W}^t \left(\tau_k^{-1}\mathbf{I} + \tilde{\mathbf{\Sigma}}^t\right)^{-1}. \tag{2.59}$$

Cette fonction de substitution $L\left(\mathbf{W}|\tau_k, \mathbf{W}^t\right)$ admet pour unique minimum

$$\mathbf{W}^{t+1} = \mathbf{L}\mathbf{H}^{-1}.$$

Propriétés de l'algorithme MLE-MM1

Cet algorithme satisfait les conditions du théorème 2 (a) de [98] (quasi-convexité des fonctions de substitutions et unicité des minimums). Il converge donc vers un maximum local de la vraisemblance.

2.6.4 Algorithme MLE-MM2 - "Eigenspace block-MM"

Cet algorithme résumé dans l'encadré 4 "MLE-MM2" vise à atteindre un maximum local de la vraisemblance. Les paramètres considérés dans les blocs sont les textures $\{\tau_k\}$, les valeurs propres de la matrice de covariance du fouillis $\{c_r\}$ et ses vecteurs propres $\{\mathbf{v}_r\}$.

Algorithm 3 MLE-MM1

1: Calculer la SCM $\hat{\mathbf{\Sigma}}_{\text{SCM}} = \frac{1}{K}\sum_{k=1}^{K} \mathbf{z}_k \mathbf{z}_k^H$.
2: Calculer la SVD $\hat{\mathbf{\Sigma}}_{\text{SCM}} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, avec $\mathbf{\Lambda} = \text{diag}\left(\lambda_1, \ldots, \lambda_M\right)$ et $\lambda_1 \geq \cdots \geq \lambda_M$.
3: Initialiser $\mathbf{W}$ comme $\mathbf{W} = \mathbf{U}\text{diag}\left(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_R}, 0, \ldots, 0\right)$, et $\{\tau_k\}_{k=1}^{K}$ aléatoirement (réels positifs).
4: **repeat**
5: Décomposer $\mathbf{\Sigma} = \mathbf{W}\mathbf{W}^H$ via SVD en $\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, avec $\mathbf{\Lambda} = \text{diag}\left(\lambda_1, \ldots, \lambda_R, 0, \ldots, 0\right)$ et $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_M]$.
6: Calculer $s_{km} = \left\|\mathbf{z}_k^H \mathbf{u}_m\right\|^2$ et $\alpha_{km} = s_{km}\frac{\tau_k^t \lambda_m}{\tau_k^t \lambda_m + 1} + 1$
7: Mettre à jour $\{\tau_k\}_{k=1}^{K}$ avec

$$\tau_k^{t+1} = \frac{1}{R} \cdot \frac{\left(\sum_{m=1}^{R} s_{km}\frac{\tau_k^t \lambda_m}{\tau_k^t \lambda_m + 1}\right) \cdot \left(\sum_{m=1}^{R} \frac{\alpha_{km}}{1 + \lambda_m \tau_k^t}\right)}{\sum_{m=1}^{R} \frac{\alpha_{km}\lambda_m}{1 + \lambda_m \tau_k^t}}. \tag{2.60}$$

8: Calculer

$$\mathbf{H} = \sum_{k=1}^{K} \left(\left(\tilde{\mathbf{\Sigma}} + \tau_k^{-1}\mathbf{I}\right)^{-1} + \left(\tau_k^{-1}\mathbf{I} + \tilde{\mathbf{\Sigma}}\right)^{-1} (\mathbf{W})^H \mathbf{z}_k \mathbf{z}_k^H \mathbf{W} \left(\tau_k^{-1}\mathbf{I} + \tilde{\mathbf{\Sigma}}\right)^{-1} \right) \tag{2.61}$$

9: Puis

$$\mathbf{L} = \sum_{k=1}^{K} \mathbf{z}_k \mathbf{z}_k^H \mathbf{W} \left(\tau_k^{-1}\mathbf{I} + \tilde{\mathbf{\Sigma}}\right)^{-1} \tag{2.62}$$

10: Mettre à jour

$$\mathbf{W} = \mathbf{L}\mathbf{H}^{-1} \tag{2.63}$$

11: **until** Convergence
12: $\hat{\mathbf{\Sigma}}_c = \mathbf{W}\mathbf{W}^H$
13: $\hat{\mathbf{\Sigma}} = \hat{\mathbb{E}}(\{\tau_k\})\mathbf{\Sigma}_c + \mathbf{I}_M$

Mise à jour de $\{\tau_k\}$ pour $\{c_r\}$ et $\{\mathbf{v}_r\}$ fixés

Pour $\{c_r\}$ et $\{\mathbf{v}_r\}$ fixés, maximiser la vraisemblance selon τ_k revient à chercher le minimum de la fonction (2.52). On applique donc la même méthode que précédemment et obtient les mises à jour suivantes :

$$\tau_k^{t+1} = \frac{1}{R} \cdot \frac{\left(\sum_{r=1}^{R} s_{kr}\frac{\tau_k^t c_r}{\tau_k^t c_r + 1}\right) \cdot \left(\sum_{r=1}^{R} \frac{\alpha_{kr}}{1 + c_r \tau_k^t}\right)}{\sum_{r=1}^{R} \frac{\alpha_{kr} c_r}{1 + c_r \tau_k^t}}, \tag{2.64}$$

avec $\alpha_{kr} = s_{kr}\frac{\tau_k^t c_r}{\tau_k^t c_r + 1} + 1$.

Mise à jour de $\{c_r\}$ pour $\{\tau_k\}$ et $\{\mathbf{v}_r\}$ fixés

Pour $\{\tau_k\}$ et $\{\mathbf{v}_r\}$ fixés, maximiser la vraisemblance selon c_r revient à chercher le minimum de la fonction :

$$L(c_r|\{\tau_k\},\{\mathbf{v}_r\}) = \sum_{k=1}^{K} \log\left(1+\tau_k c_r\right) + \sum_{k=1}^{K} \frac{s_{kr}}{1+\tau_k c_r} \tag{2.65}$$

avec $s_{kr} = \left\|\mathbf{z}_k^H \mathbf{v}_r\right\|^2$. On remarque que les c_r de (2.65) jouent un rôle symétrique aux τ_k de l'équation (2.52). Les mises à jour de $\{c_r\}$ se dérivent donc identiquement en appliquant les propositions 1 et 2 de l'annexe C :

$$c_r^{t+1} = \frac{1}{K} \cdot \frac{\left(\sum_{k=1}^{K} s_{kr} \frac{\tau_k c_r^t}{\tau_k c_r^t+1}\right) \cdot \left(\sum_{k=1}^{K} \frac{\alpha_{kr}}{1+c_r^t\tau_k}\right)}{\sum_{k=1}^{K} \frac{\alpha_{kr}\tau_k}{1+c_r^t\tau_k}}, \tag{2.66}$$

avec $\alpha_{kr} = s_{kr}\frac{\tau_k c_r^t}{\tau_k c_r^t+1} + 1$.

Mise à jour de $\{\mathbf{v}_r\}$ pour $\{c_r\}$ et $\{\tau_k\}$ fixés

Pour $\{\tau_k\}$ et $\{c_r\}$ fixés, minimiser la vraisemblance selon $\{\mathbf{v}_r\}$ sous contraintes unitaires revient à résoudre :

$$\begin{aligned} \max_{\{\mathbf{v}_r\}} \quad & \sum_{r=1}^{R} \mathbf{v}_r^H \mathbf{M}_r \mathbf{v}_r \\ \text{s.c.} \quad & \mathbf{v}_r \text{ orthonormés,} \end{aligned} \tag{2.67}$$

avec $\mathbf{M}_r = \sum_{k=1}^{K} \frac{\tau_k c_r}{\tau_k c_r+1} \mathbf{z}_k \mathbf{z}_k^H$. Comme évoqué dans les précédents algorithmes, ce problème n'admet pas de solution tractable. La fonction à maximiser admet la fonction de substitution suivante (proposition 5 annexe C) :

$$\begin{aligned} & L\left(\{\mathbf{v}_r\} \,|\, \{\tau_k\}, \{c_r\}, \{\mathbf{v}_r\}^t\right) \\ & = \sum_{r=1}^{R} \left[\left(\mathbf{v}_r^t\right)^H \mathbf{M}_r \mathbf{v}_r + \mathbf{v}_r^H \mathbf{M}_r \mathbf{v}_r^t\right] + \text{const.} \end{aligned} \tag{2.68}$$

Maximiser $L\left(\{\mathbf{v}_r\} \,|\, \{\tau_k\}, \{c_r\}, \{\mathbf{v}_r\}^t\right)$ sous contraintes unitaires revient à considérer le problème

$$\begin{aligned} \min_{\mathbf{V}} \quad & \|\mathbf{A}-\mathbf{V}\|_F^2 \\ \text{s.c.} \quad & \mathbf{V}^H\mathbf{V} = \mathbf{I}, \end{aligned} \tag{2.69}$$

avec

$$\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_R] \quad \text{et} \quad \mathbf{A} = \left[\mathbf{M}_1\mathbf{v}_1^t; \ldots; \mathbf{M}_R\mathbf{v}_R^t\right], \tag{2.70}$$

qui est un problème de Procuste [102]. Soit la SVD fine de $\mathbf{A}$ définie par $\mathbf{A} = \mathbf{V}_A\mathbf{D}_A\mathbf{U}_A^H$, la solution définissant la mise à jour de $\{\mathbf{v}_r\}$ est donnée par :

$$\mathbf{V}^{t+1} = \mathbf{V}_A\mathbf{U}_A^H. \tag{2.71}$$

Remarque : Répéter cette mise à jour (2.71) permet d'atteindre directement un maximum local de (2.67) satisfaisant les contraintes unitaires. Ces opérations offrent donc un autre algorithme pour résoudre la seconde étape des algorithmes "2-Steps" précédemment développés, sans recourir aux complexes méthodes d'optimisations sur la variété de Stiefel.

Propriétés de l'algorithme MLE-MM2

Cet algorithme produit une séquence de points augmentant la vraisemblance de manière monotone, cependant, il n'a à notre connaissance pas de garantie théorique de convergence car les contraintes unitaires sur $\{\mathbf{v}_r\}$ ne définissent pas un ensemble convexe et ne permettent pas d'appliquer les résultats de [98] comme précédemment. En pratique, on vérifie bien que cet algorithme converge. L'intérêt de MLE-MM2 réside dans ses performances, ainsi que dans son temps de calcul, plus rapide que celui de MM1, comme illustré dans la partie simulations de l'annexe C.

Algorithm 4 MLE-MM2

1: Calculer la SCM $\hat{\mathbf{\Sigma}}_{\text{SCM}} = \frac{1}{K}\sum_{k=1}^{K} \mathbf{z}_k \mathbf{z}_k^H$.
2: Initialiser $\{c_r\}_{r=1}^{R}$ avec les R valeurs propres les plus fortes de $\hat{\mathbf{\Sigma}}^{\text{SCM}}$, $\{\mathbf{v}_r\}_{r=1}^{R}$ avec les vecteurs propres correspondants et $\{\tau_k\}_{k=1}^{K}$ (réels positifs).
3: **repeat**
4: Calculer $s_{kr} = \left\|\mathbf{z}_k^H \mathbf{v}_r\right\|^2$ et $\alpha_{kr} = s_{kr}\frac{\tau_k c_r^t}{\tau_k c_r^t + 1} + 1$
5: Mettre à jour τ_k avec

$$\tau_k^{t+1} = \frac{1}{R} \cdot \frac{\left(\sum_{r=1}^{R} s_{kr}\frac{\tau_k^t c_r}{\tau_k^t c_r + 1}\right) \cdot \left(\sum_{r=1}^{R} \frac{\alpha_{kr}}{1 + c_r \tau_k^t}\right)}{\sum_{r=1}^{R} \frac{\alpha_{kr} c_r}{1 + c_r \tau_k^t}} \tag{2.72}$$

6: Mettre à jour c_r avec

$$c_r^{t+1} = \frac{1}{K} \cdot \frac{\left(\sum_{k=1}^{K} s_{kr}\frac{\tau_k c_r^t}{\tau_k c_r^t + 1}\right) \cdot \left(\sum_{k=1}^{K} \frac{\alpha_{kr}}{1 + c_r^t \tau_k}\right)}{\sum_{k=1}^{K} \frac{\alpha_{kr} \tau_k}{1 + c_r^t \tau_k}} \tag{2.73}$$

7: Calculer les matrices $\mathbf{M}_r = \sum_{k=1}^{K} \frac{\tau_k c_r}{\tau_k c_r + 1} \mathbf{z}_k \mathbf{z}_k^H$.
8: **repeat** (*Boucle optionnelle*)
9: Soit $\mathbf{V} = [\mathbf{v}_1 \ ; \ \ldots \ ; \ \mathbf{v}_r]$, calculer $\mathbf{A}$ avec

$$\mathbf{A} = [\mathbf{M}_1 \mathbf{v}_1 ; \ldots ; \mathbf{M}_R \mathbf{v}_R] \tag{2.74}$$

10: Décomposer $\mathbf{A}$ en $\mathbf{A} = \mathbf{V}_A \mathbf{D}_A \mathbf{U}_A^H$ (SVD fine).
11: Mettre à jour $\mathbf{V}$ avec $\mathbf{V} = \mathbf{V}_A \mathbf{U}_A^H$.
12: **until** Convergence
13: **until** Convergence
14: $\hat{\mathbf{\Sigma}}_c = \mathbf{V}\text{diag}(\{c_r\})\mathbf{V}^H$
15: $\hat{\mathbf{\Sigma}} = \hat{\mathbb{E}}(\{\tau_k\})\hat{\mathbf{\Sigma}}_c + \mathbf{I}_M$

2.7 Simulations

Cette section vise à illustrer les performances des algorithmes développés pour l'estimation de la matrice de covariance dans un contexte CG rang faible plus bruit blanc gaussien. Le critère considéré est la précision d'estimation de cette matrice de covariance, mesurée en termes d'erreur quadratique moyenne normalisée (notée NMSE pour "Normalized Mean Square Error"), définie par :

$$\text{NMSE}(\hat{\mathbf{\Sigma}}) = \mathbb{E}\left(\frac{||\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}||^2}{||\mathbf{\Sigma}||^2}\right) \tag{2.75}$$

pour un estimateur $\hat{\mathbf{\Sigma}}$ donné. Afin de se soustraire à l'ambiguïté d'échelle de certains estimateurs, nous calculerons le NMSE sur des estimateurs normalisés selon $\text{Tr}\,(\mathbf{\Sigma}) = M$.

2.7.1 Paramètres

Dans ces simulations, les échantillons $\mathbf{z}_k$ sont générés selon la vraisemblance décrite en section 2.2 : $\mathbf{z}_k = \mathbf{c}_k + \mathbf{n}_k$. Le bruit blanc gaussien suit $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$ avec $\sigma^2 = 1$. Le fouillis rang faible est distribué selon $(\mathbf{c}_k|\tau_k) \sim \mathcal{CN}(\mathbf{0}, \tau_k\mathbf{\Sigma})$, avec une texture aléatoire τ_k, $i.i.d.$ pour chaque échantillon. La densité de probabilité des textures est une distribution Gamma, noté $\tau \sim \Gamma(\nu, 1/\nu)$, conduisant à un fouillis K-distribué (cf. section 1.2.4). Le paramètre de forme est ν et le paramètre d'échelle $1/\nu$, la distribution Gamma vérifie donc $\mathbb{E}(\tau) = 1$. La matrice de covariance du fouillis $\mathbf{\Sigma}_c$ est construite à l'aide des R valeurs et vecteurs propres dominants d'une matrice Toeplitz, définie par $[\mathbf{\Sigma}_T]_{i,j} = \rho^{|i-j|}$ avec le paramètre de corrélation $\rho \in [0, 1]$. Cette matrice est ensuite mise à l'échelle pour fixer le rapport fouillis à bruit, $\text{CNR} = \mathbb{E}(\tau)\text{Tr}(\mathbf{\Sigma})/(R\sigma^2)$, à une valeur voulue.

2.7.2 Estimateurs considérés

Nous étudierons les estimateurs de la matrice de covariance suivants :

- SCM : la Sample Covariance Matrix, décrite section 1.3.2.
- FPE : l'estimateur du point fixe, décrit section 1.3.5. Pour les cas ou $K < M$, nous utiliserons le point fixe régularisé, décrit section 1.3.6. Comme il n'existe pas de règle de choix adaptatif "optimal" du paramètre de régularisation β pour le contexte que nous considérons, nous utiliserons la valeur minimale requise pour son existence. L'estimateur considéré se définit donc comme le SFPE (pour Shrinkage FPE) utilisant le parmètre $\beta_{min} = \max(0, 1-K/M+\epsilon)$ (on fixe $\epsilon = 0.02$).
- RC-ML : l'estimateur décrit section 1.4.3.
- MLE-2SD : le maximum de vraisemblance approché à l'aide de l'algorithme 1.
- MLE-2SR : le maximum de vraisemblance approché à l'aide de l'algorithme 2.
- MLE-MM1 : le maximum de vraisemblance calculé à l'aide de l'algorithme 3.
- MLE-MM2 : le maximum de vraisemblance calculé à l'aide de l'algorithme 4.

2.7.3 Résultats

Tout d'abord, la figure 2.1 illustre la convergence des algorithmes développés dans ce chapitre. Elle présente des réalisations typiques d'évolution de la vraisemblance négative (valeur opposée de la vraisemblance) en fonction des itérations. On observe que tous les algorithmes conduisent à une diminution

de la vraisemblance négative, donc augmentent bien la vraisemblance comme attendu. Les algorithmes EMV-MM1 et EMV-MM2 convergent vers des points conduisant à la plus grande vraisemblance. EMV-MM1 étant garanti théoriquement de converger vers un maximum local, on peut donc supposer que EMV-MM2 fait de même. Néanmoins, bien que ces deux algorithmes atteignent une valeur de l'objectif quasi identique, ils ne convergent pas nécessairement vers le même point, comme nous le verrons après. Notons aussi que EMV-2SD augmente la vraisemblance bien qu'il repose sur une heuristique de régularisation ne garantissant pas théoriquement cette propriété. A fort rapport fouillis à bruit, EMV-2SR apparaît plus performant (en termes d'optimisation) que 2SD, ce qui justifie l'approximation utilisée pour son développement. Les deux algorithmes "approchés" atteignent cependant des point sous-optimaux en comparaison de ceux obtenus avec les algorithmes MM.

Les figures 2.2 et 2.3 présentent l'évolution du NMSE en fonction du nombre d'échantillons K pour diverses configurations de fouillis. Chacune de ses figures présente les résultats pour un rapport fouillis à bruit (noté CNR pour "Clutter to Noise Ratio") faible (3dB), moyen(10dB) et fort (30dB). En figure 2.2, le paramètre ν du fouillis K-distribué est fixé à $\nu = 1$, conduisant à un fouillis modérément hétérogène. En figure 2.3, ce paramètre est fixé à $\nu = 0.1$, conduisant à un fouillis fortement hétérogène.

Décrivons ces résultats en balayant les estimateurs une par uns :

- La SCM est l'EMV sous hypothèse de bruit gaussien mais est connue pour ne pas être robuste aux distributions à queues lourdes. De plus cet estimateur ne prend pas en compte l'a priori sur la structure de la matrice de covariance. En conséquence, ce "benchmark" présente, comme on pouvait s'y attendre, les moins bonnes performances dans tous les cas de figures considérés.
- RCML correspond à la SCM ayant ses valeurs propres seuillées pour intégrer l'a priori sur la structure rang faible de la matrice de covariance. Cette opération offre de meilleures performances en termes de NMSE. Cependant, ce gain est modéré et n'est réellement observable qu'en présence de faible rapport fouillis à bruit. Ceci suggère que l'ajout de structure à la SCM ne compense pas complètement le fait d'être peu robuste aux bruits hétérogènes.
- Pour les contextes de fouillis modérément hétérogènes (figure 2.2), le FPE présente des performances comparables à la SCM et à RCML. Cependant, le FPE surpasse ces derniers lorsque la puissance ou l'hétérogénéité du fouillis augmentent (figure 2.3). Dans ces conditions et pour un nombre d'échantillons suffisant, le FPE est parfois proche des meilleures performances observables. Notons aussi que pour $K < M$, l'estimateur considéré est SFPE avec le paramètre de régularisation β minimum. Ce choix de paramètre n'étant pas nécessairement optimal il est normal d'observer un décrochement de performances entre les régimes $K < M$ et $K > M$.
- L'EMV présente généralement les meilleures performances en termes de précision d'estimation. Néanmoins ces performances dépendent fortement de l'algorithme utilisé :
  - l'heuristique EMV-2SD présente des performances similaires à RCML, excepté dans les contextes de faible rapport fouillis à bruit, ou 2SD apporte un gain quand peu d'échantillons sont disponibles. En comparaison avec les méthodes suivantes, on peut conclure que cet algorithme n'est pas le plus adapté au problème de reconstruction de la matrice de covariance totale.
  - Les performances de l'algorithme EMV-2SR varient selon la puissance du fouillis. Ceci est logique car son développement repose sur l'hypothèse de fort rapport fouillis à bruit, contexte où cet algorithme atteint les meilleures performances.

- Hormis les contextes de très fort rapport fouillis à bruit, l'algorithme EMV-MM2 semble généralement le plus performant en termes de précision sur la reconstruction de la matrice de covariance totale.
- L'algorithme EMV-MM1 présente des performances toujours en dessous de EMV-MM2 pour le critère considéré. Ces deux algorithmes sont initialisés au même point et ne convergent pas vers la même solution, ce qui vient confirmer la présence de minimum locaux dans la vraisemblance.

Pour résumer les algorithmes EMV-MM2 et EMV-2SR (pour les contextes de fort rapport fouillis à bruit) semblent les plus prometteurs pour estimer la matrice de covariance totale dans le contexte considéré. Les estimateurs associés à ces algorithmes présentent en effet les meilleures performances comparées à l'état de l'art.

Notons toutefois que le NMSE est un critère qui n'est généralement pas lié aux performances d'une application donnée. Il a donc ici un caractère illustratif et les performances des estimateurs considérés seront aussi illustrées sur une application de détection adaptative au cours du chapitre 4.

2.8 Synthèse du Chapitre 2

Dans ce chapitre, nous avons présenté le modèle considéré au cours de cette thèse. Motivé principalement par des applications de type radar, nous considérons les vecteurs aléatoires comme étant la somme d'un bruit CG de rang faible (le fouillis) et d'un bruit blanc gaussien (le bruit thermique). Nous avons ensuite détaillé la vraisemblance associé à ce modèle et exprimé l'EMV de la matrice de covariance du fouillis. Cet EMV n'étant pas directement tractable, c'est pourquoi nous avons développé différents algorithmes afin de l'atteindre, ou de s'en approcher.

Les premiers algorithmes présentés visent à maximiser la vraisemblance de manière alternée (2-Step) entre les paramètres textures et valeurs propres $\{\{\tau_k\}, \{c_r\}\}$, et les paramètres vecteurs propres $\{\mathbf{v}_r\}$. Néanmoins, les textures τ_k et valeurs propres c_r n'ont pas d'EMV tractable, c'est pourquoi nous avons proposé deux relaxations du modèle pour les estimer. La première relaxation introduit des paramètres intermédiaires $d_r^k = \tau_k c_r$ dont l'EMV est tractable, puis propose une régularisation pour recouvrer les paramètres d'origine. La seconde relaxation suppose un fort rapport fouillis à bruit. Sous cette condition, l'EMV des paramètres recherchés peut être obtenu à l'aide de l'estimateur du Point Fixe.
Afin de résoudre l'étape de maximisation de la vraisemblance selon les paramètres $\{\mathbf{v}_r\}$, nous avons eu recours à une descente de gradient sur la variété de Stiefel [73], permettant de maximiser la vraisemblance sous contrainte unitaire.

Les seconds algorithmes sont développés via la méthodologie Majorization-Minimization et ne nécessitent pas de relaxations afin de converger vers un EMV local.

Au cours de simulations de validation, nous avons montré que les estimateurs ainsi développés présentaient des performances (en termes de NMSE sur la matrice de covariance totale) supérieures à l'état de l'art. Néanmoins ces performances dépendent fortement de l'algorithme utilisé. De manière générale, l'algorithme EMV-MM2 semble présenter une bonne solution, adaptée à la majeure partie des contextes. Pour les contextes de fort rapport fouillis à bruit (présents dans de nombreuses applications) l'algorithme EMV-2SR semble présenter la meilleur alternative.

Au cours de ce chapitre, nous avons vu qu'un paramètre récurrent apparaît : le sous-espace fouillis $\mathbf{\Pi}_c$ (ou sa base orthonormée $\{\mathbf{v}_r\}$). Il correspond au sous-espace vectoriel engendré par les R vecteurs propres dominants la matrice de covariance totale. Beaucoup traitements (MUSIC, annulation d'interférence...) reposent uniquement sur l'estimation de ce sous-espace, et non de la matrice de covariance totale. C'est pourquoi il semble intéressant de se focaliser sur la problématique d'estimation de ce paramètre. Dans le prochain chapitre, nous montrerons qu'il est possible de construire des estimateurs directs de ce paramètre en relâchant certaines contraintes sur la vraisemblance. Ceci conduit à de nouveaux algorithmes, plus simples que ceux précédemment développés.

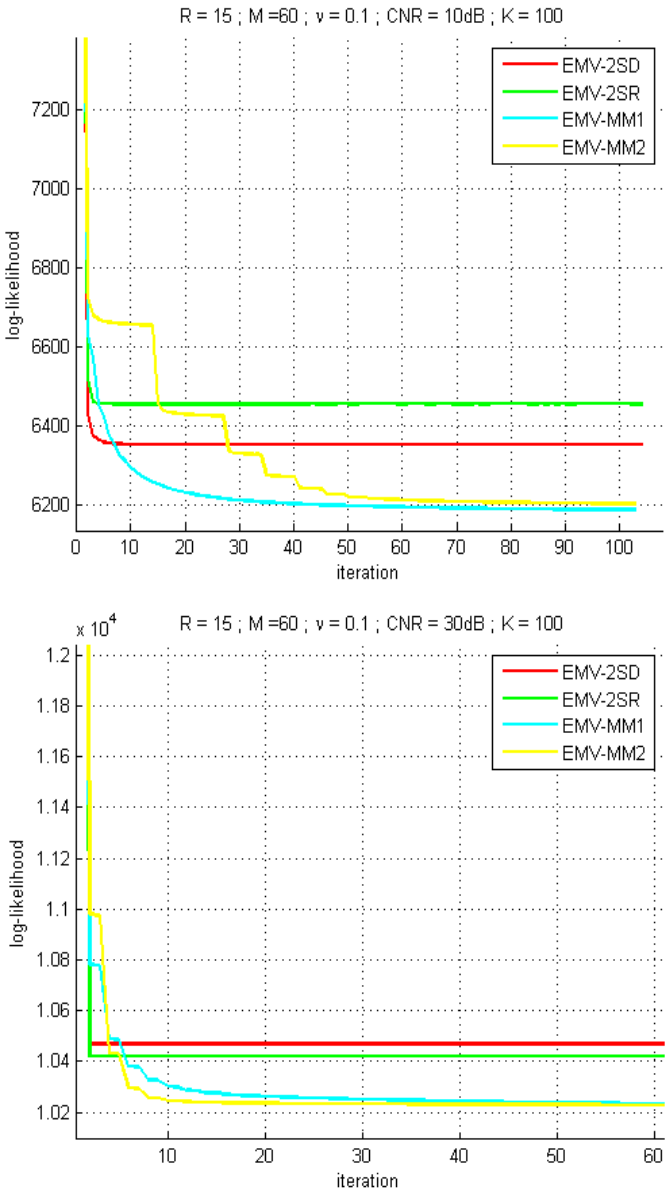

FIGURE 2.1 – Évolution de la vraisemblance négative (valeur opposée de la vraisemblance) en fonction des itérations pour les différents algorithmes "EMV". $M = 60$, $R = 15$, $K = 100$, $\nu = 1$. Haut : CNR$= 10$, bas : CNR$= 30$dB.

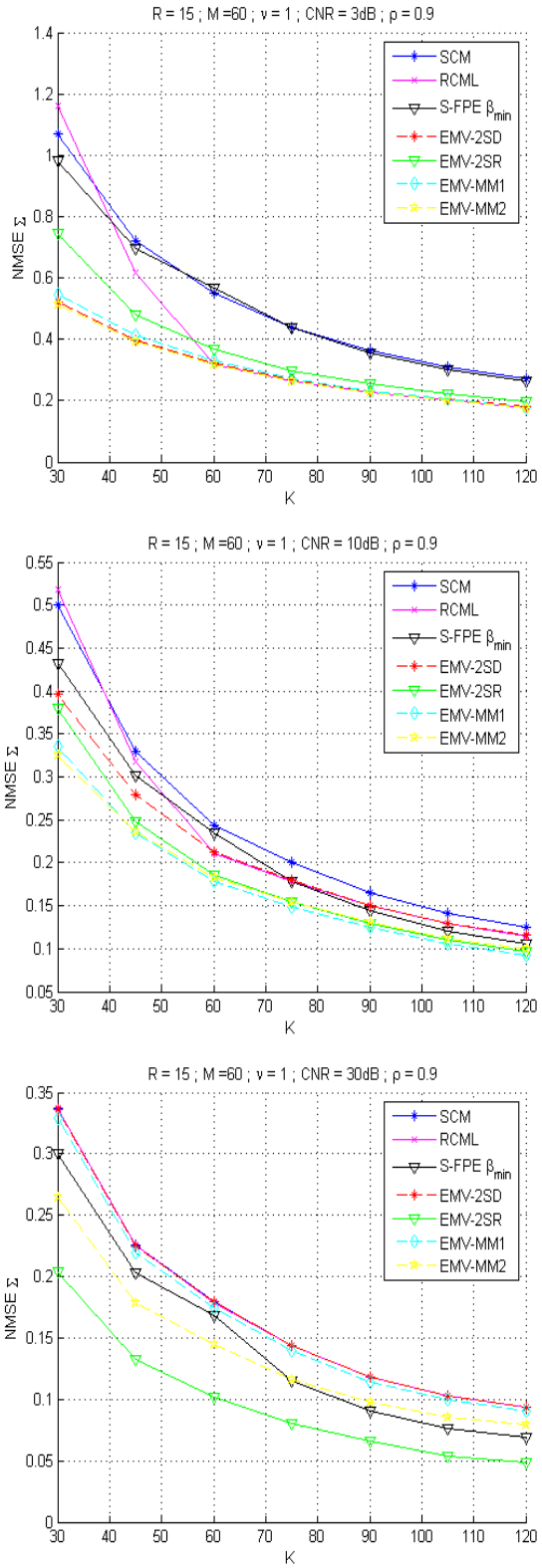

FIGURE 2.2 – NMSE sur $\mathbf{\Sigma}_{tot}$ des différents estimateurs. $M = 60$, $R = 15$, $\nu = 1$. De haut en bas : CNR$= [3, 10, 30]$dB.

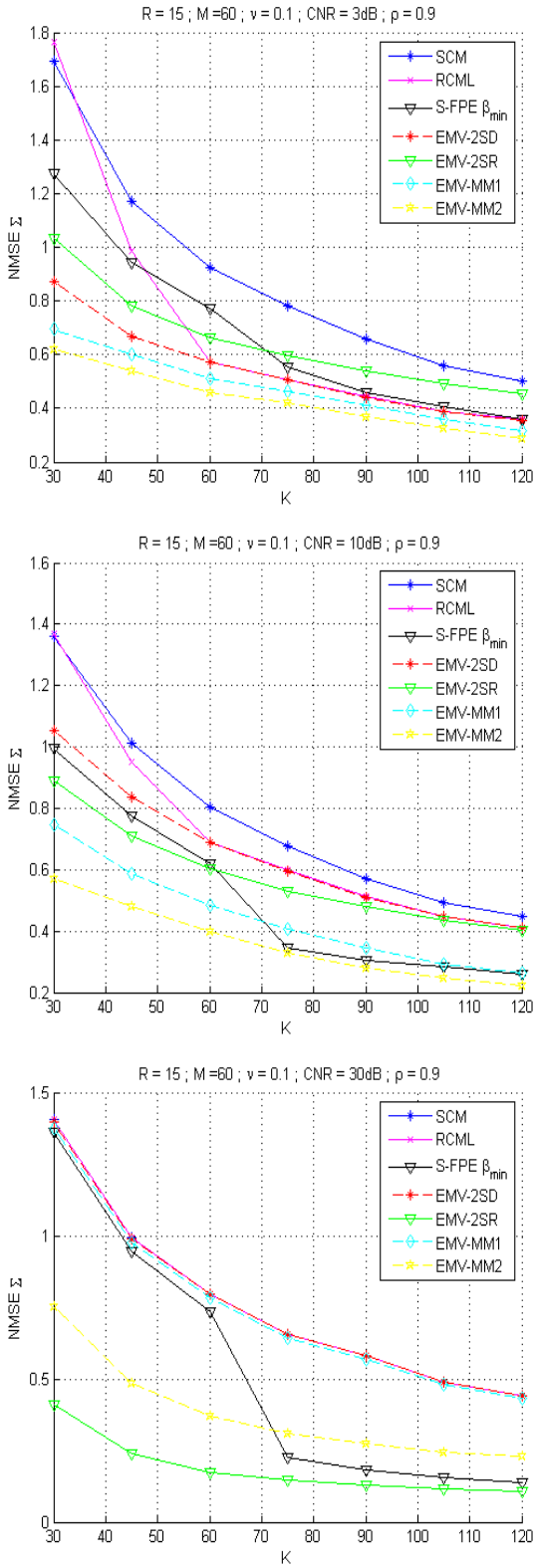

FIGURE 2.3 – NMSE sur $\mathbf{\Sigma}_{tot}$ des différents estimateurs. $M = 60$, $R = 15$, $\nu = 0.1$. De haut en bas : CNR$= [3, 10, 30]$dB.

Annexe A

Preuves du chapitre 2

A.1 Preuve du Théorème 2.3.1

La preuve du Théorème 2.3.1 se déroule comme suit : on dérive dans un premier temps les relations satisfaites par les EMV des paramètres τ_k et c_r, fonction des $\mathbf{v}_r$. Nous considérons ensuite la vraisemblance réduite où ces paramètres sont remplacés par leur EMV respectifs. L'expression obtenue est différentiée selon les $\mathbf{v}_r$ pour identifier la relation géométrique que satisfait leur EMV. On utilise alors le Lemme A.1.2, permettant de relier cette relation à la solution d'un problème d'optimisation sous contrainte.

Avant de commencer la preuve de ce théorème, la contrainte unitaire sur les paramètres $\{\mathbf{v}_r\}$ doit être exprimée. En effet, l'EMV de $\{\mathbf{v}_r\}$ doit former une base orthonormée, donc satisfaire les contraintes :

- $\mathbf{v}_r^H \mathbf{v}_r = 1$, for $r \in [\![1, R]\!]$
- $\mathbf{v}_r^H \mathbf{v}_j = 0$, for $r, j \in [\![1, R]\!]$ and $r \neq j$

Ces contraintes seront imposées à l'aide de multiplicateurs de Lagrange. Soit $\mathcal{L}(\{\mathbf{v}_r\})$ la fonction de Lagrange associée aux contraintes unitaires sur $\{\mathbf{v}_r\}$:

$$\mathcal{L}(\{\mathbf{v}_r\}) = \sum_{r=1}^{R} \lambda_r (\mathbf{v}_r^H \mathbf{v}_r - 1) + \sum_{r=1}^{R} \sum_{l=r+1}^{R} \mu_{r,l}\, \mathbf{v}_r^H \mathbf{v}_l + \sum_{r=1}^{R} \sum_{l=r+1}^{R} \mu_{r,l}^H\, \mathbf{v}_l^H \mathbf{v}_r \;, \tag{A.1}$$

où λ_r sont les multiplicateurs de Lagrange associés à la contrainte de normalisation et où $\mu_{r,l}$ sont les multiplicateurs de Lagrange associés à la contrainte d'orthogonalité. La présence des variables complexes conjuguées associées aux $\mu_{r,l}$ est due au fait que nous traitons des variables complexes. Cette fonction de Lagrange vérifie le lemme suivant :

Lemme A.1.1 *La dérivée de $\mathcal{L}(\{\mathbf{v}_r\})$ selon $\mathbf{v}_j^H$ pour $j \in [\![1, R]\!]$ est :*

$$\frac{\partial \mathcal{L}(\{\mathbf{v}_r\})}{\partial \mathbf{v}_j^H} = \mathbf{V}\mathbf{\Lambda}\mathbf{e}_j \;, \tag{A.2}$$

où $\mathbf{V}$ est la concaténation des vecteurs $\mathbf{v}_r$, $\mathbf{e}_j$ est le j^{ieme} vecteur de la base canonique, et $\mathbf{\Lambda}$ est la matrice hermitienne définie par (A.4).

Preuve A.1.1 *On a*

$$\frac{\partial \mathcal{L}(\{\mathbf{v}_r\})}{\partial \mathbf{v}_j^H} = \lambda_j \mathbf{v}_j + \sum_{l=j+1}^{R} \mu_{j,l} \mathbf{v}_l + \sum_{r=1}^{j-1} \mu_{r,j}^H \mathbf{v}_r \tag{A.3}$$

Cette expression peut être ré-écrite en utilisant la matrice des contraintes de Lagrange $\mathbf{\Lambda} \in R \times R$, définie par :

$$[\mathbf{\Lambda}]_{i,j} = \begin{cases} \mu_{i,j}^H & si\ i < j \\ \lambda_j & si\ i = j \\ \mu_{j,i} & si\ i > j \end{cases}, \tag{A.4}$$

Il est important de noter que la matrice $\mathbf{\Lambda}$ est hermitienne car $[\mathbf{\Lambda}]_{i,j} = [\mathbf{\Lambda}]_{j,i}^H$.

Nous dérivons ensuite le lemme A.1.2 :

Lemme A.1.2 *Soit $\{\mathbf{M}_r\}$ un jeu de R matrices hermitiennes de dimension $M \times M$. Les solutions locales du problème d'optimisation suivant :*

$$\begin{aligned} \underset{\{\mathbf{v}_r\}}{max} \quad & f_0(\{\mathbf{v}_r\}) = \sum_{r=1}^{R} \mathbf{v}_k^H \mathbf{M}_r \mathbf{v}_k \\ s.c. \quad & \mathbf{v}_r^H \mathbf{v}_r = 1\ ,\ r \in [\![1, R]\!] \\ & \mathbf{v}_r^H \mathbf{v}_j = 0\ ,\ r, j \in [\![1, R]\!]\ ,\ r \neq j \end{aligned}$$

satisfont

$$\mathbf{V}^H \left[\mathbf{M}_1 \mathbf{v}_1 | ... | \mathbf{M}_R \mathbf{v}_R\right] = \mathbf{\Lambda} \tag{A.5}$$

où $\mathbf{V}$ est la concaténation des $\{\mathbf{v}_r\}$, et $\mathbf{\Lambda}$ est la matrice hermitienne des multiplicateurs de Lagrange associées aux contraintes unitaires.

Preuve A.1.2 *En utilisant la méthode des multiplicateurs de Lagrange, la fonctionnelle à maximiser est $f_0(\{\mathbf{v}_r\}) - \mathcal{L}(\{\mathbf{v}_r\})$. La dérivée de cette fonction selon $\mathbf{v}_j$ est annulée pour :*

$$\frac{\partial (f_0 - \mathcal{L})(\{\mathbf{v}_r\})}{\partial \mathbf{v}_j^H} = 0 \Leftrightarrow \mathbf{M}_j \mathbf{v}_j = \mathbf{V} \mathbf{\Lambda} \mathbf{e}_j, \tag{A.6}$$

Donc, les R relations satisfaites par les solutions peuvent être concaténées en :

$$\left[\mathbf{M}_1 \mathbf{v}_1 | ... | \mathbf{M}_R \mathbf{v}_R\right] = \mathbf{V} \mathbf{\Lambda} \tag{A.7}$$

Multiplier A.7 à gauche par $\mathbf{V}^H$ et utiliser $\mathbf{V}^H \mathbf{V} = \mathbf{I}_R$ (vérifié sous contrainte unitaire) permet de conclure le lemme.

Tournons nous maintenant vers la preuve du théorème 2.3.1.

Preuve A.1.3 *Dans un premier temps, établissons les relations que vérifient les EMV des paramètres τ_k et c_r. La dérivée de la log-vraisemblance selon tous ces paramètres est :*

$$\frac{\partial \ln f}{\partial \tau_k} = \sum_{R=1}^{R} \left[\frac{c_r}{(1+\tau_k c_r)^2} \mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k - \frac{c_r}{1+\tau_k c_r} \right] , \tag{A.8}$$

$$\frac{\partial \ln f}{\partial c_r} = \sum_{k=1}^{K} \left[\frac{\tau_k}{(1+\tau_k c_r)^2} \mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k - \frac{\tau_k}{1+\tau_k c_r} \right] , \tag{A.9}$$

qui s'annule pour

$$\begin{cases} \sum_{R=1}^{R} \left[\frac{\hat{c}_r}{(1+\hat{\tau}_k \hat{c}_r)^2} \mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k - \frac{\hat{c}_r}{1+\hat{\tau}_k \hat{c}_r} \right] = 0 \quad \forall k \\ \sum_{k=1}^{K} \left[\frac{\hat{\tau}_k}{(1+\hat{\tau}_k \hat{c}_r)^2} \mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k - \frac{\hat{\tau}_k}{1+\hat{\tau}_k \hat{c}_r} \right] = 0 \quad \forall r \end{cases}$$

Ce système comprend $K+R$ polynômes de degré élevés et n'admet pas de simplification. L'EMV des paramètres τ_k et c_r n'a donc pas de forme tractable. Cependant, il est important de noter qu'ils annulent les équations ci dessus. On écrit alors la vraisemblance réduite (les paramètres τ_k et c_r remplacés par leur EMV) :

$$\log \hat{f} = \sum_{k=1}^{K} \sum_{r=1}^{R} \left[\frac{\hat{\tau}_k \hat{c}_r}{1+\hat{\tau}_k \hat{c}_r} \mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k - \log(1+\hat{\tau}_k \hat{c}_r) \right] + \mathcal{L} \tag{A.10}$$

avec $\mathcal{L}$ les contraintes de Lagrange. On dérive cette expression selon $\mathbf{v}_j$, j fixé :

$$\begin{aligned} \frac{\partial \hat{f}}{\partial \mathbf{v}_j} = \sum_{k=1}^{K} \frac{\hat{\tau}_k \hat{c}_r}{1+\hat{c}_r \hat{\tau}_k} \mathbf{z}_k \mathbf{z}_k^H \mathbf{v}_j + \frac{\partial \mathcal{L}}{\partial \mathbf{v}_j} + \\ \sum_{k=1}^{K} \sum_{r=1}^{R} \mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k \frac{\partial (\hat{c}_r \hat{\tau}_k)/\partial \mathbf{v}_j}{(1+\hat{c}_r \hat{\tau}_k)^2} - \sum_{k=1}^{K} \sum_{r=1}^{R} \frac{\partial (\hat{c}_r \hat{\tau}_k)/\partial \mathbf{v}_j}{1+\hat{c}_r \hat{\tau}_k} . \end{aligned} \tag{A.11}$$

En regroupant les deux derniers termes, on obtient :

$$\begin{aligned} \sum_{k=1}^{K} \sum_{r=1}^{R} \left(\frac{1}{1+\hat{c}_r \hat{\tau}_k} - \frac{\mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k}{(1+\hat{c}_r \hat{\tau}_k)^2} \right) \frac{\partial (\hat{d}_k \hat{\tau}_i)}{\partial \mathbf{v}_j} = \\ + \sum_{k=1}^{K} \sum_{r=1}^{R} \left(\frac{\hat{c}_r \mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k}{(1+\hat{c}_r \hat{\tau}_k)^2} - \frac{\hat{c}_r}{1+\hat{c}_r \hat{\tau}_k} \right) \frac{\partial (\hat{\tau}_k)}{\partial \mathbf{v}_j} \\ + \sum_{r=1}^{R} \sum_{k=1}^{K} \left(\frac{\hat{\tau}_k \mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k}{(1+\hat{c}_r \hat{\tau}_k)^2} - \frac{\hat{\tau}_k}{1+\hat{c}_r \hat{\tau}_k} \right) \frac{\partial (\hat{c}_r)}{\partial \mathbf{v}_j} . \end{aligned} \tag{A.12}$$

Toutes ces expressions sont annulées d'après les relations vérifiées par les EMV $\hat{c}_r$ et $\hat{\tau}_k$. On retrouve donc la relation :

$$\frac{\partial \hat{f}}{\partial \mathbf{v}_j^H} = \sum_{k=1}^{K} \frac{\hat{\tau}_k \hat{c}_r}{1+\hat{c}_r \hat{\tau}_k} \mathbf{z}_k \mathbf{z}_k^H \mathbf{v}_j + \frac{\partial \mathcal{L}}{\partial \mathbf{v}_j} \tag{A.13}$$

Puis, en notant

$$\hat{\mathbf{M}}_r = \sum_{k=1}^{K} \frac{\hat{\tau}_k \hat{c}_r}{1 + \hat{c}_r \hat{\tau}_k} \mathbf{z}_k \mathbf{z}_k^H \ , \tag{A.14}$$

Les R solutions selon chaque $\hat{\mathbf{v}}_r$ sont concaténées en :

$$\left[\hat{\mathbf{M}}_1 \hat{\mathbf{v}}_1 | ... | \hat{\mathbf{M}}_R \hat{\mathbf{v}}_R \right] = \hat{\mathbf{V}} \mathbf{\Lambda} \ . \tag{A.15}$$

En utilisant $\hat{\mathbf{V}}^H \hat{\mathbf{V}} = \mathbf{I}_R$ on obtient enfin :

$$\hat{\mathbf{V}}^H \left[\hat{\mathbf{M}}_1 \hat{\mathbf{v}}_1 | ... | \hat{\mathbf{M}}_R \hat{\mathbf{v}}_R \right] = \mathbf{\Lambda} \ . \tag{A.16}$$

Le lemme A.1.2 s'applique alors pour identifier les $\mathbf{v}_r$ comme solutions du problème d'optimisation contraint.

A.2 Preuve du Théorème 2.4.1

Preuve A.2.1 *La dérivée de la vraisemblance (2.16) par rapport à d_r^k, pour $r \in [\![1, R]\!]$ et $k \in [\![1, K]\!]$ fixés est :*

$$\frac{\partial \ln f(\{\mathbf{z}_k\} | \{\mathbf{v}_r\}, \{d_r^k\})}{\partial d_r^k} = -\frac{1}{1 + d_r^k} + \frac{\mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k}{(1 + d_r^k)^2} \ . \tag{A.17}$$

cette expression est annulée pour identifier $\hat{d}_r^k$, l'EMV de d_r^k. Cependant, ce facteur de puissance doit être une valeur positive. Puisque la vraisemblance est strictement croissante après le maximum $\hat{d}_r^k$, l'EMV sous contrainte de positivité est donc bien celui donné en théorème 2.4.1.

A.3 Preuve du Théorème 2.5.1

Preuve A.3.1 *Considérons les variables indépendantes de dimension réduite $\mathbf{z}_c$ and $\mathbf{z}_g$*

$$\tilde{\mathbf{z}} = \mathbf{V}^H \mathbf{z} = \begin{bmatrix} \mathbf{V}_c^H \mathbf{z} \\ \mathbf{V}_c^{\perp H} \mathbf{z} \end{bmatrix} = \begin{bmatrix} \mathbf{z}_c \\ \mathbf{z}_g \end{bmatrix} \ . \tag{A.18}$$

Chacune est distribuée selon :

$$\begin{cases} (\mathbf{z}_c | \tau_k) \sim \ \mathcal{CN}(\mathbf{0}, \tau_k \mathbf{C}_c) \ , \\ \qquad \mathbf{z}_g \sim \ \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_{M-R}) \ , \end{cases} . \tag{A.19}$$

Le résultat provient directement du fait que $\mathbf{U}_c^H \mathbf{z}$ est le vecteur $\mathbf{z}_c$ exprimé dans une base qui engendre le même espace que $\mathbf{V}_c$. En conséquence, $\mathbf{U}_c^H \mathbf{z}$ est un vecteur aléatoire distribué comme un CG dont la texture à une densité de probabilité inconnue. L'EMV correspond donc à l'estimateur du Point Fixe sur les variables de dimension réduite ([84, 89]).

Annexe B

Article : Développement des Algorithmes MM1 et MM2

Low-Complexity Algorithms for Low Rank Clutter Parameters Estimation in Radar Systems

Ying Sun, Arnaud Breloy, *Student Member, IEEE*, Prabhu Babu, Daniel Palomar, *Fellow, IEEE*, Frédéric Pascal, *Senior Member, IEEE*, Guillaume Ginolhac, *Member, IEEE*

***Abstract*—This paper addresses the problem of the clutter subspace projector estimation in the context of a disturbance composed of a low rank heterogeneous (Compound Gaussian) clutter and a white Gaussian noise. In such context, adaptive processing based on an estimated orthogonal projector onto the clutter subspace (instead of an estimated covariance matrix) requires less samples than classical methods. The clutter subspace estimate is usually derived from the Singular Value Decomposition of a covariance matrix estimate. However, it has been shown previously that a direct Maximum Likelihood Estimator of the clutter subspace projector can be obtained for the considered context. In this paper, we derive two algorithms based on the block Majorization-Minimization framework to reach this estimator. These algorithms are shown to be computationally faster than the state of the art with guaranteed convergence. Finally, the performance of the related estimators is illustrated on realistic Space Time Adaptive Processing for airborne radar simulations.**

***Index terms*—** Covariance Matrix and Projector estimation, Maximum Likelihood Estimator, Low Rank structure, Compound Gaussian, Majorization-Minimization.

I. Introduction

In array processing, many applications require the use of the covariance matrix (CM) of the noise: source localization techniques [3], [4], radar and sonar detection methods [5], [6] and filters [7]. In practice, the CM is unknown and has to be estimated from a set of samples $\mathbf{z}_k \in \mathbb{C}^M$, $k \in [\![1, K]\!]$, which are K signal-free independent realizations of the noise. The CM estimate is then used to perform the so-called adaptive process. In radar systems, the noise is composed of a correlated noise, referred to as *clutter* (caused by the response of the environment to the emitted signal), and a White Gaussian Noise (WGN, the thermal noise due to electronics). The total covariance of this disturbance is therefore:

$$\mathbf{\Sigma}_{tot} = \mathbf{\Sigma} + \sigma^2\mathbf{I}, \tag{1}$$

where $\mathbf{\Sigma}$ is the clutter CM and $\sigma^2\mathbf{I}$ is the CM of the WGN. In most cases, the clutter belongs to a subspace of limited dimension, meaning that the clutter CM $\mathbf{\Sigma}$ has rank $R < M$. Consequently, the total CM is structured as a Low Rank (LR) plus a scaled identity matrix.

When the clutter corresponds to a strong interference contained in a low dimensional subspace ($R \ll M$), one can use the following LR approximation [8], [9]:

$$\mathbf{\Sigma}_{tot}^{-1} \simeq \frac{1}{\sigma^2}\mathbf{\Pi}_c^{\perp} \simeq \frac{1}{\sigma^2}\left(\mathbf{I} - \mathbf{\Pi}_c\right),$$

where $\mathbf{\Pi}_c$, named clutter subspace projector, is the orthogonal projector onto the clutter subspace. This subspace is spanned by the R eigenvectors $\mathbf{v}_r$ associated to the R largest eigenvalues of the matrix $\mathbf{\Sigma}$, *i.e.*, $\mathbf{\Pi}_c = \sum_{r=1}^{R} \mathbf{v}_r\mathbf{v}_r^H$. This approximation allows developing adaptive process (*e.g.*, filters or detectors) that relies on a clutter subspace projector estimate $\hat{\mathbf{\Pi}}_c$ rather than a total CM estimate $\hat{\mathbf{\Sigma}}_{tot}$.

The practical use of this approximation is that adaptive LR process requires less samples to reach equivalent performance as the classical ones, which is valuable since the number of available samples is often limited. For example, if we consider the case of a Gaussian distributed clutter, the optimal adaptive filter is built from the Sample Covariance Matrix (SCM), which is the Maximum Likelihood Estimator (MLE) in that scenario. In this case, $K = 2M$ samples are required to ensure a 3dB loss of the output Signal to Noise Ratio (SNR) compared to the optimal non-adaptive filter [10]. If instead we use the MLE of the clutter subspace projector, which corresponds to the subspace spanned by the R largest eigenvectors of the SCM and can be obtained from its Singular Value Decomposition (SVD), one can build adaptive LR filter that reaches equivalent performance as the previous scheme with only $K = 2R \ll 2M$ samples [11].

However, it is now well-known that most modern radar clutter measurements are not Gaussian and behave heterogeneously. Therefore, the SCM may not provide the optimal solution since it is not an accurate estimator of the CM for heavy-tailed distributions or in the presence of outliers. To account for the heterogeneity of the clutter, one can model it with a Compound Gaussian distribution, which is a sub-familly of the Complex Elliptically Symmetric distribution [12]. The Compound Gaussian family covers a large panel of well-known and useful distributions, notably Gaussian, Weibull, K-distribution, t-distribution, etc. Moreover, it has a well-founded physical interpretation and presents good agreement to several real data sets [13]–[16]. Eventually, the total disturbance will be modeled in this paper as a LR Compound Gaussian clutter plus a WGN (as done in [9], [17]–[19]).

Ying Sun, Prabhu Babu, and Daniel P. Palomar are with the Hong Kong University of Science and Technology (HKUST), Hong Kong. e-mail: ysunac, eeprabhubabu, palomar@ust.hk. Arnaud Breloy is with the SATIE, ENS Cachan, CNRS, F-94230 Cachan, and with SONDRA, Centrale-Supélec, F-91192 Gif-sur-Yvette Cedex, France (e-mail: abreloy@satie.ens-cachan.fr). Guillaume Ginolhac is with the LISTIC, Université de Savoie Mont-Blanc, 74944 Annecy le Vieux Cedex, France (e-mail: guillaume.ginolhac@univ-savoie.fr). Frédéric Pascal is with L2S, Centrale-Supélec, F-91192 Gif-sur-Yvette Cedex, France (e-mail: frederic.pascal@centralesupelec.fr).

The work of Ying Sun, Prabhu Babu, and Daniel P. Palomar was supported by the Hong Kong RGC 617312 research grant. The work of Arnaud Breloy was supported by a DGA grant and the SONDRA PA10-DSOCO11127.

We point out that the sum of a Compound Gaussian and a WGN cannot be represented as a simple Compound Gaussian vector with different distribution parameters (except for the trivial Gaussian case). Nevertheless, most of the related work consider the case of a full rank Compound Gaussian clutter and neglect the possible LR plus WGN structure. Under this framework, robust estimation[1] of the CM can be performed using the so-called M-estimators [20]–[22] that are seen as generalized MLE for Complex Elliptically Symmetric distributions. A detailed review of this framework can be found in [12]. These estimators have been studied in details and successfully applied in modern detection/estimation literature due to their interest both from a theoretical and an application points of view [23]–[29]. We also remark that regularization of these estimators to handle under-sampled scenario [30]–[33], or taking advantage of prior knowledge on the CM structure [34]–[38], is currently an attractive topic of research. Instead of the SCM, it is possible to derive clutter subspace projector estimators from the SVD of a (regularized or not) M-estimator. However, despite their robustness properties, M-estimators may not achieve the optimal clutter subspace estimation performance as they do not take into account the noise structure that we consider in this work.

For this noise model, the seminal work [18] derived the MLE of the clutter subspace projector under the assumptions that the CM of the LR Compound Gaussian clutter has identical eigenvalues, and the Probability Density Function (PDF) of the texture is known. The assumption of known texture PDF has been relaxed in [39] by treating the texture as an unknown deterministic parameter. The expression of the MLE of the clutter CM parameters (eigenvectors and eigenvalues) was derived in [1], which does not have a simple closed-form and requires therefore an iterative algorithm to be reached. Initially, [1] focused on the clutter subspace projector MLE and proposed an ad-hoc algorithm to obtain it. This leads to an accurate clutter subspace projector estimator, but a poor recovery of the total CM. In [2], the high Clutter to Noise Ratio (CNR) assumption was made to derive a general "2-Step MLE" algorithm (with several possible adaptations) that was shown to provide accurate estimates of the CM. Most algorithms developed [1], [2] require the use of a gradient descent algorithm on manifold proposed in [40], which can be difficult to tune and computationally costly.

In this paper, we propose to apply the block Majorization-Minimizaton algorithm framework to the problem of computing the considered MLE. We derive two new algorithms that enjoy the following properties:

- They do not rely on any heuristic (as [1]) or high CNR assumption (as [2]).
- They are computationally faster than the ones derived in [1], [2], hence more suitable for implementation.
- Compared to the gradient descent algorithm [40], which requires many trials to find the descent step, the proposed algorithms only require basic matrix operations or SVD.

The paper is organized as follows. Section II states the problem formulation and briefly reviews the block Majorization-Minimizaton methodology. Then, Sections III and IV derive the two new algorithms for computing the MLE of the considered problem. The difference between the two algorithms lies in the choice of the variables with respect to which the likelihood function is cyclically optimized. Section V validates the performance of the proposed algorithms with simulations. Moreover, the proposed algorithms are applied to a realistic Space Time Adapive Processing (STAP) for airborne radar simulation. Finally, Section VI draws conclusions of this study.

Throughout the paper the following convention is adopted: italic indicates a scalar quantity, lower case boldface indicates a vector quantity and upper case boldface a matrix. ^{H} denotes the transpose conjugate operator or the simple conjugate operator for a scalar quantity. ^{T} denotes the transpose operator. $\mathcal{CN}(\mathbf{a}, \mathbf{\Sigma})$ is a complex-valued Gaussian distribution of mean $\mathbf{a}$ and covariance matrix $\mathbf{\Sigma}$. $\mathbf{I}$ is the identity matrix of appropriate dimension. $\det(\cdot)$ and $\mathrm{Tr}(\cdot)$ stand for the determinant and the trace of a matrix, respectively. $\hat{d}$ is an estimate of the parameter d. $\{w_n\}_{n\in[\![1,N]\!]}$ denotes the set of n elements w_n with $n \in [\![1, n]\!]$, and will often be abbreviated to $\{w_n\}$ in the sequel. $\mathbf{e}_i$ is the i^{th} vector of the canonical basis of appropriate dimension.

II. Background

A. Problem Formulation

We assume that K samples $\{\mathbf{z}_k\}_{k\in[\![1,K]\!]}$ are available. Each of these data $\mathbf{z}_k \in \mathbb{C}^M$ corresponds to a realization of a proper circular LR Compound Gaussian process $\mathbf{c}_k$ plus an independent additive zero-mean complex WGN $\mathbf{n}_k$:

$$\mathbf{z}_k = \mathbf{n}_k + \mathbf{c}_k. \tag{2}$$

The WGN $\mathbf{n}_k$ follows the distribution:

$$\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}), \tag{3}$$

where the power of the WGN σ^2 is assumed to be known[2] and fixed to be $\sigma^2 = 1$ without loss of generality. The LR Compound Gaussian [12] $\mathbf{c}_k$ is a M-dimensional zero-mean complex Gaussian vector (the *speckle*) with CM $\mathbf{\Sigma}$, multiplied by the square root of an independent positive random power factor (the *texture*) τ_k. Each $\mathbf{c}_k$ follows then, conditionally to τ_k:

$$(\mathbf{c}_k|\tau_k) \sim \mathcal{CN}(\mathbf{0}, \tau_k\mathbf{\Sigma}), \tag{4}$$

where the Compound Gaussian clutter CM $\mathbf{\Sigma}$ has a Low Rank structure with $\mathrm{rank}(\Sigma) = R < M$. The rank R is assumed to be known and a discussion on this assumption will be provided in the remark below. In this work, we do not assume the knowledge of the texture PDF, and treat each τ_k as an unknown deterministic variable. The likelihood function is then:

$$f(\{\mathbf{z}_k\}|\mathbf{\Sigma}, \{\tau_k\}) = \prod_{k=1}^{K} \frac{e^{-\mathbf{z}_k^H\mathbf{\Sigma}_k^{-1}\mathbf{z}_k}}{\pi^M \det(\mathbf{\Sigma}_k)}, \tag{5}$$

with $\mathbf{\Sigma}_k = \tau_k\mathbf{\Sigma} + \mathbf{I}$.

[1] *i.e.*, not highly sensitive to the underlying distribution

[2] This hypothesis is made for describing a valid theoretical framework. In practice, presented results can be applied with an estimate of σ^2 used as its actual value.

The MLE of the clutter CM $\mathbf{\Sigma}$ is therefore defined as the minimizer of the following problem (equivalent to the maximizer of the log-likelihood function):

$$\begin{array}{ll}\underset{\mathbf{\Sigma}_k,\{\tau_k\},\mathbf{\Sigma}}{\text{minimize}} & \sum_{k=1}^{K}\log\det\left(\mathbf{\Sigma}_k\right)+\sum_{k=1}^{K}\mathbf{z}_k^H\mathbf{\Sigma}_k^{-1}\mathbf{z}_k \\ \text{subject to} & \mathbf{\Sigma}_k=\tau_k\mathbf{\Sigma}+\mathbf{I} \\ & \tau_k\geq 0 \\ & \text{rank}\left(\mathbf{\Sigma}\right)\leq R,\end{array} \tag{A}$$

where $\mathbf{\Sigma}$ is of size $M\times M$. Denote the objective function by $L\left(\mathbf{\Sigma},\{\tau_k\}\right)$.

Remark: The known rank assumption is made for deriving a valid and tractable theoretical framework. Results derived in this work can be applied with a plug-in estimate of the rank. There are several methods of rank estimation present in the literature ($e.g.$ [41] and references therein), but this topic is beyond the scope of the paper. Moreover, in some applications, the clutter rank can be obtained by the geometry of the system (such as in STAP [42] thanks to the Brennan rule [43]).

B. Block MM principle

To solve Problem (A), we adopt the block Majorization-Minimization (MM) algorithm framework, which is stated below briefly.

Consider the following optimization problem

$$\begin{array}{ll}\underset{\mathbf{x}}{\text{minimize}} & f\left(\mathbf{x}\right)\\ \text{subject to} & \mathbf{x}\in\mathcal{X},\end{array} \tag{6}$$

where the optimization variable $\mathbf{x}$ can be partitioned into m blocks as $\mathbf{x}=\left(\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(m)}\right)$, with each n_i-dimensional block $\mathbf{x}^{(i)}\in\mathcal{X}_i$ and $\mathcal{X}=\prod_{i=1}^{m}\mathcal{X}_i$.

At the $(t+1)$-th iteration, the i-th block $\mathbf{x}^{(i)}$ is updated by solving the following problem:

$$\begin{array}{ll}\underset{\mathbf{x}^{(i)}}{\text{minimize}} & g_i\left(\mathbf{x}^{(i)}|\mathbf{x}_t\right)\\ \text{subject to} & \mathbf{x}^{(i)}\in\mathcal{X}_i\end{array} \tag{7}$$

with $i=(t \bmod m)+1$ (so blocks are updated in cyclic order) and the continuous surrogate function $g_i\left(\mathbf{x}^{(i)}|\mathbf{x}_t\right)$ satisfying the following properties:

$$\begin{aligned} f\left(\mathbf{x}_t\right) &= g_i\left(\mathbf{x}_t^{(i)}|\mathbf{x}_t\right),\\ f\left(\mathbf{x}_t^{(1)},\ldots,\mathbf{x}^{(i)},\ldots,\mathbf{x}_t^{(m)}\right) &\leq g_i\left(\mathbf{x}^{(i)}|\mathbf{x}_t\right)\ \forall\mathbf{x}^{(i)}\in\mathcal{X}_i,\\ f'\left(\mathbf{x}_t;\mathbf{d}_i^0\right) &= g_i'\left(\mathbf{x}_t^{(i)};\mathbf{d}_i|\mathbf{x}_t\right)\\ &\quad\ \forall\mathbf{x}_t^{(i)}+\mathbf{d}_i\in\mathcal{X}_i,\\ &\quad\ \mathbf{d}_i^0\triangleq\left(\mathbf{0};\ldots;\mathbf{d}_i;\ldots;\mathbf{0}\right),\end{aligned}$$

where $f'(\mathbf{x};\mathbf{d})$ stands for the directional derivative at $\mathbf{x}$ along $\mathbf{d}$. In short, at each iteration, the block MM algorithm updates the variables in one block by minimizing a tight upperbound of the function while keeping the value of the other blocks fixed.

In practice, the surrogate functions are usually designed so that each sub-problem (7) can be solved easily, for example in closed-form.

III. Direct Block Majorization-Minimization Algorithm

As $\text{rank}\left(\mathbf{\Sigma}\right)\leq R$, the variable $\mathbf{\Sigma}$ can be reparametrized as $\mathbf{\Sigma}=\mathbf{W}\mathbf{W}^H$ with $\mathbf{W}\in\mathbb{C}^{M\times R}$. Problem (A) can then be written equivalently as:

$$\begin{array}{ll}\underset{\mathbf{\Sigma}_k,\{\tau_k\},\mathbf{W}}{\text{minimize}} & \sum_{k=1}^{K}\log\det\left(\mathbf{\Sigma}_k\right)+\sum_{k=1}^{K}\mathbf{z}_k^H\mathbf{\Sigma}_k^{-1}\mathbf{z}_k\\ \text{subject to} & \mathbf{\Sigma}_k=\tau_k\mathbf{W}\mathbf{W}^H+\mathbf{I}\\ & \tau_k\geq 0.\end{array} \tag{B}$$

Following the block MM methodology, we partition the variables as $\left\{\{\tau_k\},\mathbf{W}\right\}$ and derive an algorithm that updates the blocks in cyclic order (note that variables $\mathbf{\Sigma}_k$ are implicitly optimized at every iteration while optimizing either τ_k or $\mathbf{W}$). In each iteration, an upperbound of the objective function is minimized, which guarantees a monotonic decrement of the objective value.

To be precise, given a starting point $\left\{\{\tau_k\}^{t=0},\mathbf{W}^{t=0}\right\}$, one iteratively

- updates $\{\tau_k\}^{t+1}$ for fixed $\mathbf{W}=\mathbf{W}^t$ by minimizing a set of surrogate functions $L\left(\tau_k|\tau_k^t,\mathbf{W}\right)$ for $k\in[\![1,K]\!]$,
- updates $\mathbf{W}^{t+1}$ for fixed $\{\tau_k\}=\{\tau_k\}^{t+1}$ by minimizing a surrogate function $L\left(\mathbf{W}|\left\{\tau_k\right\},\mathbf{W}^t\right)$,

until convergence, which will produce a stationary point of Problem (B). This procedure is summed up in the box Algorithm 1.

A. Update $\{\tau_k\}$ with fixed $\mathbf{W}$

Let $\mathbf{W}=\mathbf{W}^t$ and $\mathbf{\Sigma}=\mathbf{W}^t\mathbf{W}^{tH}$. To lighten notation, we omit the reference on t for these variables in this part. The objective function is separable in the τ_k's, and for each of them, the following problem should be solved:

$$\begin{array}{ll}\underset{\tau_k}{\text{minimize}} & \log\det\left(\tau_k\mathbf{\Sigma}+\mathbf{I}\right)+\mathbf{z}_k^H\left(\tau_k\mathbf{\Sigma}+\mathbf{I}\right)^{-1}\mathbf{z}_k\\ \text{subject to} & \tau_k\geq 0.\end{array} \tag{B1}$$

Eigendecompose $\mathbf{\Sigma}$ as $\mathbf{\Sigma}=\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, with $\mathbf{\Lambda}=\text{diag}\left(\lambda_1,\ldots,\lambda_R,0,\ldots,0\right)$ and $\mathbf{U}=[\mathbf{u}_1,\ldots,\mathbf{u}_M]$. The objective function of (B1) can be simplified to:

$$\begin{aligned} L\left(\tau_k|\mathbf{W}\right) &= \sum_{m=1}^{M}\log\left(\tau_k\lambda_m+1\right)+\sum_{m=1}^{M}s_{km}\left(\tau_k\lambda_m+1\right)^{-1}\\ &= \sum_{m=1}^{R}\log\left(\tau_k\lambda_m+1\right)+\sum_{m=1}^{R}s_{km}\left(\tau_k\lambda_m+1\right)^{-1}+\text{const.},\end{aligned} \tag{8}$$

where $s_{km}=\left\|\mathbf{z}_k^H\mathbf{u}_m\right\|^2$. This function is the sum of quasi-convex functions, which is not necessarily quasi-convex, and has no closed-form minimizer.

Applying the block MM algorithm, we find an upperbound (surrogate function) of $L\left(\tau_k|\mathbf{W}\right)$, with equality achieved at

$\tau_k = \tau_k^t$, and minimize this surrogate function. The update of τ_k is derived based on the following two propositions:

Proposition 1. *The function $L(\tau_k|\mathbf{W})$ in (8) can be upperbounded by the surrogate function $L(\tau_k|\tau_k^t, \mathbf{W})$ defined as*

$$L\left(\tau_k|\tau_k^t,\mathbf{W}\right) = -\beta_k \log \tau_k + \left(\sum_{m=1}^{R}\alpha_{km}\right)\log\left(\frac{\left(\sum_{m=1}^{R}\frac{\alpha_{km}\lambda_m}{1+\lambda_m\tau_k^t}\right)}{\sum_{m=1}^{R}\alpha_{km}}\tau_k + \frac{\sum_{m=1}^{R}\frac{\alpha_{km}}{1+\lambda_m\tau_k^t}}{\sum_{m=1}^{R}\alpha_{km}}\right) + \text{const.}, \tag{9}$$

where

$$\alpha_{km} = s_{km}\frac{\tau_k^t\lambda_m}{\tau_k^t\lambda_m+1} + 1 \tag{10}$$

and

$$\beta_k = \sum_{m=1}^{R} s_{km}\frac{\tau_k^t\lambda_m}{\tau_k^t\lambda_m+1}. \tag{11}$$

The equality is achieved at $\tau_k = \tau_k^t$.

Proof: See Appendix A. ∎

Proposition 2. *The surrogate function $L(\tau_k|\tau_k^t, \mathbf{W})$ is quasi-convex and has a unique minimizer given by*

$$\tau_k^{t+1} = \frac{1}{R}\cdot\frac{\left(\sum_{m=1}^{R} s_{km}\frac{\tau_k^t\lambda_m}{\tau_k^t\lambda_m+1}\right)\cdot\left(\sum_{m=1}^{R}\frac{\alpha_{km}}{1+\lambda_m\tau_k^t}\right)}{\sum_{m=1}^{R}\frac{\alpha_{km}\lambda_m}{1+\lambda_m\tau_k^t}}. \tag{12}$$

Proof: See Appendix B. ∎

B. Update $\mathbf{W}$ with Fixed $\{\tau_k\}$

Let now fix $\{\tau_k\} = \{\tau_k\}^{t+1}$. To lighten the notation, we omit the reference on t for this set of variables in this part. To obtain the update of $\mathbf{W}$, we need to solve:

$$\begin{aligned}\underset{\mathbf{\Sigma}_k,\mathbf{W}}{\text{minimize}} \quad & \sum_{k=1}^{K}\log\det\left(\mathbf{\Sigma}_k\right) + \sum_{k=1}^{K}\mathbf{z}_k^H\mathbf{\Sigma}_k^{-1}\mathbf{z}_k \\ \text{subject to} \quad & \mathbf{\Sigma}_k = \tau_k\mathbf{W}\mathbf{W}^H + \mathbf{I}.\end{aligned} \tag{B2}$$

Without loss of generality assume that $\tau_k > 0$, otherwise the terms with $\tau_k = 0$ can be deleted from the summation in the objective function as they are constants.

Problem (B2) has no closed-form minimizer. As in the previous part, we are going to derive a tight upperbound of the objective function $L(\mathbf{W}|\tau_k)$, with equality achieved at $\mathbf{W} = \mathbf{W}^t$.

Proposition 3. *The function*

$$L\left(\mathbf{W}|\tau_k\right) = \sum_{k=1}^{K}\log\det\left(\mathbf{\Sigma}_k\right) + \sum_{k=1}^{K}\mathbf{z}_k^H\mathbf{\Sigma}_k^{-1}\mathbf{z}_k, \tag{13}$$

where $\mathbf{\Sigma}_k = \tau_k\mathbf{W}\mathbf{W}^H + \mathbf{I}$, can be upperbounded by the convex quadratic function

$$L\left(\mathbf{W}|\tau_k,\mathbf{W}^t\right) = \mathrm{Tr}\left(\mathbf{W}\mathbf{H}\mathbf{W}^H\right) - \mathrm{Tr}\left(\mathbf{L}\mathbf{W}^H\right) - \mathrm{Tr}\left(\mathbf{L}^H\mathbf{W}\right), \tag{14}$$

with equality achieved at $\mathbf{W} = \mathbf{W}^t$, where the matrices $\mathbf{H}$ and $\mathbf{L}$ are defined according to (40)*,* (41)*,* (45)*, and* (46)*.*

Algorithm 1 Direct Block Majorization-Minimization Algorithm

1: Compute the sample covariance matrix $\hat{\mathbf{\Sigma}}^{\mathrm{SCM}} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{z}_k\mathbf{z}_k^H$.
2: Eigendecompose $\hat{\mathbf{\Sigma}}^{\mathrm{SCM}}$ as $\hat{\mathbf{\Sigma}}^{\mathrm{SCM}} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, where $\mathbf{\Lambda} = \mathrm{diag}\left(\lambda_1,\ldots,\lambda_M\right)$ with $\lambda_1 \geq \cdots \geq \lambda_M$.
3: Initialize $\mathbf{W}$ to be $\mathbf{W} = \mathbf{U}\mathrm{diag}\left(\sqrt{\lambda_1},\ldots,\sqrt{\lambda_R},0,\ldots,0\right)$, and $\{\tau_k\}_{k=1}^K$ to be arbitrary positive real numbers.
4: **repeat**
5: Update $\{\tau_k\}_{k=1}^K$ with (10) and (12).
6: Update $\mathbf{W}$ with (17).
7: **until** Some convergence criterion is met
8: Eigendecompose $\mathbf{W}\mathbf{W}^H$ as $\mathbf{V}\mathbf{D}\mathbf{V}^H$.
9: $\hat{\mathbf{\Pi}}_c = \mathbf{V}\mathbf{V}^H$.

Proof: See Appendix C. ∎

The matrix $\mathbf{H}$ is positive definite by definition, hence the upperbound $L(\mathbf{W}|\tau_k,\mathbf{W}^t)$ has a unique minimizer given by

$$\mathbf{W}^{t+1} = \mathbf{L}\mathbf{H}^{-1}.$$

To efficiently compute this solution, let the SVD of $\mathbf{W}^t$ be $\mathbf{W}^t = \mathbf{U}\mathbf{S}\mathbf{V}^H$. Then

$$\mathbf{L} = \sum_{k=1}^{K}\mathbf{L}_k^t = \left(\sum_{k=1}^{K}\mathbf{z}_k\mathbf{z}_k^H\mathbf{U}\mathbf{S}\left(\tau_k^{-1}\mathbf{I}+\mathbf{S}^2\right)^{-1}\right)\mathbf{V}^H, \tag{15}$$

and

$$\begin{aligned}\mathbf{H} &= \sum_{k=1}^{K}\left(\left(\mathbf{W}^t\right)^H\mathbf{W}^t + \tau_k^{-1}\mathbf{I}\right)^{-1} + \sum_{k=1}^{K}\mathbf{H}_k^t \\ &= \sum_{k=1}^{K}\mathbf{V}\left(\tau_k^{-1}\mathbf{I}+\mathbf{S}^2\right)^{-1}\mathbf{S}^H\mathbf{U}^H\mathbf{z}_k\mathbf{z}_k^H\mathbf{U}\mathbf{S}\left(\tau_k^{-1}\mathbf{I}+\mathbf{S}^2\right)^{-1}\mathbf{V}^H \\ &\quad + \sum_{k=1}^{K}\mathbf{V}\left(\tau_k^{-1}\mathbf{I}+\mathbf{S}^2\right)^{-1}\mathbf{V}^H,\end{aligned} \tag{16}$$

where $\mathbf{S}^2 = \mathbf{S}^H\mathbf{S}$ (since $\mathbf{S}$ is a real diagonal matrix). $\mathbf{W}^{t+1}$ can therefore be computed as:

$$\begin{aligned}\mathbf{W}^{t+1} = &\left(\sum_{k=1}^{K}\mathbf{z}_k\mathbf{z}_k^H\mathbf{U}\mathbf{S}\left(\tau_k^{-1}\mathbf{I}+\mathbf{S}^2\right)^{-1}\right)\times \\ &\left(\sum_{k=1}^{K}\left(\tau_k^{-1}\mathbf{I}+\mathbf{S}^2\right)^{-1} + \right. \\ &\left.\left(\tau_k^{-1}\mathbf{I}+\mathbf{S}^2\right)^{-1}\mathbf{S}^H\mathbf{U}^H\mathbf{z}_k\mathbf{z}_k^H\mathbf{U}\mathbf{S}\left(\tau_k^{-1}\mathbf{I}+\mathbf{S}^2\right)^{-1}\right)^{-1}\mathbf{V}^H,\end{aligned} \tag{17}$$

where only scalar inversion and matrix multiplication operations are involved in the computation.

C. Convergence Analysis

Proposition 4. *Any limit point of the pair $\{\{\tau_k^t\}, \mathbf{W}^t\}$ generated by the Algorithm 1 is a stationary point of Problem* (B)*.*

Proof: We have proved the quasi-convexity and the uniqueness of the minimizer of the surrogate functions $L\left(\tau_k|\tau_k^t,\mathbf{W}^t\right)$ and $L\left(\mathbf{W}|\tau_k,\mathbf{W}^t\right)$ in the previous subsections. The algorithm convergence is a direct application of Theorem 2 (a) in [44]. ∎

IV. Eigenspace Block Majorization-Minimization Algorithm

As $\operatorname{rank}(\mathbf{\Sigma}) \leq R$, the variable $\mathbf{\Sigma}$ can be reparametrized by its eigendecomposition:

$$\mathbf{\Sigma} = \sum_{r=1}^{R} c_r \mathbf{v}_r \mathbf{v}_r^H.$$

Problem (A) can then be rewritten as

$$\begin{aligned} \underset{\{\tau_k\},\{c_r\},\{\mathbf{v}_r\}}{\text{minimize}} \quad & -\sum_{k=1}^{K}\sum_{r=1}^{R} \frac{\tau_k c_r}{\tau_k c_r + 1} \mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k \\ & + \sum_{k=1}^{K}\sum_{r=1}^{R} \log\left(1+\tau_k c_r\right) \\ \text{subject to} \quad & \tau_k \geq 0,\ c_r \geq 0,\ \text{orthonormal } \mathbf{v}_r\text{'s}. \end{aligned} \tag{C}$$

Following the same methodology as in the previous section, we partition the variables as $\left\{\{\tau_k\},\{c_r\},\{\mathbf{v}_r\}\right\}$ and derive an algorithm that updates the blocks in cyclic order by minimizing an upperbound of the objective.

Given a starting point $\left\{\{\tau_k\}^{t=0},\{c_r\}^{t=0},\{\mathbf{v}_r\}^{t=0}\right\}$, one iteratively

- updates $\{\tau_k\}^{t+1}$ for fixed $\{c_r\} = \{c_r\}^t$ and $\{\mathbf{v}_r\} = \{\mathbf{v}_r\}^t$ by minimizing a set of surrogates functions $L\left(\tau_k|\tau_k^t,\{c_r\},\{\mathbf{v}_r\}\right)$ for $k \in [\![1,K]\!]$,
- updates $\{c_r\}^{t+1}$ for fixed $\{\tau_k\} = \{\tau_k\}^{t+1}$ and $\{\mathbf{v}_r\} = \{\mathbf{v}_r\}^t$ by minimizing a set of surrogates functions $L\left(c_k|\{\tau_k\},c_r^t,\{\mathbf{v}_r\}\right)$ for $r \in [\![1,R]\!]$,
- updates $\{\mathbf{v}_r\}^{t+1}$ for fixed $\{\tau_k\} = \{\tau_k\}^{t+1}$ and $\{c_r\} = \{c_r\}^{t+1}$ by minimizing the objective $w.r.t.$ $\{\mathbf{v}_r\}$ under orthonormality constraints. This step is done by iteratively minimizing surrogate functions $L\left(\{\mathbf{v}_r\}|\{\tau_k\}\{c_r\},\{\mathbf{v}_r\}^t\right)$ (so there is an inner-loop in the block MM iterations).

This procedure is summed up in the box Algorithm 2.

A. Update $\{\tau_k\}$ with Fixed $\{c_r\}$ and $\{\mathbf{v}_r\}$

Let $\{c_r\} = \{c_r\}^t$ and $\{\mathbf{v}_r\} = \{\mathbf{v}_r\}^t$. To lighten notation, we omit the reference on t for these variables in this part. The objective function is separable in the τ_k's, and for each of them, the following problem (noticing that $\mathbf{\Sigma}$ has $M-R$ zero eigenvalues) should be solved:

$$\begin{aligned} \underset{\tau_k}{\text{minimize}} \quad & \sum_{r=1}^{R} \log\left(1+\tau_k c_r\right) + \sum_{r=1}^{R} \frac{s_{kr}}{1+\tau_k c_r} \\ \text{subject to} \quad & \tau_k \geq 0, \end{aligned} \tag{C1}$$

with $s_{kr} = \left\|\mathbf{z}_k^H \mathbf{v}_r\right\|^2$. Notice that the objective function of (C1) has the same form as (8), therefore we can apply Propositions 1 and 2 to obtain the τ_k's updates as:

$$\tau_k^{t+1} = \frac{1}{R} \cdot \frac{\left(\sum_{r=1}^{R} s_{kr} \frac{\tau_k^t c_r}{\tau_k^t c_r + 1}\right) \cdot \left(\sum_{r=1}^{R} \frac{\alpha_{kr}}{1+c_r \tau_k^t}\right)}{\sum_{r=1}^{R} \frac{\alpha_{kr} c_r}{1+c_r \tau_k^t}}, \tag{18}$$

where $\alpha_{kr} = s_{kr} \frac{\tau_k^t c_r}{\tau_k^t c_r + 1} + 1$.

B. Update $\{c_r\}$ with Fixed $\{\tau_k\}$ and $\{\mathbf{v}_r\}$

Let $\{\tau_k\} = \{\tau_k\}^{t+1}$ and $\{\mathbf{v}_r\} = \{\mathbf{v}_r\}^t$. As before we omit the reference on t for these variables in this part. The objective function is separable in the c_r's, and for each of them, the following problem should be solved:

$$\begin{aligned} \underset{c_r}{\text{minimize}} \quad & \sum_{k=1}^{K} \log\left(1+\tau_k c_r\right) + \sum_{k=1}^{K} \frac{s_{kr}}{1+\tau_k c_r} \\ \text{subject to} \quad & c_r \geq 0, \end{aligned} \tag{C2}$$

with $s_{kr} = \left\|\mathbf{z}_k^H \mathbf{v}_r\right\|^2$. Notice that the c_r's in (C2) play a similar role as the τ_k's in (C1). Similar to the update of τ_k, we can apply Propositions 1 and 2 to obtain the c_r's updates as:

$$c_r^{t+1} = \frac{1}{K} \cdot \frac{\left(\sum_{k=1}^{K} s_{kr} \frac{\tau_k c_r^t}{\tau_k c_r^t + 1}\right) \cdot \left(\sum_{k=1}^{K} \frac{\alpha_{kr}}{1+c_r^t \tau_k}\right)}{\sum_{k=1}^{K} \frac{\alpha_{kr} \tau_k}{1+c_r^t \tau_k}}, \tag{19}$$

where $\alpha_{kr} = s_{kr} \frac{\tau_k c_r^t}{\tau_k c_r^t + 1} + 1$.

C. Update $\{\mathbf{v}_r\}$ with Fixed $\{\tau_k\}$ and $\{c_r\}$

With $\{\tau_k\}$ and $\{c_r\}$ fixed, minimizing the objective $w.r.t.$ $\{\mathbf{v}_r\}$ is equivalent to solving the problem:

$$\begin{aligned} \underset{\{\mathbf{v}_r\}}{\text{maximize}} \quad & \sum_{r=1}^{R} \mathbf{v}_r^H \mathbf{M}_r \mathbf{v}_r \\ \text{subject to} \quad & \text{orthonormal } \mathbf{v}_r\text{'s}, \end{aligned} \tag{C3}$$

where

$$\mathbf{M}_r = \sum_{k=1}^{K} \frac{\tau_k c_r}{\tau_k c_r + 1} \mathbf{z}_k \mathbf{z}_k^H. \tag{20}$$

We start with the following proposition.

Proposition 5. *The function*

$$L\left(\{\mathbf{v}_r\}|\{\tau_k\},\{c_r\}\right) = \sum_{r=1}^{R} \mathbf{v}_r^H \mathbf{M}_r \mathbf{v}_r \tag{21}$$

can be lowerbounded by the surrogate function

$$\begin{aligned} & L\left(\{\mathbf{v}_r\}|\{\tau_k\},\{c_r\},\{\mathbf{v}_r\}^t\right) \\ & = \sum_{r=1}^{R} \left[\left(\mathbf{v}_r^t\right)^H \mathbf{M}_r \mathbf{v}_r + \mathbf{v}_r^H \mathbf{M}_r \mathbf{v}_r^t\right] + \text{const}. \end{aligned} \tag{22}$$

with equality achieved at $\{\mathbf{v}_r\} = \{\mathbf{v}_r\}^t$.

Proof: As the matrices $\mathbf{M}_r$ are Hermitian positive semidefinite, the objective function is convex, and therefore

Algorithm 2 Eigenspace Block Majorization-Minimization Algorithm

1: Compute the sample covariance matrix $\hat{\mathbf{\Sigma}}^{\mathrm{SCM}} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{z}_k\mathbf{z}_k^H$.
2: Initialize $\{c_r\}_{r=1}^{R}$ to be the first R leading eigenvalues of $\hat{\mathbf{\Sigma}}^{\mathrm{SCM}}$, $\{\mathbf{v}_r\}_{r=1}^{R}$ to be the corresponding eigenvectors, and $\{\tau_k\}_{k=1}^{K}$ to be arbitrary positive real numbers.
3: **repeat**
4: Update τ_k with (18).
5: Update c_r with (19).
6: Compute $\mathbf{M}_r$ with (20).
7: **repeat** (*optional inner loop*)
8: Compute $\mathbf{A}$ with (25).
9: Decompose $\mathbf{A}$ as $\mathbf{A} = \mathbf{V}_A\mathbf{D}_A\mathbf{U}_A^H$ (Thin SVD).
10: Update $\mathbf{V}$ as $\mathbf{V} = \mathbf{V}_A\mathbf{U}_A^H$.
11: **until** Some convergence criterion is met
12: **until** Some convergence criterion is met
13: $\hat{\mathbf{\Pi}}_c = \mathbf{V}\mathbf{V}^H$.

is minorized by its first order Taylor expansion at $\mathbf{v}_r^t$, which is the considered surrogate function (22). ■

Maximizing $L\left(\{\mathbf{v}_r\} \,|\, \{\tau_k\}, \{c_r\}, \{\mathbf{v}_r\}^t\right)$, under orthonormality constraints on the $\{\mathbf{v}_r\}$, is equivalent to solving

$$\begin{aligned} \underset{\mathbf{V}}{\text{maximize}} \quad & \mathrm{Tr}\left(\mathbf{A}^H\mathbf{V}\right) + \mathrm{Tr}\left(\mathbf{V}^H\mathbf{A}\right) \\ \text{subject to} \quad & \mathbf{V}^H\mathbf{V} = \mathbf{I}, \end{aligned} \tag{23}$$

where

$$\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_R] \tag{24}$$

and

$$\mathbf{A} = \left[\mathbf{M}_1\mathbf{v}_1^t; \ldots; \mathbf{M}_R\mathbf{v}_R^t\right], \tag{25}$$

i.e., $\mathbf{A}$ is constructed by stacking the column vectors $\mathbf{M}_r\mathbf{v}_r^t$. Problem (23) is equivalent to:

$$\begin{aligned} \underset{\mathbf{V}}{\text{minimize}} \quad & \|\mathbf{A} - \mathbf{V}\|_F^2 \\ \text{subject to} \quad & \mathbf{V}^H\mathbf{V} = \mathbf{I}, \end{aligned} \tag{26}$$

which is the Procrustes problem [45]. Let the thin SVD of $\mathbf{A}$ be $\mathbf{A} = \mathbf{V}_A\mathbf{D}_A\mathbf{U}_A^H$, the optimal $\mathbf{V}$ is given by:

$$\mathbf{V}^{t+1} = \mathbf{V}_A\mathbf{U}_A^H. \tag{27}$$

Note that in this step, we maximize the objective function of (C3) by iteratively maximizing a series of surrogates (steps 7-11 in Algorithm 2). These iterations lead to a local solution of (C3). While the block MM framework indicates a cyclic update of the variables $\{\tau_k\}, \{c_r\}, \{\mathbf{v}_r\}$, our numerical tests reveal that including this inner loop provides a faster decreasing rate of the objective value than updating the $\mathbf{v}_r$'s only once.

D. Convergence Analysis

Since the constraint set of Problem (C) is non-convex, the convergence result provided in [44] cannot be applied here. To the best of our knowledge, there is no convergence result of general block descent type algorithms with a non-convex constraint set. The sequence of objective values generated by the Algorithm 2 will converge because of monotonicity, but the convergence of the points $\left(\{\tau_k\}^t, \{c_r\}^t, \{\mathbf{v}_r\}^t\right)$ remains unknown. Nevertheless, section V will show that the numerical performance of the Algorithm 2 is satisfactory.

V. Numerical Results

This section is devoted to numerical simulations to illustrate the performance of different CM estimators in the considered context. In particular, we will study the following clutter subspace projector estimators:

- $\hat{\mathbf{\Pi}}_{SCM}$: the clutter subspace projector estimator derived from the SVD of the SCM defined as $\hat{\mathbf{\Sigma}}_{SCM} = \sum_{k=1}^{K}\mathbf{z}_k\mathbf{z}_k^H/K$.
- $\hat{\mathbf{\Pi}}_{SFPE}$: the clutter subspace projector estimator derived from the SVD of the Shrinkage-FPE (SFPE) [31]–[33], which is a regularized Tyler's estimator [21] defined as the unique solution of the following fixed-point equation:

$$\hat{\mathbf{\Sigma}}_{S-FPE}(\beta) = \frac{(1-\beta)M}{K}\sum_{k=1}^{K}\frac{\mathbf{z}_k\mathbf{z}_k^H}{\mathbf{z}_k^H\hat{\mathbf{\Sigma}}_{S-FPE}^{-1}(\beta)\mathbf{z}_k} + \beta\mathbf{I},$$

  for $\beta \in (\ \max(0, 1-K/M)\ ,\ 1\ ]$. This estimator can be computed with simple fixed point iterations provided in [31]–[33]. Since there is no rule to adaptively select the optimal shrinkage parameter β for the considered problem, we test the following values: $\beta_1 = \max\left(1 - K/M + \epsilon\ ,\ 0\right)$, which is the lowest β allowed for the under-sampled cases and corresponds to Tyler's estimator (referred to as FPE) for the over-sampled cases, $\beta_2 = (\beta_1 + \beta_3)/2$ and $\beta_3 = 1 - \epsilon$. We set $\epsilon = 10^{-2}$.
- $\hat{\mathbf{\Pi}}_{MLE-MM1}$: the clutter subspace projector MLE computed with the direct block-MM Algorithm in Section III .
- $\hat{\mathbf{\Pi}}_{MLE-MM2}$: the clutter subspace projector MLE computed with the Eigenspace block-MM Algorithm in Section IV.
- $\hat{\mathbf{\Pi}}_{MLE}$: the clutter subspace projector MLE under high CNR assumption, computed with Algorithm 1 provided in reference [2]. This algorithm will be referred to as "Algorithm 3" in the rest of the paper.
- $\hat{\mathbf{\Pi}}_{A-MLE}$: the clutter subspace projector Approached MLE under high CNR assumption, computed with Algorithm 2 provided in reference [2]. This algorithm will be referred to as "Algorithm 4" in the rest of the paper.

A. Validation Simulations and Algorithm Complexity

Simulation parameters: Samples $\mathbf{z}_k$ are generated according to the LR Compound Gaussian plus WGN model described in section II: $\mathbf{z}_k = \mathbf{c}_k + \mathbf{n}_k$. The WGN is distributed as $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$ and $\sigma^2 = 1$. The LR Compound Gaussian clutter is distributed as $(\mathbf{c}_k|\tau_k) \sim \mathcal{CN}(\mathbf{0}, \tau_k\mathbf{\Sigma})$,

with a random texture τ_k, $i.i.d.$ generated for each sample. The texture PDF is a Gamma distribution (leading to a K-distributed clutter) of shape parameter ν and scale parameter $1/\nu$, denoted $\tau \sim \Gamma(\nu, 1/\nu)$, which satisfies $\mathbb{E}(\tau) = 1$. The rank R clutter CM $\mathbf{\Sigma}_c$ is constructed with the largest R eigenvalues and the corresponding eigenvectors of a Toeplitz matrix of correlation parameter $\rho \in [0, 1]$. This matrix is then scaled to set the CNR, defined as CNR $= \mathbb{E}(\tau)\text{Tr}(\mathbf{\Sigma})/(R\sigma^2)$, to a given value.

Fig.1 displays a typical realization of the objective value versus the number of iterations of different algorithms. The objective value at each inner loop is also displayed for the proposed algorithms. One can observe that MLE and A-MLE algorithms from [2] converge to a sub-optimal point of the problem. This was to be expected since these algorithms are optimizing a modified likelihood (assuming High CNR, the WGN is ignored over the clutter subspace). MLE Algorithm provides a slightly better objective value than A-MLE, thanks to the use of the modified gradient descent algorithm [40] instead of a SVD relaxation for updating the subspace estimate. Contrary to these algorithms, the Block MM algorithm (Algorithm 1) converges to a critical point with a smaller objective value. We also notice that Algorithm 2 converges in practice to the same point as Algorithm 1, *i.e.*, a critical point.

Fig.2 displays a typical realization of the objective value versus the time of computation. It illustrates that despite a fast convergence, Algorithm 3 has a slow computation due to the use of the modified gradient descent [40]. It also shows that Algorithm 2 is less computationally intensive than Algorithm 1 since it requires less time to converge. This can be explained by the fact that constructing the update of $\mathbf{W}$ (step 6) in Algorithm 1 has a complexity that grows linearly with the sample size. On the contrary, the inner loops (steps 7-11) in Algorithm 2 involve the SVD of a matrix of fixed dimension, which is not costly in comparison.

Fig.3 displays the mean NMSE criterion ($\mathbb{E}(||\hat{\mathbf{\Pi}}_c - \mathbf{\Pi}_c||^2)/R$ for a given estimator $\hat{\mathbf{\Pi}}_c$) versus K of the estimators $\hat{\mathbf{\Pi}}_{SCM}$, $\hat{\mathbf{\Pi}}_{MLE-MM1}$, $\hat{\mathbf{\Pi}}_{MLE-MM2}$, $\hat{\mathbf{\Pi}}_{MLE}$ and $\hat{\mathbf{\Pi}}_{A-MLE}$ for a given configuration (computed over 100 Monte-Carlo simulations). It illustrates the performance of the proposed methods: block MM Algorithms 1 and 2 reach identically the lowest NMSE and algorithms from [2] lead to slightly higher NMSE, yet better than the SCM and the state-of-the-art methods, as illustrated in [1], [2].

Fig.4 displays the mean computation time (over 100 Monte Carlo simulations) of the algorithms for the same set of parameters as in previous figures. The stop criterion for each algorithm is when of the minimum achievable objective value is reached up to 1%. The computation time of MLE Algorithms 3 is not displayed since this algorithm (relying on a costly gradient descent) always reaches a maximum time stop criterion, greatly larger than the presented values (as seen in Fig.2). One can observe that Algorithm 2 and 4 are the "fastest" algorithms , which is due to their low computational cost at each step. Algorithm 4 is the less computationally intensive since it has a fast convergence, as observed in Fig.1 and Fig.2. Nevertheless, Algorithm 1 and 2

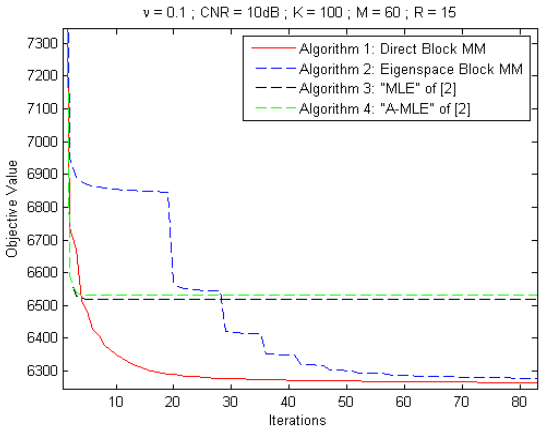

Figure 1. Objective value versus number of iterations for different MLE algorithms: Direct Block-MM (red), Eigenspace Block-MM (blue), "MLE" algorithm of [2] (black), "A-MLE" algorithm of [2] (green). $K = 100$, $R = 15$, $\nu = 0.1$, $\rho = 0.9$, CNR $= 10$dB

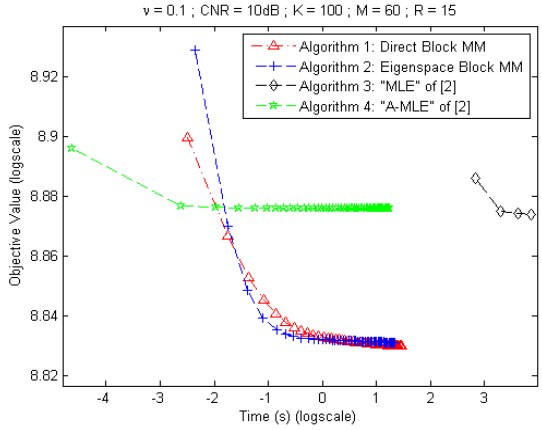

Figure 2. Objective value versus computation time for different MLE algorithms: Direct Block-MM (red), Eigenspace Block-MM (blue), "MLE" algorithm of [2] (black), "A-MLE" algorithm of [2] (green). $K = 100$, $R = 15$, $\nu = 0.1$, $\rho = 0.9$, CNR $= 10$dB

offer better performance (see Fig.3). We also point out that, though Algorithm 2 is computationally faster than Algorithm 1, the latter has more theoretical guarantees (convergence to a critical point). Both Fig.3 and Fig.4 illustrate the applicative interest of the two proposed computation methods. Indeed, they reach the best estimation performance for the considered model at a low computational cost.

B. Application to LR-STAP filtering

STAP is a technique used in airborne phased array radar to detect moving target embedded in an interference background such as jamming or strong clutter [42]. The radar receiver consists in an array of Q antenna elements processing P pulses in a coherent processing interval. The received signal is $\mathbf{z} = \alpha\mathbf{p} + \mathbf{c} + \mathbf{n}$, where α is the target power and $\mathbf{p}$ is the so-called STAP steering vector, $\mathbf{c}$ is the heterogeneous ground clutter and $\mathbf{n}$ is the thermal noise. It is important to notice that application fits the model considered in this paper

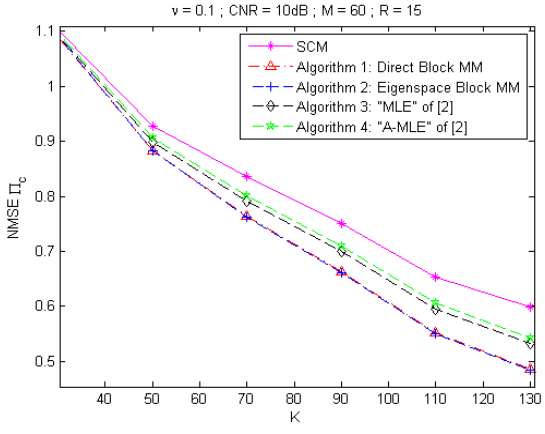

Figure 3. Mean NMSE of the estimators $\hat{\mathbf{\Pi}}_{SCM}$ (magenta), $\hat{\mathbf{\Pi}}_{MLE}$ (black), $\hat{\mathbf{\Pi}}_{A-MLE}$ (green), $\hat{\mathbf{\Pi}}_{MLE-MM1}$ (red), $\hat{\mathbf{\Pi}}_{MLE-MM2}$ (blue), versus the number of samples K. $M = 60$, $R = 15$, $\nu = 0.1$, $\rho = 0.9$, CNR $= 10$dB.

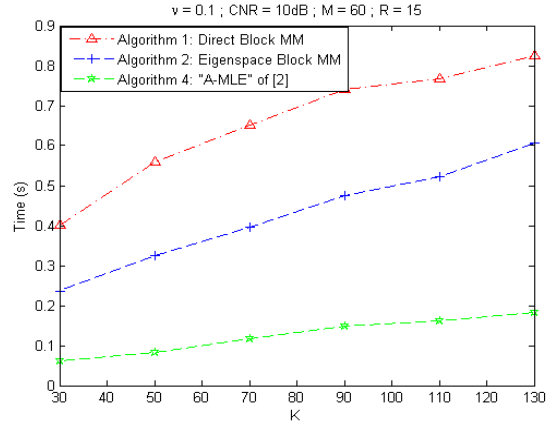

Figure 4. Mean Computation time of each MLE algorithm over 100 Monte-Carlo Simulation, versus K. Same parameters as Fig.3. Algorithms stop criterion: minimum achievable objective value is reached up to 1%. Stopping criterion for Algorithm 2 inner loop : $\mathbf{V}$: $\frac{\|\mathbf{V}^{t+1}-\mathbf{V}^{t}\|_F^2}{M\times R} \leq 10^{-6}$. Computer specifics: Intel(R) Core(TM) i5-3230M CPU @ 2.6 Ghz.

since in side looking STAP, the clutter CM is known to be LR. Moreover, the rank of the clutter CM can be evaluated thanks to the Brennan Rule [43]. Since the clutter could behave heterogeneously, the total interference can be modeled as LR Compound Gaussian (the ground clutter) plus WGN (the thermal noise).

The theoretical optimal STAP filter is $\mathbf{w}_{opt} = \mathbf{\Sigma}_{tot}^{-1}\mathbf{p}$ [42]. In the context of a LR clutter, it is well-known that a correct sub-optimal filter [8], [11] is $\mathbf{w}_{lr} = \mathbf{\Pi}_c^{\perp}\mathbf{p} = (\mathbf{I} - \mathbf{\Pi}_c)\mathbf{p}$. In practice, $\mathbf{\Pi}_c^{\perp}$ is unknown and has to be estimated using the samples $\{\mathbf{z}_k\}$ to perform adaptive filtering. The adaptive LR filter is then $\hat{\mathbf{w}}_{lr} = \hat{\mathbf{\Pi}}_c^{\perp}\mathbf{d} = \left(\mathbf{I} - \hat{\mathbf{\Pi}}_c\right)\mathbf{p}$, with $\hat{\mathbf{\Pi}}_c$ being an estimate of the clutter subspace projector. Consequently, the performance of the LR filters directly relies on the estimation accuracy of $\mathbf{\Pi}_c$. To evaluate the performance, we use the SINR-Loss criterion [42], which is the expected ratio between the $SINR_{out}$, computed for $\hat{\mathbf{w}}_{lr}$, and $SINR_{max}$ computed for the optimal filter $\mathbf{w} = \mathbf{\Sigma}_{tot}^{-1}\mathbf{d}$. For an estimate of the clutter subspace $\hat{\mathbf{\Pi}}_c$, the SINR-Loss expression is given by :

$$\rho_{\hat{\mathbf{\Pi}}_c} = \frac{SINR_{out}}{SINR_{max}} = \mathbb{E}\left(\frac{(\mathbf{p}^H\hat{\mathbf{\Pi}}_c^{\perp}\mathbf{p})^2}{\mathbf{p}^H\hat{\mathbf{\Pi}}_c^{\perp}\mathbf{\Sigma}_{tot}\hat{\mathbf{\Pi}}_c^{\perp}\mathbf{p}}\right). \quad (28)$$

We consider the following STAP configuration. The number Q of sensors is 8 and the number P of coherent pulses is also 8. The center frequency and the bandwidth are equal to $f_0 = 450$ MHz and $B = 4$ MHz, respectively. The radar velocity is 100 m/s. The inter-element spacing is $d = \frac{c}{2f_0}$ (c is the celerity of light) and the pulse repetition frequency is $f_r = 600$ Hz. The rank of the clutter CM $\mathbf{\Sigma}$ is computed from Brennan rule and is equal to $r = 15 \ll 64$, therefore, the LR assumption is valid. The texture PDF is a Gamma distribution of shape parameter $\nu = 0.1$ and scale parameter $1/\nu$, so the clutter follows a K-distribution. The target $\mathbf{p}$ has a celerity of $V = 35$ m/s and is at $+10°$ Azimuth. CNR is defined as CNR $= \mathbb{E}(\tau)\text{Tr}(\mathbf{\Sigma}_c)/(\sigma^2 R)$, and we set $\sigma^2 = 1$.

Since the proposed algorithms have been shown to reach identical performance previously, we only display the results for $\hat{\mathbf{\Pi}}_{MLE}$ computed with the fastest algorithm (Eigenspace Block MM) in this simulation. Fig.5 displays the SINR-Loss versus K for various clutter configurations: from average to high CNR (10dB, 20dB, and 30dB) and for mildly ($\nu = 1$) and highly ($\nu = 0.1$) heterogeneous clutter. First, one can state that the MLE reaches the best SINR-Loss for all configurations. Under standard conditions (20dB and 30dB of CNR and $\nu = 1$), the $\hat{\mathbf{\Pi}}_{SCM}$ provides a clutter subspace projector estimate that is close to the $\hat{\mathbf{\Pi}}_{MLE}$, hence they have similar performance. One can observe that under these conditions, the classical -3dB SINR-Loss of the filter build from $\hat{\mathbf{\Pi}}_{SCM}$ is reached close to $K \simeq 2R$ samples, as theoretically expected from [11], [19]. $\hat{\mathbf{\Pi}}_{FPE}$ also reaches performance close to $\hat{\mathbf{\Pi}}_{MLE}$ but requires $K > M$ samples to be computed. $\hat{\mathbf{\Pi}}_{SFPE}$ can be computed with a smaller K, but its SINR-Loss decreases as β increases (which is expected since a larger β implies a higher bias). When standard conditions are not met (low/average CNR or highly heterogeneous clutter), one can see that the performance of $\hat{\mathbf{\Pi}}_{FPE}$ and $\hat{\mathbf{\Pi}}_{SFPE}$ greatly drops compared to others estimators. On the contrary, $\hat{\mathbf{\Pi}}_{MLE}$ offers a better resistance to these conditions and also outperforms $\hat{\mathbf{\Pi}}_{SCM}$.

VI. Conclusions

In this paper, we derived two algorithms based on the block MM framework for computing the MLE of the CM parameter when the samples are modeled as the sum of a LR Compound Gaussian (with known rank) and a WGN. This complex problem was initially considered in [1], [2], where several algorithms were proposed. The new algorithms proposed in this paper enjoy two major advantages: firstly, they do not rely on any heuristic (as [1]) or high CNR assumption (as [2]). Thus they reach the exact MLE (at least locally) of the considered problem with no approximation; secondly, they are computationally faster and easier to implement than the ones derived in [1], [2], as they do not require the use of the modified gradient descent algorithm [40], which can

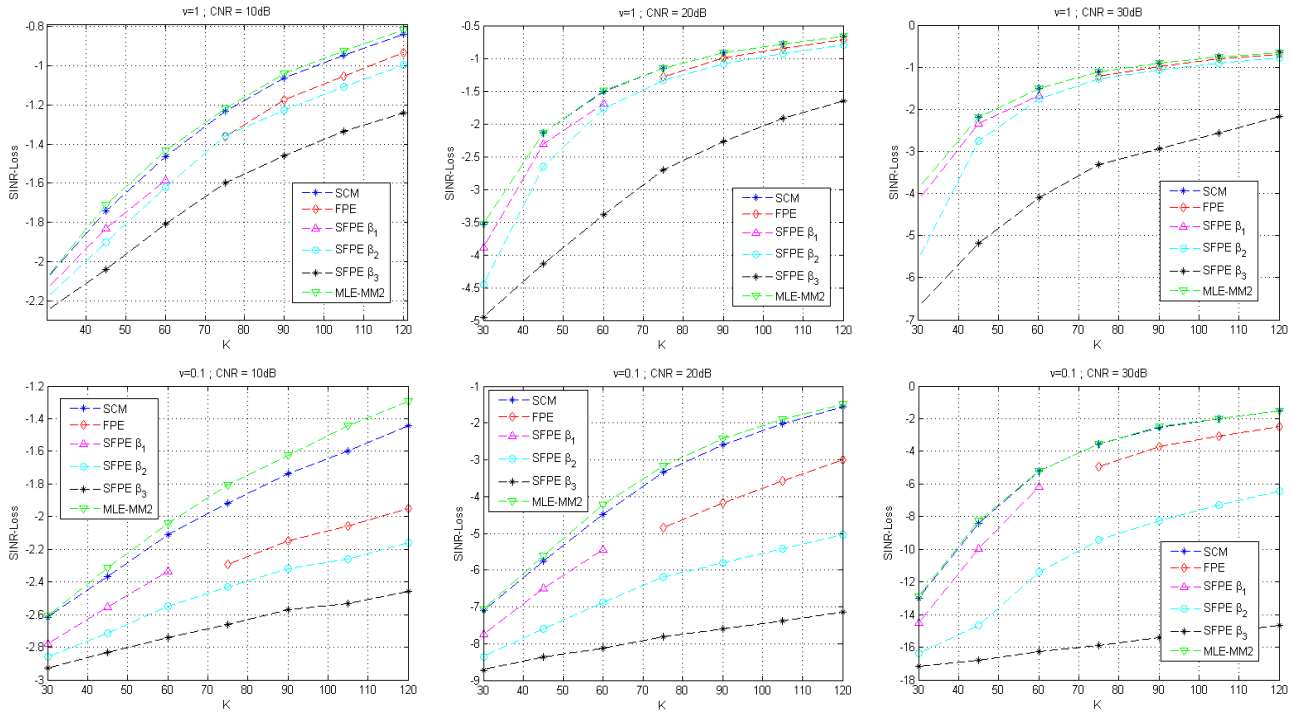

Figure 5. Mean SINR-Loss of the adaptive LR filters build from various clutter subspace projector estimators: $\hat{\mathbf{\Pi}}_{SCM}$ (blue), $\hat{\mathbf{\Pi}}_{FPE}$ (red), $\hat{\mathbf{\Pi}}_{SFPE}$ for β_1 (magenta), β_2 (cyan) and β_3 (black), $\hat{\mathbf{\Pi}}_{MLE}$ computed with Algorithm 2 (green). The SINR-Loss are presented for various clutter configurations. Columns from the left to the right: CNR = 10dB, CNR = 20dB, CNR = 30dB. Up row: $\nu = 1$ (mildly heterogeneous clutter), Down row: $\nu = 0.1$ (highly heterogeneous clutter).

be computationally expensive. The performance of the related estimators was illustrated on Space Time Adaptive Processing filtering for airborne radar simulations.

As side benefit of this study, we emphasize that the proposed algorithms allow a full estimation of the CM parameters, while presented simulations only focus on the clutter subspace projector estimation. They could therefore be suitable for other applications that involve the estimation of Low Rank structured Covariance Matrices.

Appendix A
Proof of Proposition 1

Before going to the formal proof of Proposition 1, we need to first state the two following lemmas:

Lemma 6 (Eq.(5) of [46]). *For $x_i > 0$ and $\alpha_i \in \mathbb{R}$, the function $\prod_{i=1}^{n} x_i^{\alpha_i}$ can be lower bounded by*

$$\prod_{i=1}^{n} x_i^{\alpha_i} \geq \prod_{j=1}^{n} \left(x_j^t\right)^{\alpha_j} \left(1 + \sum_{i=1}^{n} \alpha_i \log x_i - \sum_{i=1}^{n} \alpha_i \log x_i^t\right),$$

where x_i^t is some arbitrary positive real number. The equality is achieved at $x_i = x_i^t$. More specifically, in the uni-dimensional case $n = 1$ and for $\alpha_1 = 1$, one has:

$$x \geq x^t \left(1 + \log(x) - \log(x^t)\right) \tag{29}$$

Lemma 7. *The function $\sum_{i=1}^{n} \alpha_i \log(1 + c_i x)$ with $\alpha_i > 0$ can be upperbounded by*

$$\sum_{i=1}^{n} \alpha_i \log(1 + c_i x) \leq \sum_{i=1}^{n} \alpha_i \log\left(1 + c_i x^t\right) + \left(\sum_{i=1}^{n} \alpha_i\right) \log\left(\frac{\sum_{i=1}^{n} \alpha_i \cdot \frac{1+c_i x}{1+c_i x^t}}{\sum_{i=1}^{n} \alpha_i}\right), \tag{30}$$

where x^t is some arbitrary real number such that $1 + c_i x^t > 0$, $\forall i$. The equality is achieved at $x = x^t$

Proof: The $\log(\cdot)$ function is concave, and Jensen's inequality states that:

$$\log\left(\frac{\sum_{i=1}^{n} \alpha_i z_i}{\sum_{i=1}^{n} \alpha_i}\right) \geq \frac{\sum_{i=1}^{n} \alpha_i \log z_i}{\sum_{i=1}^{n} \alpha_i} \tag{31}$$

for $\alpha_i > 0$ and $z_i > 0$, and equality is achieved if the z_i's are equal. Let $z_i = \frac{1+c_i x}{1+c_i x^t}$, we have

$$\begin{aligned} &\frac{\sum_{i=1}^{n} \alpha_i \log\left(\frac{1+c_i x}{1+c_i x^t}\right)}{\sum_{i=1}^{n} \alpha_i} \\ =&\frac{\sum_{i=1}^{n} \alpha_i \log(1 + c_i x) - \sum_{i=1}^{n} \alpha_i \log\left(1 + c_i x^t\right)}{\sum_{i=1}^{n} \alpha_i} \\ \leq &\log\left(\frac{\sum_{i=1}^{n} \alpha_i \cdot \frac{1+c_i x}{1+c_i x^t}}{\sum_{i=1}^{n} \alpha_i}\right). \end{aligned} \tag{32}$$

Rearranging the terms leads to the inequality (30), with equality achieved at $x = x^t$, which corresponds to $z_i = 1$, $\forall i = 1, \ldots, n$. ■

We can now turn to the proof of Proposition 1:

Proof: Ignoring the constant term, $L(\tau_k|\mathbf{W})$ can be written as:

$$\begin{aligned} & L(\tau_k|\mathbf{W}) \\ &= \sum_{m=1}^{R} \log(\tau_k\lambda_m + 1) + \sum_{m=1}^{R} s_{km}\left(1 - \frac{\tau_k\lambda_m}{\tau_k\lambda_m + 1}\right) \\ &= \sum_{m=1}^{R} \log(\tau_k\lambda_m + 1) - \sum_{m=1}^{R} s_{km}\frac{\tau_k\lambda_m}{\tau_k\lambda_m + 1} + \text{const.} \end{aligned} \tag{33}$$

First, applying Eq.(29) of Lemma 6 to the second term of (33) (with x parametrized as $\frac{\tau_k\lambda_m}{\tau_k\lambda_m+1}$) yields:

$$\begin{aligned} & L\left(\tau_k|\mathbf{W}^t\right) \\ &\leq \sum_{m=1}^{R} \log(\tau_k\lambda_m + 1) + \text{const.} \\ &\quad - \sum_{m=1}^{R} s_{km}\frac{\tau_k^t\lambda_m}{\tau_k^t\lambda_m + 1}\left(1 + \log\tau_k\lambda_m - \log(\tau_k\lambda_m + 1)\right) \\ &\leq \sum_{m=1}^{R}\left(s_{km}\frac{\tau_k^t\lambda_m}{\tau_k^t\lambda_m + 1} + 1\right)\log(1 + \lambda_m\tau_k) \\ &\quad - \left(\sum_{m=1}^{R} s_{km}\frac{\tau_k^t\lambda_m}{\tau_k^t\lambda_m + 1}\right)\log\tau_k + \text{const.} \end{aligned} \tag{34}$$

where the equality is achieved at $\tau_k = \tau_k^t$. This upperbound is convex in $\tilde{\tau}_k \triangleq \log\tau_k$, but the minimizer cannot be computed in closed-form. Denote:

$$\alpha_{km} = s_{km}\frac{\tau_k^t\lambda_m}{\tau_k^t\lambda_m + 1} + 1 \quad \text{and} \quad \beta_k = \sum_{m=1}^{R} s_{km}\frac{\tau_k^t\lambda_m}{\tau_k^t\lambda_m + 1}\ .$$

We now apply Eq.(30) of Lemma 7 to the first term of the upperbound in (34), which leads to:

$$\begin{aligned} & L(\tau_k|\mathbf{W}) \\ &\leq \sum_{m=1}^{R} \alpha_{km}\log(1 + \lambda_m\tau_k) - \beta_k\log\tau_k + \text{const.} \\ &\leq \left(\sum_{m=1}^{R}\alpha_{km}\right)\log\left(\frac{\sum_{m=1}^{R}\alpha_{km}\frac{1+\lambda_m\tau_k}{1+\lambda_m\tau_k^t}}{\sum_{m=1}^{R}\alpha_{km}}\right) - \beta_k\log\tau_k \\ &\quad + \text{const.} \\ &\leq \left(\sum_{m=1}^{R}\alpha_{km}\right)\log\left(\frac{\left(\sum_{m=1}^{R}\frac{\alpha_{km}\lambda_m}{1+\lambda_m\tau_k^t}\right)}{\sum_{m=1}^{R}\alpha_{km}}\tau_k + \frac{\sum_{m=1}^{R}\frac{\alpha_{km}}{1+\lambda_m\tau_k^t}}{\sum_{m=1}^{R}\alpha_{km}}\right) \\ &\quad - \beta_k\log\tau_k + \text{const.} \end{aligned} \tag{35}$$

Ignoring the constant term, we arrive at the surrogate function $L(\tau_k|\tau_k^t, \mathbf{W})$ defined in Proposition 1. Note that the equality still holds at $\tau_k = \tau_k^t$. ■

APPENDIX B
PROOF OF PROPOSITION 2

Proof: The surrogate function, ignoring the constant term, has the form:

$$L\left(\tau_k|\tau_k^t, \mathbf{W}^t\right) \triangleq a\log(b\tau_k + c) - \beta_k\log\tau_k.$$

The gradient of this surrogate function is

$$\frac{\partial L(\tau_k|\tau_k^t, \mathbf{W})}{\partial\tau_k} = \frac{ab}{b\tau_k + c} - \beta_k\frac{1}{\tau_k},$$

and a zero of the gradient can be solved in closed-form as:

$$\tau_k^{t+1} = \frac{\beta_k c}{(a - \beta_k)\, b}.$$

Since

$$a = \sum_{m=1}^{R}\alpha_{km} = \sum_{m=1}^{R}\left(s_{km}\frac{\tau_k^t\lambda_m}{\tau_k^t\lambda_m + 1} + 1\right)$$

and

$$\beta_k = \sum_{m=1}^{R} s_{km}\frac{\tau_k^t\lambda_m}{\tau_k^t\lambda_m + 1},$$

we have $a - \beta_k = R$. Therefore,

$$\tau_k^{t+1} = \frac{1}{R}\cdot\frac{\left(\sum_{m=1}^{R} s_{km}\frac{\tau_k^t\lambda_m}{\tau_k^t\lambda_m+1}\right)\cdot\left(\sum_{m=1}^{R}\frac{\alpha_{km}}{1+\lambda_m\tau_k^t}\right)}{\sum_{m=1}^{R}\frac{\alpha_{km}\lambda_m}{1+\lambda_m\tau_k^t}} \geq 0.$$

It remains to be shown that τ_k^{t+1} is the unique minimizer of $L(\tau_k|\tau_k^t, \mathbf{W}^t)$ on $\mathbb{R}_+$.

First assume that $\beta_k \neq 0$. Since c is positive, $L(\tau_k|\tau_k^t, \mathbf{W}^t) \to +\infty$ as $\tau_k \to 0$. Besides, as $a > \beta_k$, $L(\tau_k|\tau_k^t, \mathbf{W}^t) \to +\infty$ as $\tau_k \to +\infty$. Since $L(\tau_k|\tau_k^t, \mathbf{W}^t)$ is finite, τ_k^{t+1} is the unique minimizer by the continuity of $L(\tau_k|\tau_k^t, \mathbf{W}^t)$. If $\beta_k = 0$, then $\tau_k^{t+1} = 0$. Function $a\log(b\tau_k + c)$ is monotonically increasing on $\mathbb{R}_+$, therefore $\tau_k^{t+1} = 0$ is the unique minimizer. Moreover, as $L(\tau_k|\tau_k^t, \mathbf{W}^t)$ is a one-dimensional function, it has to be quasi-convex [47]. ■

APPENDIX C
PROOF OF PROPOSITION 3

Proof: By the matrix inversion lemma we have the following equality:

$$\begin{aligned} (\tau_k\boldsymbol{\Sigma} + \mathbf{I})^{-1} &= \mathbf{I} - \mathbf{W}\left(\tau_k^{-1}\mathbf{I} + \mathbf{W}^H\mathbf{W}\right)^{-1}\mathbf{W}^H \\ &= \mathbf{I} - \mathbf{W}\left(\tau_k^{-1}\mathbf{I} + \tilde{\boldsymbol{\Sigma}}\right)^{-1}\mathbf{W}^H. \end{aligned} \tag{36}$$

where $\tilde{\boldsymbol{\Sigma}} \triangleq \mathbf{W}^H\mathbf{W}$. Therefore, the second term of $L(\mathbf{W}|\tau_k)$ can be written as

$$\begin{aligned} & \mathbf{z}_k^H\left(\tau_k\boldsymbol{\Sigma} + \mathbf{I}\right)^{-1}\mathbf{z}_k \\ =& \mathbf{z}_k^H\mathbf{z}_k - \mathbf{z}_k^H\mathbf{W}\left(\tau_k^{-1}\mathbf{I} + \tilde{\boldsymbol{\Sigma}}\right)^{-1}\mathbf{W}^H\mathbf{z}_k, \end{aligned} \tag{37}$$

which is jointly concave in $\left\{\mathbf{W}, \tilde{\boldsymbol{\Sigma}}\right\}$.

The concavity of (37) implies that it can be upperbounded by its first order Taylor expansion around the pair $\left\{\mathbf{W}^t, \tilde{\mathbf{\Sigma}}^t\right\}$, i.e.,

$$\begin{aligned}
&\mathbf{z}_k^H \left(\tau_k \mathbf{\Sigma} + \mathbf{I}\right)^{-1} \mathbf{z}_k \\
\leq &- \mathrm{Tr}\left(\mathbf{z}_k \mathbf{z}_k^H \mathbf{W}^t \left(\tau_k^{-1}\mathbf{I} + \tilde{\mathbf{\Sigma}}^t\right)^{-1} \mathbf{W}^H\right) \\
&- \mathrm{Tr}\left(\left(\tau_k^{-1}\mathbf{I} + \tilde{\mathbf{\Sigma}}^t\right)^{-1} \left(\mathbf{W}^t\right)^H \mathbf{z}_k \mathbf{z}_k^H \mathbf{W}\right) \\
&+ \mathrm{Tr}\left(\left(\tau_k^{-1}\mathbf{I} + \tilde{\mathbf{\Sigma}}^t\right)^{-1} \left(\mathbf{W}^t\right)^H \mathbf{z}_k \mathbf{z}_k^H \mathbf{W}^t \left(\tau_k^{-1}\mathbf{I} + \tilde{\mathbf{\Sigma}}^t\right)^{-1} \tilde{\mathbf{\Sigma}}\right) \\
&+ \mathrm{const.}
\end{aligned} \tag{38}$$

Substituting $\tilde{\mathbf{\Sigma}} = \mathbf{W}^H\mathbf{W}$ into (38) leads to the following quadratic upperbound for $\mathbf{z}_k^H \left(\tau_k \mathbf{\Sigma} + \mathbf{I}\right)^{-1} \mathbf{z}_k$:

$$\mathrm{Tr}\left(\mathbf{W}\mathbf{H}_k^t\mathbf{W}^H\right) - \mathrm{Tr}\left(\mathbf{L}_k^t\mathbf{W}^H\right) - \mathrm{Tr}\left(\left(\mathbf{L}_k^t\right)^H \mathbf{W}\right), \tag{39}$$

where

$$\mathbf{H}_k^t = \left(\tau_k^{-1}\mathbf{I} + \tilde{\mathbf{\Sigma}}^t\right)^{-1} \left(\mathbf{W}^t\right)^H \mathbf{z}_k \mathbf{z}_k^H \mathbf{W}^t \left(\tau_k^{-1}\mathbf{I} + \tilde{\mathbf{\Sigma}}^t\right)^{-1} \tag{40}$$

and

$$\mathbf{L}_k^t = \mathbf{z}_k \mathbf{z}_k^H \mathbf{W}^t \left(\tau_k^{-1}\mathbf{I} + \tilde{\mathbf{\Sigma}}^t\right)^{-1}. \tag{41}$$

As for the $\log\det\left(\cdot\right)$ term in $L\left(\mathbf{W}|\tau_k\right)$, we can apply the Sylvester's determinant theorem

$$\log\det\left(\tau_k \mathbf{W}\mathbf{W}^H + \mathbf{I}\right) = \log\det\left(\tau_k \mathbf{W}^H\mathbf{W} + \mathbf{I}\right) \tag{42}$$

and get the upperbound

$$\mathrm{Tr}\left(\tau_k \left(\tau_k \tilde{\mathbf{\Sigma}}^t + \mathbf{I}\right)^{-1} \mathbf{W}^H\mathbf{W}\right) + \mathrm{const.} \tag{43}$$

by linearization over $\tilde{\mathbf{\Sigma}}$.

Together with (39) we have an upperbound (ignoring the constant terms) for $L\left(\mathbf{W}|\tau_k\right)$ at $\mathbf{W}^t$ as

$$\begin{aligned}
&\mathrm{Tr}\left(\mathbf{W}\left(\sum_{k=1}^K \left(\tilde{\mathbf{\Sigma}}^t + \tau_k^{-1}\mathbf{I}\right)^{-1}\right)\mathbf{W}^H\right) \\
&+ \sum_{k=1}^K \left(\mathrm{Tr}\left(\mathbf{W}\mathbf{H}_k^t\mathbf{W}^H\right) - \mathrm{Tr}\left(\mathbf{L}_k^t\mathbf{W}^H\right) - \mathrm{Tr}\left(\left(\mathbf{L}_k^t\right)^H\mathbf{W}\right)\right) \\
=&\mathrm{Tr}\left(\mathbf{W}\mathbf{H}\mathbf{W}^H\right) - \mathrm{Tr}\left(\mathbf{L}\mathbf{W}^H\right) - \mathrm{Tr}\left(\mathbf{L}^H\mathbf{W}\right),
\end{aligned} \tag{44}$$

where

$$\mathbf{H} = \sum_{k=1}^K \left(\left(\tilde{\mathbf{\Sigma}}^t + \tau_k^{-1}\mathbf{I}\right)^{-1} + \mathbf{H}_k^t\right) \tag{45}$$

and

$$\mathbf{L} = \sum_{k=1}^K \mathbf{L}_k^t. \tag{46}$$

■

REFERENCES

[1] A. Breloy, G. Ginolhac, F. Pascal, and P. Forster, "Clutter subspace estimation in low rank heterogeneous noise context," *IEEE Trans. Signal Process.*, vol. 63, no. 9, pp. 2173–2182, May 2015.

[2] ——, "Robust covariance matrix estimation in heterogeneous low-rank context," *IEEE Trans. Signal Process., submitted*.

[3] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 3, pp. 276–280, March 1986.

[4] R. Roy and T. Kailath, "ESPRIT-Estimation of signal parameters via rotational invariant techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, July 1989.

[5] F. Robey, D. Fuhrmann, E. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 2, pp. 208 – 216, 1992.

[6] S. Kraut, L. Scharf, and L. McWhorter, "Adaptive subspace detectors," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 1–16, January 2001.

[7] S. Haykin, *Adaptive filter theory*, ser. Prentice-Hall information and system sciences series. Prentice Hall, 2002.

[8] I. Kirsteins and D. Tufts, "Adaptive detection using a low rank approximation to a data matrix," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 30, pp. 55 – 67, 1994.

[9] M. Rangaswamy, F. Lin, and K. Gerlach, "Robust adaptive signal processing methods for heterogeneous radar clutter scenarios," *Signal Processing*, vol. 84, pp. 1653 – 1665, 2004.

[10] I. Reed, J. Mallett, and L. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-10, no. 6, pp. 853 – 863, November 1974.

[11] A. Haimovich, "Asymptotic distribution of the conditional signal-to-noise ratio in an eigenanalysis-based adaptive array," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, pp. 988 – 997, 1997.

[12] E. Ollila, D. Tyler, V. Koivunen, and H. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5597–5625, 2012.

[13] J. Billingsley, "Ground clutter measurements for surface-sited radar," MIT, Tech. Rep. 780, February 1993.

[14] E. Conte, A. De Maio, and C. Galdi, "Statistical analysis of real clutter at different range resolutions," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 40, no. 3, pp. 903–918, July 2004.

[15] M. Greco, F. Gini, and M. Rangaswamy, "Statistical analysis of measured polarimetric clutter data at different range resolutions," *Radar, Sonar and Navigation, IEE Proceedings* -, vol. 153, no. 6, pp. 473–481, December 2006.

[16] E. Ollila, D. Tyler, V. Koivunen, and H. Poor, "Compound-gaussian clutter modeling with an inverse Gaussian texture distribution," *IEEE Signal Process. Lett.*, vol. 19, no. 12, pp. 876–879, Dec 2012.

[17] M. Rangaswamy, I. Kirsteins, B. Freburger, and D. Tufts, "Signal detection in strong low rank Compound-Gaussian interference," in *Proc. IEEE 2000 Sensor Array and Multichannel Signal Processing Workshop (SAM)*, 2000, pp. 144–148.

[18] R. Raghavan, "Statistical interpretation of a data adaptive clutter subspace estimation algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 2, pp. 1370 – 1384, 2012.

[19] G. Ginolhac, P. Forster, F. Pascal, and J.-P. Ovarlez, "Performance of two low-rank STAP filters in a heterogeneous noise," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 57–61, 2013.

[20] R. A. Maronna, "Robust M-estimators of multivariate location and scatter," *Ann. Statist.*, vol. 4, no. 1, pp. 51–67, January 1976.

[21] D. Tyler, "A distribution-free M-estimator of multivariate scatter," *Ann. Statist.*, vol. 15, no. 1, pp. 234–251, 1987.

[22] F. Pascal, Y. Chitour, J. Ovarlez, P. Forster, and P. Larzabal, "Existence and characterization of the covariance matrix maximum likelihood estimate in spherically invariant random processes," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 34 – 48, January 2008.

[23] E. Conte, A. D. Maio, and G. Ricci, "Recursive estimation of the covariance matrix of a compound-Gaussian process and its application to adaptive CFAR detection," *IEEE Trans. Signal Process.*, vol. 50, no. 8, pp. 1908 – 1915, August 2002.

[24] F. Gini and A. Farina, "Vector subspace detection in compound-Gaussian clutter. Part I: survey and new results," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 4, pp. 1295–1311, 2002.

[25] F. Gini, A. Farina, and M. Montanari, "Vector subspace detection in compound-Gaussian clutter. Part II: performance analysis," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 4, pp. 1312–1323, 2002.

[26] F. Gini and M. Greco, "Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter," *Signal Processing, special section on SP with Heavy Tailed Distributions*, vol. 82, no. 12, pp. 1847–1859, December 2002.

[27] E. Ollila and D. Tyler, "Distribution-free detection under complex elliptically symmetric clutter distribution," in *Proc. IEEE 7th Sensor Array and Multichannel Signal Process. Workshop (SAM)*, June 2012, pp. 413–416.

[28] O. Besson and Y. Abramovich, "Adaptive detection in elliptically distributed noise and under-sampled scenario," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1531–1535, Dec 2014.

[29] M. Greco, S. Fortunati, and F. Gini, "Maximum likelihood covariance matrix estimation for complex elliptically symmetric distributions under mismatched condition," *Signal Processing Elsevier*, vol. 104, pp. 381–386, Nov 2014.

[30] Y. Chen, A. Wiesel, and A. O. Hero, "Robust shrinkage estimation of high-dimensional covariance matrices," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4097–4107, 2011.

[31] F. Pascal, Y. Chitour, and Y. Quek, "Generalized robust shrinkage estimator and its application to STAP detection problem," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5640–5651, 2014.

[32] Y. Sun, P. Babu, and D. Palomar, "Regularized Tyler's scatter estimator: existence, uniqueness, and algorithms," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 5143–5156, Oct 2014.

[33] E. Ollila and D. Tyler, "Regularized M-estimators of scatter matrix," *IEEE Trans. Signal Process.*, vol. 62, no. 22, pp. 6059–6070, Nov 2014.

[34] A. Wiesel, "Unified framework to regularized covariance estimation in scaled Gaussian models," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 29–38, 2012.

[35] ——, "Geodesic convexity and covariance estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6182–6189, Dec 2012.

[36] I. Soloveychik and A. Wiesel, "Group symmetry and non-Gaussian covariance estimation," in *Global Conference on Signal and Information Processing (GlobalSIP), 2013 IEEE*, Dec 2013, pp. 1105–1108.

[37] ——, "Tyler's covariance matrix estimator in elliptical models with convex structure," *IEEE Trans. Signal Process.*, vol. 62, no. 20, pp. 5251–5259, Oct 2014.

[38] Y. Sun, P. Babu, and D. Palomar, "Robust estimation of structured covariance matrix for heavy-tailed elliptical distributions," *IEEE Trans. Signal Process., submited*.

[39] A. Breloy, L. L. Magoarou, G. Ginolhac, F. Pascal, and P. Forster, "Maximum likelihood estimation of clutter subspace in non homogeneous noise context," in *Proceedings of EUSIPCO*, Marrakech, Morocco, September 2013.

[40] J. Manton, "Optimization algorithms exploiting unitary constraints," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 635–650, Mar 2002.

[41] P. Perry and P. Wolfe, "Minimax rank estimation for subspace tracking," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 3, pp. 504–513, June 2010.

[42] J. Ward, "Space-time adaptive processing for airborne radar," Lincoln Lab., MIT, Lexington, Mass., USA, Tech. Rep., December 1994.

[43] L. E. Brennan and F. Staudaher, "Subclutter visibility demonstration," RL-TR-92-21, Adaptive Sensors Incorporated, Tech. Rep., March 1992.

[44] M. Razaviyayn, M. Hong, and Z. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 1126–1153, 2013.

[45] P. H. Schönemann, "A generalized solution of the orthogonal Procrustes problem," *Psychometrika*, vol. 31, no. 1, pp. 1–10, 1966.

[46] K. Lange and H. Zhou, "MM algorithms for geometric and signomial programming," *Mathematical programming*, vol. 143, no. 1-2, pp. 339–356, 2014.

[47] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge university press, 2004.

Chapitre 3

Estimation de projecteur sur le sous-espace fouillis en contexte hétérogène rang faible

Ce chapitre porte sur l'estimation de sous-espace fouillis dans le contexte d'un fouillis hétérogène de rang faible plus un bruit blanc gaussien (modèle décrit au chapitre 2). Dans un premier temps, nous présentons l'approximation rang faible et son intérêt pratique, justifiant que l'on se focalise sur l'estimation de sous-espace seul. Nous dérivons ensuite deux nouveaux estimateurs du projecteur sur le sous-espace fouillis pour le contexte considéré. Le premier découle d'une heuristique développée en relaxant les contraintes d'orthogonalité entre le sous-espace fouillis et son complémentaire. Le second résulte d'une approximation négligeant l'influence des valeurs propres et permet d'obtenir un EMV approché du projecteur sur le sous-espace fouillis (AEMV). Enfin, suite à une interprétation sur l'expression de l'AEMV, nous mettons en avant le problème d'estimation robuste de sous-espace et soulignons sa différence avec l'estimation robuste de la matrice de covariance. Ces discussions conduisent à proposer un dernier estimateur modifié, répondant à cette problématique.

3.1 Motivations

3.1.1 L'approximation rang faible et ses motivations

Considérons une fois de plus le cas d'un fouillis plus un bruit blanc gaussien indépendant de variance unitaire. La matrice de covariance totale du bruit est :

$$\mathbf{\Sigma}_{tot} = \mathbf{\Sigma}_c + \mathbf{I}_M = \sum_{r=1}^{R} (c_r + 1)\, \mathbf{v}_r \mathbf{v}_r^H + \sum_{r=R+1}^{M} \mathbf{v}_r \mathbf{v}_r^H \tag{3.1}$$

Lorsque le fouillis est de rang faible $R \ll M$, l'approximation rang faible consiste en :

$$\mathbf{\Sigma}_{tot}^{-1} \approx \mathbf{\Pi}_c^{\perp} \approx \mathbf{I}_M - \mathbf{\Pi}_c \tag{3.2}$$

où $\mathbf{\Pi}_c = \sum_{r=1}^{R} \mathbf{v}_r \mathbf{v}_r^H$ est le projecteur sur le sous-espace fouillis. Les traitements classiques de TS (comme les filtres ou détecteurs) requièrent l'inverse de la matrice de covariance afin de blanchir le bruit

corrompant l'échantillon reçu. Recourir à l'approximation rang faible s'apparente donc, conceptuellement, à de l'annulation d'interférence plutôt qu'à du blanchiment. Cette approche s'avère donc utile en contexte de fort rapport fouillis à bruit (car mathématiquement vérifiée si $1 \ll c_r$ et si la variance du bruit blanc est supposée unitaire), ainsi que pour perturbation fortement hétérogènes (ou le blanchiment n'est pas toujours approprié).

L'approximation rang faible offre de plus un avantage dans le cadre adaptatif. En effet, les traitements adaptatifs rang faible reposent sur un estimateur du sous-espace fouillis et non de la covariance totale. L'approche est asymptotiquement sous-optimale, cependant elle permet d'atteindre des performances équivalentes au traitements classiques avec moins d'échantillons. Par exemple, pour un fouillis gaussien, le filtre adaptatif optimal est celui construit à partir de la SCM. Dans ce cas, $K = 2M$ échantillons sont nécessaires afin d'atteindre des performances satisfaisantes (-3dB de perte en rapport signal à bruit de sortie par rapport au filtre optimal non adaptatif) [99]. En comparaison, le filtre adaptatif rang faible, construit avec le projecteur sur les R vecteurs propres dominants de la SCM, atteint les mêmes performances pour $K = 2R \ll 2M$ échantillons [54].

Les traitements rang faible ont été suggérés pour des applications radar et sonar [66, 54], où il est connu que le fouillis réside dans un sous-espace de dimension réduite, notamment en STAP [95, 96] où le rang du fouillis peut être évalué grâce à la formule de Brennan [18]. Leur utilisation est aussi utile pour contrer l'effet de brouilleurs ou d'interférences.

Néanmoins, les précédentes références considèrent des projecteurs dérivés à partir d'estimateurs de la matrice de covariance, ce qui ne permet pas de tirer partie du faible nombre d'échantillons nécessaires, notamment dans le cas des estimateurs robustes puisqu'il faut $K > M$ pour les obtenir (cf. section 1.3.5). Dans ce chapitre, nous allons développer des estimateurs directs de ce sous-espace, adaptés au contexte considéré. On peut dégager deux intérêts principaux de ces travaux :

- En se focalisant uniquement sur le sous-espace, on propose des algorithmes plus simples et compacts que ceux dérivés dans le chapitre précédent, ayant néanmoins des performances comparables.
- Leur analyse apporte une interprétation permettant de considérer et de proposer une solution à un problème plus complexe qu'est la robustesse à la contamination des données.

3.2 Relaxation sur l'orthogonalité entre sous-espaces : l'heuristique LR-FPE

Dans ces travaux, nous avons développé un algorithme basé sur une approche 2-Step relaxée, présentant une structure et des propriétés intéressantes. Nous considérons le modèle relaxé présenté à la section 2.5.1 (fort rapport fouillis à bruit) et dérivons un algorithme 2-Step en relâchant la contrainte d'orthogonalité entre le sous-espace fouillis et son complémentaire. Soit $(\mathbf{z}_k|\tau_k) \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_k)$ avec

$$\mathbf{\Sigma}_k \approx \left[\mathbf{V}_c \;\; \mathbf{V}_c^{\perp} \right] \begin{pmatrix} \tau_k \mathbf{C}_c & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{M-R} \end{pmatrix} \left[\mathbf{V}_c \;\; \mathbf{V}_c^{\perp} \right]^H , \tag{3.3}$$

conduisant à la vraisemblance :

$$f(\{\mathbf{z}_k\}) = \sum_{k=1}^{K}\sum_{r=1}^{R} \frac{1}{c_r \tau_k} \mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k + \sum_{k=1}^{K}\sum_{r=R+1}^{M} \mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k + R\ln(\tau_k) + \sum_{r=1}^{R} \ln(c_r) \tag{3.4}$$

Dans la section 2.5.4, l'orthogonalité entre le sous-espace fouillis et le sous-espace bruit blanc est imposée en utilisant le changement de variable $\sum_{r=R+1}^{M} \mathbf{v}_r \mathbf{v}_r^H = \mathbf{I} - \sum_{r=1}^{R} \mathbf{v}_r \mathbf{v}_r^H$. Si l'on n'impose pas cette contrainte, les paramètres $\{\mathbf{v}_r\}_{r\in[\![R+1,M]\!]}$ sont considérés indépendants des paramètres d'intérêt. La vraisemblance devient alors :

$$f(\{\mathbf{z}_k\} = \sum_{k=1}^{K}\sum_{r=1}^{R} \frac{1}{c_r \tau_k} \mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k + R\ln(\tau_k) + \sum_{r=1}^{R} \ln(c_r) + \gamma\,. \tag{3.5}$$

Afin de maximiser cette vraisemblance, nous proposons l'algorithme itératif suivant, basé sur une approche 2-Step :

– Étape 1 : la vraisemblance est maximisée selon $\{\tau_k\}$ pour $\{c_r\}$ et $\{\mathbf{v}_r\}$ fixés. L'EMV des $\{\tau_k\}$ est obtenu par :

$$\hat{\tau}_k = \frac{\sum_{r=1}^{R} \mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k / c_r}{R} \tag{3.6}$$

En notant $\mathbf{\Sigma}_c = \sum_{r=1}^{R} c_r \mathbf{v}_r \mathbf{v}_r^H$ et la pseudo-inverse $\mathbf{\Sigma}_c^{\dagger} = \sum_{r=1}^{R} 1/c_r \mathbf{v}_r \mathbf{v}_r^H$, ceci conduit à :

$$\hat{\tau}_k = \mathbf{z}_k^H \mathbf{\Sigma}_c^{\dagger} \mathbf{v}_r / R\,. \tag{3.7}$$

– Étape 2 : la vraisemblance est maximisée selon $\{c_r\}$ et $\{\mathbf{v}_r\}$ pour $\{\tau_k\}$ fixé. L'EMV de ces paramètres sont directement identifiés dans (3.5) comme les R vecteurs et valeurs propres dominants de la SCM des données pondérées $\frac{1}{K}\sum_{k=1}^{K} \frac{\mathbf{z}_k \mathbf{z}_k^H}{\tau_k}$.

L'alternance de ces deux étapes peut enfin se résumer en une seule opération :

$$\mathbf{\Sigma}_{(n+1)} = \frac{R}{K} \sum_{k=1}^{K} \frac{\mathbf{z}_k \mathbf{z}_k^H}{\mathbf{z}_k^H \mathbf{\Sigma}_{(n)}^{\dagger_R} \mathbf{z}_k} \tag{3.8}$$

où ^{†_R} définit la pseudo-inverse de rang R (ou pseudo-inverse tronquée au rang R). On remarque que ces itérations correspondent à l'algorithme du point fixe, mais utilisant une pseudo-inverse plutôt que l'inverse totale de la matrice, ce qui semble approprié dans un contexte de fouillis rang faible. On dérive ensuite l'estimateur de la base du sous-espace fouillis comme les R vecteurs propres dominants de la matrice obtenue après itérations.

Notons que les preuves de convergence de l'algorithme du point fixe [89, 84] ne se généralisent pas directement à l'algorithme proposé. L'existence et l'unicité de la solution considérée reste donc une question ouverte. Néanmoins nous proposons cet estimateur car il présente une formulation et des propriétés de robustesse à la contamination de données intéressantes.

3.3 Relaxation sur les valeurs propres : estimateur AEMV

Dans cette section, nous présentons des estimateurs de sous-espace fouillis dérivés sous l'hypothèse que les valeurs propres de la matrice de covariance du fouillis $\mathbf{\Sigma}_c$ soient égales. Dans ce cas, elle peut être directement définie comme une matrice de projection sur le sous-espace recherché $\mathbf{\Sigma}_c = \mathbf{\Pi}_c = \sum_{r=1}^{M} \mathbf{v}_r \mathbf{v}_r^H$. On vérifiera en effet que les valeurs propres ont peu d'influence sur l'estimation du sous-espace fouillis, et que s'en affranchir permet d'obtenir des estimateurs ayant une forme simple à calculer et manipuler.

3.3.1 Densité de probabilité de textures connue

Dans le travail séminal sur ce sujet, [94] dérive le maximum de vraisemblance du projecteur sur le sous-espace fouillis pour un fouillis CG plus un bruit blanc gaussien. Pour ce faire, [94] suppose que la matrice de covariance du fouillis est un projecteur et suppose de plus la densité de probabilité des textures τ_k, notée f_τ connue. La log-vraisemblance de l'ensemble des données s'exprime alors :

$$\ln(f(\{\ \mathbf{z}_k\}|\mathbf{\Pi}_c)) = \sum_{k=1}^{K} \ln \int_0^\infty \frac{e^{-\mathbf{z}_k^H \mathbf{\Sigma}_k^{-1} \mathbf{z}_k}}{\mathbf{\Pi}^N |\Sigma_k|} f_\tau(\tau_k) d\tau_k \ , \tag{3.9}$$

L'EMV de l'ensemble des vecteurs propres du sous-espace fouillis $\{\hat{\mathbf{v}}_k\}_{k=1,\ldots,R}$ est alors définit comme les R premiers vecteurs propres de la matrice $\mathbf{M}_f$:

$$\mathbf{M}_f = \sum_{k=1}^{K} \left[1 + \frac{h'\left(\sum_{r=1}^{R} \mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k\right)}{h\left(\sum_{r=1}^{R} \mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k\right)} \right] \mathbf{z}_k \mathbf{z}_k^H \ , \tag{3.10}$$

avec

$$h(q) = \int_0^\infty \frac{\exp -q/(1+\tau_k)}{(1+\tau_k)^r} f_\tau(\tau_k) d\tau_k \ . \tag{3.11}$$

On remarque que dans ces équations, $\mathbf{\Pi}_c$ (ou $\{\mathbf{v}_r\}_{r=1,\ldots,R}$) intervient dans l'expression même de son estimation. [94] résout alors cette équation de point fixe (3.10) par un algorithme EM (*Expectation-Maximization*), correspondant exactement à l'approche 2-Step adoptée au chapitre 2 (ainsi que dans la section suivante).

3.3.2 Densité de probabilité de textures inconnue

Dans ces travaux, nous avons voulu dériver un estimateur similaire à [94] ne dépendant pas de la connaissance de la densité de probabilité des textures. Pour cette situation la texture de chaque réalisation τ_k est considérée ici comme un paramètre déterministe inconnu (absorbant l'éventuel facteur de puissance de $\mathbf{\Sigma}_c$). Chaque donnée est alors, conditionnellement τ_k, distribuée selon $(\mathbf{z}_k|\tau_k) \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_k)$, avec

$$\mathbf{\Sigma}_k = \tau_k \mathbf{\Pi}_c + \mathbf{I}_M \ , \tag{3.12}$$

avec la matrice de covariance du fouillis rang faible $\mathbf{\Pi}_c = \mathbf{\Sigma}_c = \sum_{r=1}^{R} \mathbf{v}_r \mathbf{v}_r^H$ La log-vraisemblance de l'ensemble des données correspondante est alors (dérivée identiquement à la section 2.2) :

$$\ln(f(\{\mathbf{z}_k\}|\mathbf{\Pi}_c, \{\tau_i\})) = \sum_{k=1}^{K} \sum_{r=1}^{R} \left[\frac{\tau_k}{1+\tau_k} \mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k - \ln(\tau_k + 1) \right] + \gamma \ , \tag{3.13}$$

où γ est une constante indépendante de $\{\tau_k\}$ et $\{\mathbf{v}_r\}$.

Pour ce modèle, l'EMV du sous-espace fouillis satisfait le théorème suivant :

Théorème 3.3.1 *L'EMV de la base du sous-espace fouillis pour le modèle 3.13 est défini par les* $\{\hat{\mathbf{v}}_k\}$ *étant le* R *vecteurs propres les plus forts de la matrice* $\hat{\mathbf{M}}(\hat{\mathbf{\Pi}}_c)$ *:*

$$\hat{\mathbf{M}}(\hat{\mathbf{\Pi}}_c) = \sum_{k=1}^{K} \frac{\hat{\tau}_k}{\hat{\tau}_k + 1} \mathbf{z}_k \mathbf{z}_k^H \tag{3.14}$$

où $\hat{\tau}_k$ *est l'EMV sous contrainte de positivité des paramètres de textures, défini par :*

$$\hat{\tau}_k = \begin{cases} \frac{1}{R} \mathbf{z}_k^H \hat{\mathbf{\Pi}}_c \mathbf{z}_k - 1 & si ||\hat{\mathbf{\Pi}}_c \mathbf{z}_k||^2 > R \\ 0 & sinon \end{cases} \tag{3.15}$$

Preuve 3.3.1 *c.f Annexe D.*

Comme dans les cas précédents, la matrice $\hat{\mathbf{M}}(.)$ définissant le sous-espace recherché $\hat{\mathbf{\Pi}}_c$ est définie sous forme de point fixe puisque son expression dépend de $\hat{\mathbf{\Pi}}_c$ au travers de l'EMV des textures $\hat{\tau}_k$.

Pour atteindre cet estimateur, nous utiliserons, comme dans la section précédente un algorithme de type 2-Step, maximisant la vraisemblance de manière alternée entre les différents paramètres $\{\tau_k\}$ et $\{\mathbf{v}_r\}$.

- Étape 1 : pour un sous-espace fouillis $\mathbf{\Pi}_c = \sum_{r=1}^{R} \mathbf{v}_r \mathbf{v}_r^H$ fixé supposé connu, l'EMV des texture s'obtient avec l'expression (3.15).
- Étape 2 : pour les textures $\{\tau_k\}$ fixées supposées connues, l'EMV de la base du sous-espace fouillis est solution de :

$$\begin{aligned} \max_{\{\mathbf{v}_r\}} \quad & \sum_{r=1}^{R} \mathbf{v}_r^H \left(\sum_{k=1}^{K} \frac{\tau_k}{1+\tau_k} \mathbf{z}_k^H \mathbf{z}_k^H \right) \mathbf{v}_r \\ \text{s.c.} \quad & \mathbf{v}_r^H \mathbf{v}_r = 1 \ , \ r \in [\![1, R]\!] \\ & \mathbf{v}_r^H \mathbf{v}_j = 0 \ , \ r, j \in [\![1, R]\!] \ , \ r \neq j \end{aligned}$$

  qui s'exprime comme un problème de SVD. La solution est donc définie par les R vecteurs propres dominants de la matrice $\mathbf{M} = \sum_{k=1}^{K} \frac{\tau_k}{\tau_k+1} \mathbf{z}_k \mathbf{z}_k^H$.

L'algorithme associé à ces itérations est récapitulé dans l'encadré 5. Comme cet algorithme correspond à une maximisation alternée de la vraisemblance (3.13), il converge vers un maximum local de celle-ci. Néanmoins, il n'existe pas de garantie concernant l'unicité de l'EMV pour ce contexte et on peut supposer l'algorithme sensible à une mauvaise initialisation.

Algorithm 5 AEMV du sous-espace fouillis

1: Initialiser $\{\mathbf{v}_r\}$ avec les R vecteurs propres les plus forts de la SCM.
2: **repeat**
3: Mettre à jour les textures avec

$$\tau_k = \begin{cases} \frac{1}{R}\mathbf{z}_k^H \mathbf{\Pi}_c \mathbf{z}_k - 1 & \text{si } ||\mathbf{\Pi}_c \mathbf{z}_k||^2 > R \\ 0 & \text{sinon} \end{cases} \tag{3.16}$$

4: Calculer la matrice

$$\mathbf{M} = \sum_{k=1}^{K} \frac{\tau_k}{\tau_k + 1} \mathbf{z}_k \mathbf{z}_k^H \tag{3.17}$$

5: Mettre à jour $\{\mathbf{v}_r\}$ avec les R vecteurs propres les plus forts de $\mathbf{M}$ et $\mathbf{\Pi}_c = \sum_{r=1}^{R} \mathbf{v}_r \mathbf{v}_r^H$.
6: **until** convergence

3.3.3 Interprétations de AEMV

Classiquement, un estimateur du projecteur sur le sous-espace fouillis est dérivé au travers de la SVD d'un estimateur de la matrice de covariance $\hat{\mathbf{\Sigma}}$:

$$\hat{\mathbf{\Sigma}} = \sum_{r=1}^{M} \hat{c}_r \hat{\mathbf{v}}_r \hat{\mathbf{v}}_r^H \quad \longrightarrow \quad \hat{\mathbf{\Pi}}_c = \sum_{r=1}^{R} \hat{\mathbf{v}}_r \hat{\mathbf{v}}_r^H \tag{3.18}$$

Une première intuition voudrait qu'un estimateur précis de la matrice de covariance conduise (par SVD) à un estimateur précis du projecteur. Nous avons par exemple observé que les estimateurs robustes étaient plus précis que la SCM pour la matrice de covariance en contexte de bruit hétérogène rang faible (cf. section 2.7.3). Nous pourrions donc attendre que cette propriété se transpose en termes de sous-espace estimé.

Cependant, dans le contexte fouillis hétérogène rang faible, nous avons vu que l'AEMV de la base du sous-espace fouillis est défini comme les vecteurs propres de la matrice :

$$\mathbf{M} = \sum_{k=1}^{K} \frac{\tau_k}{1 + \tau_k} \mathbf{z}_k \mathbf{z}_k^H \tag{3.19}$$

où τ_k est la texture du fouillis CG de l'échantillon $\mathbf{z}_k$ [1].

Une première remarque est que cette matrice, bien que définissant l'EMV du sous-espace d'intérêt, n'est pas un estimateur de la matrice de covariance pour le contexte considéré. Il est donc possible de dériver des estimateurs de sous-espaces directs, sans recourir à la matrice de covariance. Cet intermédiaire correspond à la SCM des données multipliées par des facteurs correspondants à la proportion de puissance du fouillis. Ces facteurs $\tau_k/(1+\tau_k) \in [0, 1]$ augmentent avec les textures τ_k, ce qui signifie que les réalisations $\mathbf{z}_k$ contenant plus de puissance sur le sous-espace d'intérêt sont favorisées dans le

1. L'EMV exact est défini au travers d'un jeu de matrices $\{\mathbf{M}_r\}$ faisant intervenir les valeurs propres. Néanmoins, leur formulation fait apparaître les même types de facteurs, c'est pourquoi nous nous focaliserons sur l'interprétation de l'AEMV.

processus d'estimation de celui-ci.

Une interprétation de l'expression de cet EMV vient modérer notre première intuition : les estimateurs robustes ont tendance à pénaliser les échantillons contenant beaucoup de puissance, ce qui n'est peut-être pas le procédé d'estimation le plus approprié en termes de sous-espace. Au contraire, supposons que le rapport fouillis à bruit tende vers l'infini pour chaque échantillon $\tau_k/(1+\tau_k) \longrightarrow 1,\ \forall r, k$, alors :

$$\mathbf{M} \longrightarrow \sum_{k=1}^{K} \mathbf{z}_k \mathbf{z}_k^H \propto \hat{\mathbf{\Sigma}}_{SCM} \tag{3.20}$$

montrant que la SCM est proche de contenir l'EMV approché du sous-espace fouillis même si elle n'est pas un estimateur du précis de la covariance.

Il convient donc, pour notre contexte particulier, de dissocier l'estimation de projecteur de celle de la matrice de covariance. Dans la section suivante, nous allons considérer le problème de la robustesse à la contamination de données dans une optique d'estimation de sous-espace.

3.4 AEMV sous hypothèse de données contaminées

3.4.1 Problème de robustesse à la contamination : un bref état de l'art

Classiquement, on fait l'hypothèse que les signaux ne sont pas contaminés par la présence de données aberrantes (présence de cibles par exemple). Les échantillons sont donc considérés comme $i.i.d.$ et ne contenant que du bruit et sont utilisés pour estimer les paramètres recherchés. Si la seconde hypothèse n'est pas vérifiée en pratique, appliquer les méthodes d'estimation classiques peut conduire à de mauvaises performances des traitements adaptatifs. La robustesse à cette condition non standard est un problème classique de traitement du signal, que nous allons explorer dans le contexte d'estimation de sous-espace.

Considérons le modèle de données secondaires :

$$\mathbf{z}_k = \mathbf{c}_k + \mathbf{n}_k + \mathbf{p}_k\ , \quad k \in [\![1, K]\!] \tag{3.21}$$

où $\mathbf{c}_k$ est le fouillis hétérogène rang faible et $\mathbf{n}_k$ le bruit blanc gaussien indépendant, et où $\mathbf{p}_k$ est une corruption potentielle. La présence de cette corruption a pour effet de distordre le sous-espace dominant de la SCM donc d'impacter l'estimation du sous-espace, traditionnellement dérivé de la SVD de cette SCM. Notons par exemple, pour un problème de détection, que si les $\mathbf{p}_k$ sont proportionnels à une cible $\mathbf{p}_0$ présente dans l'échantillon testé, le projecteur estimé à partir de la SCM est susceptible d'englober le projecteur de rang un $\mathbf{p}_0\mathbf{p}_0^H$ (phénomène aussi appelé "subspace swap" dans la littérature [112]). Cette mauvaise estimation conduira à des traitements rang faible annulant la réponse de la cible, donc à un mauvais traitement de détection.

Une première solution est de se tourner vers des estimateurs robustes de la matrice de covariance, et supposer que cette robustesse se transpose en termes d'estimation de sous-espace dominant. Du point de

vue de l'estimation de la matrice de covariance, plusieurs approches permettent de prendre en compte une potentielle corruption des échantillons :

- L'utilisation du "diagonal-loading" : c'est à dire régulariser un estimateur $\hat{\mathbf{\Sigma}}$ par $\hat{\mathbf{\Sigma}} = \hat{\mathbf{\Sigma}} + \beta \mathbf{I}_M$. Utile en termes d'estimation de la matrice de covariance, le diagonal-loading ne modifie cependant pas les vecteurs propres de l'estimateur et donc n'est pas adapté pour la robustification d'estimateurs de sous-espace.
- Prendre en compte la présence de données aberrantes dans la vraisemblance du modèle. Le problème est alors de maximiser cette vraisemblance en estimant les paramètres et détectant conjointement la présence des aberrations. La solution s'obtient au travers d'algorithmes effectuant des sous-partitionnements et une sélection adaptative d'échantillons ([43] et les références contenues). On note que dans ces algorithmes, la quantité $\mathbf{z}_k^H \mathbf{\Sigma}^{-1} \mathbf{z}_k$ (détecteur d'anomalies de Kelly [39]) joue un rôle important dans la détection d'outliers. Néanmoins, la résolution de cet algorithme est combinatoire.
- De récents travaux ont considéré ce modèle dans un contexte bayésien [77, 15], nécessitant alors à un a priori sur la distribution de la contamination.
- L'utilisation de la médiane sur des estimateurs construits à partir de sous-partitions des échantillons.
- Utiliser un estimateur robuste de la matrice de covariance (M-estimateur)

$$\hat{\mathbf{M}} = \sum_{k=1}^{K} \psi \left(\mathbf{z}_k^H \hat{\mathbf{M}}^{-1} \mathbf{z}_k^H \right) \mathbf{z}_k \mathbf{z}_k^H \tag{3.22}$$

  En effet, ces points-fixes sont définis via des poids adaptatifs pénalisant les fortes valeurs $\mathbf{z}_k^H \mathbf{\Sigma}^{-1} \mathbf{z}_k$, et donc rejetant les outliers. Cette solution a attiré beaucoup d'attention ces dernières années. En effet, ces estimateurs sont maniables et facilement calculables, et ne nécessitent pas de sous-partitionnement des échantillons. De plus, ils ne nécessitent pas d'a priori. Leurs performances théoriques ainsi qu'une étude théorique sur la robustesse (fonction d'influence, breakdown point) ont été dérivées [84, 71].

Cependant, comme évoqué précédemment (et confirmé en simulations), les M-estimateurs ne semblent pas appropriés au contexte considéré puisque la quantité $\mathbf{z}_k^H \mathbf{\Sigma}^{-1} \mathbf{z}_k$, croissante avec les textures du fouillis, peut conduire à rejeter des échantillons alors devraient être favorisés dans le processus d'estimation du sous-espace (comme suggéré par l'EMV ou l'AEMV). D'autre part, la précédente analyse suggère que l'EMV est proche d'être contenu dans la SCM et est donc fortement sensible aux données aberrantes. Le but de ces travaux est donc de développer un estimateur de sous-espace fouillis offrant un meilleur compromis performance-robustesse.

3.4.2 Estimateur AEMV modifié

Considérons le modèle de données suivant :

$$\mathbf{z}_k = \mathbf{c}_k + \mathbf{p}_k + \mathbf{n}_k \ , \tag{3.23}$$

où $\mathbf{c}_k$ est le fouillis de rang faible dont on veut estimer le sous-espace correspondant et $\mathbf{n}_k$ est le bruit blanc. $\mathbf{p}_k$ est une corruption potentielle, modélisée comme un vecteur CG indépendant contenu dans le sous-espace orthogonal à $\mathbf{c}_k$. Pour les raisons de simplification justifiées précédemment, on néglige encore la différence entre les valeurs propres des matrices de covariance de chaque contribution. Soit alors le modèle :

$$\mathbf{z}_k \sim \mathcal{CN}\left(\mathbf{0}, \tau_k \mathbf{\Pi}_c + \beta_k \mathbf{\Pi}_c^{\perp} + \mathbf{I}_M\right) , \tag{3.24}$$

avec τ_k et β_k des variables déterministes inconnues.

Comme précédemment, on construit l'EMV via l'algorithme 2-Step visant à atteindre un maximum local de la vraisemblance. Les dérivations étant strictement identiques à celles de A-MLE, nous présentons simplement les résultats :

– Étape 1 : pour $\{\mathbf{v}_r\}$ fixé, l'EMV sous contrainte de positivité des paramètres τ_k et β_k est obtenu avec :

$$\hat{\tau}_k = \begin{cases} \left(\mathbf{z}_k^H \mathbf{\Pi}_c \mathbf{z}_k\right)/R - 1 & \text{si } ||\mathbf{\Pi}_c \mathbf{z}_k||^2 > R \\ 0 & \text{sinon} \end{cases} , \tag{3.25}$$

et

$$\hat{\beta}_k = \begin{cases} \left(\mathbf{z}_k^H \mathbf{\Pi}_c^{\perp} \mathbf{z}_k\right)/(M-R) - 1 & \text{si } ||\mathbf{\Pi}_c^{\perp} \mathbf{z}_k||^2 > M - R \\ 0 & \text{sinon} \end{cases} . \tag{3.26}$$

– Étape 2 : pour $\{\tau_k\}$ et $\{\beta_k\}$ fixés, l'EMV sous contrainte unitaire de la base du sous-espace fouillis $\{\mathbf{v}_r\}$ est obtenu par les R vecteurs propres dominants de la matrice :

$$\mathbf{R} = \sum_{k=1}^{K} \left(\frac{\max(\ 0\ ,\ \tau_k - \beta_k\ )}{(\beta_k + 1)\,(\tau_k + 1)} \right) \mathbf{z}_k \mathbf{z}_k^H , \tag{3.27}$$

où le seuillage sur les facteurs $\tau_k - \beta_k$ intervient pour garantir des facteurs positifs.

Les itérations de ces deux étapes, résumées dans l'encadré 6, convergent alors vers un maximum local de la vraisemblance associée au modèle (3.24). Cependant, comme dans les sections précédentes, il n'existe pas de preuve d'unicité d'un tel EMV, c'est pourquoi nous notons que l'algorithme proposé est susceptible d'être sensible à l'initialisation.

L'estimateur ainsi dérivé présente une formulation intéressante puisqu'il fait intervenir des poids adaptatifs α_k :

$$\alpha_k = \frac{\max(\ 0\ ,\ \tau_k - \beta_k\ )}{(\beta_k + 1)\,(\tau_k + 1)} . \tag{3.28}$$

Supposons que $\beta_k \approx 0$, alors $\alpha_k \approx \tau_k/(1+\tau_k)$ se comporte comme les facteurs de pondérations d'AEMV : les données ayant une forte puissance dans le sous-espace d'intérêt sont favorisées pour l'estimer. Pour β_k non nul (une corruption est présente), on distingue deux cas. Si $\beta_k > \tau_k$ alors $\alpha_k = 0$, l'échantillon est simplement rejeté. Dans le cas contraire, supposons $1 \ll \tau_k$ alors $\alpha_k \approx 1/(1+\beta_k)$: la contribution de l'échantillon n'est pas annulée mais diminuée proportionnellement à la puissance de la corruption. L'estimateur proposé pénalise donc les données potentiellement nuisibles au processus d'estimation du sous-espace d'intérêt, ce que ne fait pas AEMV.

Algorithm 6 R-AEMV du sous-espace fouillis

1: Initialiser $\{\mathbf{v}_r\}$ avec les R vecteurs propres les plus forts de la SCM.
2: **repeat**
3: Mettre à jour les textures avec

$$\hat{\tau}_k = \begin{cases} \left(\mathbf{z}_k^H \mathbf{\Pi}_c \mathbf{z}_k\right)/R - 1 & \text{si } ||\mathbf{\Pi}_c \mathbf{z}_k||^2 > R \\ 0 & \text{sinon} \end{cases} \tag{3.29}$$

et

$$\hat{\beta}_k = \begin{cases} \left(\mathbf{z}_k^H \mathbf{\Pi}_c^{\perp} \mathbf{z}_k\right)/(M-R) - 1 & \text{si } ||\mathbf{\Pi}_c^{\perp} \mathbf{z}_k||^2 > M - R \\ 0 & \text{sinon} \end{cases} \tag{3.30}$$

4: Calculer la matrice

$$\mathbf{R} = \sum_{k=1}^{K} \left(\frac{\tau_k - \beta_k}{(\beta_k + \sigma_c)(\tau_k + \sigma_c)} \right) \mathbf{z}_k \mathbf{z}_k^H \tag{3.31}$$

5: Mettre à jour $\{\mathbf{v}_r\}$ avec les R vecteurs propres les plus forts de $\mathbf{R}$ et $\mathbf{\Pi}_c = \sum_{r=1}^{R} \mathbf{v}_r \mathbf{v}_r^H$.
6: **until** Convergence

3.5 Simulations

Cette section vise à illustrer les performances des algorithmes développés pour l'estimation du projecteur sur le sous-espace fouillis dans un contexte CG rang faible plus bruit blanc gaussien. Le critère considéré est la précision d'estimation de cette matrice de covariance, mesuré en termes d'erreur quadratique moyenne normalisée (notée NMSE pour "Normalized Mean Square Error") :

$$\text{NMSE}(\hat{\mathbf{\Pi}}_c) = \mathbb{E}\left(\frac{||\hat{\mathbf{\Pi}}_c - \mathbf{\Pi}_c||^2}{||\mathbf{\Pi}_c||^2} \right) \tag{3.32}$$

pour un estimateur $\hat{\mathbf{\Pi}}_c$ donné.

Nous étudierons les estimateurs de projecteur sur le sous-espace fouillis suivants :

- SCM : le projecteur estimé au travers des R vecteurs propres dominants de la Sample Covariance Matrix, décrite section 1.3.2.
- FPE : le projecteur estimé au travers des R vecteurs propres dominants du point fixe, décrit à la section 1.3.5. Pour les cas où $K < M$, nous utiliserons l'estimateur régularisé SFPE, décrit à la section 1.3.6. Comme il n'existe pas de règle de choix adaptatif "optimal" pour le paramètre de régularisation β pour le contexte que nous considérons, nous utiliserons la valeur minimale requise pour son existence. L'estimateur considéré se définit donc comme le SFPE utilisant le parmètre $\beta_{min} = \max(0, 1 - K/M + \epsilon)$ (on fixe $\epsilon = 0.02$).
- EMV-MM2 : le projecteur construit avec maximum de vraisemblance de $\{\mathbf{v}_r\}$, calculé à l'aide de l'algorithme 4. Pour éviter une certaine redondance, nous ne considérons que cet algorithme parmi ceux précédemment développés. Ce choix est justifié par la figure 3.1, comparant les NMSE des différents algorithmes "EMV" pour une configuration donnée. Cette figure montre que les

différents algorithmes conduisent des performances relativement comparables. Cependant, les algorithmes MM1 et MM2 offrent (identiquement) un gain en précision comparé à 2SD et 2SR.
- AEMV : le maximum de vraisemblance approché dérivé section 3.4.2 et calculé avec l'algorithme 5.
- LR-FPE : l'estimateur du projecteur dérivé via les itérations définies 3.2.
- R-AMEV : l'estimateur du maximum de vraisemblance du modèle modifié décrit section 3.4.2.

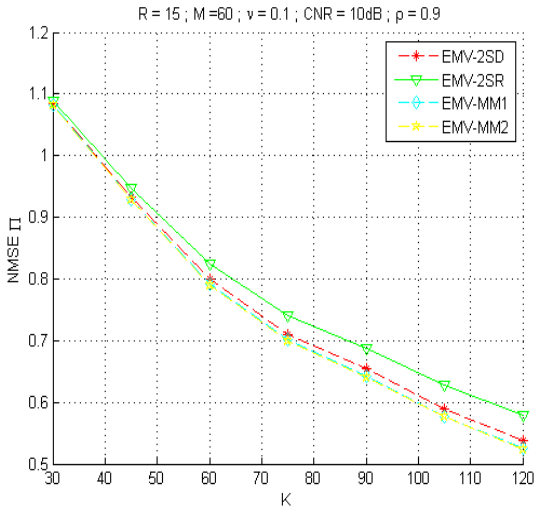

FIGURE 3.1 – NMSE sur $\mathbf{\Pi}_c$ des estimateurs EMV obtenus par les différents algorithmes dérivés au chapitre 2. $M = 60$, $R = 15$, $\nu = 0.1$, CNR$= 10$.

3.5.1 Paramètres

Dans ces simulations, les échantillons $\mathbf{z}_k$ sont générés selon la vraisemblance décrite à la section 2.2 : $\mathbf{z}_k = \mathbf{c}_k + \mathbf{n}_k$. Le bruit blanc gaussien suit $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_M)$ avec $\sigma^2 = 1$. Le fouillis rang faible est distribué selon $(\mathbf{c}_k|\tau_k) \sim \mathcal{CN}(\mathbf{0}, \tau_k\mathbf{\Sigma})$, avec une texture aléatoire τ_k, $i.i.d.$ pour chaque échantillon. La densité de probabilité des textures est une distribution Gamma, noté $\tau \sim \Gamma(\nu, 1/\nu)$, conduisant à un fouillis K-distribué (cf. section 1.2.4). Le paramètre de forme est ν et le paramètre d'échelle $1/\nu$, la distribution Gamma vérifie donc $\mathbb{E}(\tau) = 1$. La matrice de covariance du fouillis $\mathbf{\Sigma}_c$ est construite à l'aide des R valeurs et vecteurs propres dominants d'une matrice Toeplitz de paramètre de corrélation $\rho \in [0, 1]$. Cette matrice est ensuite mise à l'échelle pour fixer le rapport fouillis à bruit (noté CNR), CNR $= \mathbb{E}(\tau)\text{Tr}(\mathbf{\Sigma}_c)/(R\sigma^2)$, à une valeur voulue. Si non précisé, les données ne sont pas corrompues $\mathbf{p}_k = \mathbf{0}$.

3.5.2 Résultats

Les figures 3.2 et 3.3 présentent l'évolution du NMSE en fonction du nombre d'échantillons K pour diverses configurations de fouillis. Chacune de ses figures présente les résultats pour un rapport fouillis à bruit faible (3dB), moyen (10dB) et fort (30dB). En figure 3.2, le paramètre ν du fouillis K-distribué est fixé à $\nu = 1$, conduisant à un fouillis modérément hétérogène. En figure 3.3, ce paramètre est fixé à

$\nu = 0.1$, conduisant à un fouillis fortement hétérogène.

Voici la description de ces résultats en considérant les estimateurs un par un :

- On observe que la SCM permet de construire un estimateur de $\mathbf{\Pi}_c$ souvent proche de l'optimum observé. Ces résultats confirment l'interprétation faite en section 3.3.3, justifiant que la SCM peut contenir un bon estimateur de projecteur même si elle n'est pas un bon estimateur de la matrice de covariance dans le contexte considéré (comme observé dans les simulations du chapitre 2).
- L'estimateur de $\mathbf{\Pi}_c$ dérivé du FPE présente de moins bonnes performances que celui issu la SCM. Or, pour le modèle considéré, le point fixe est un estimateur plus précis de la matrice de covariance que la SCM (cf. simulations chapitre 2). Le résultat semble contre intuitif, cependant, il vient bien confirmer l'interprétation section 3.3.3 : un estimateur précis de la matrice totale n'est pas nécessairement le plus efficace en termes d'estimation de projecteur. En effet, le point fixe $\hat{\mathbf{\Sigma}}_{FPE}$ pondère les données par la quantité $(\mathbf{z}_k^H \hat{\mathbf{\Sigma}}_{FPE}^{-1} \mathbf{z}_k)^{-1}$, croissante avec les textures du fouillis. Bien que cette opération donne lieu à une meilleure estimation de la matrice de covariance, elle peut conduire à rejeter des échantillons pertinents pour le processus d'estimation du sous-espace (comme suggéré par la formulation de l'EMV). Comme observé au chapitre 2, pour $K < M$, l'estimateur considéré est SFPE avec le paramètre de régularisation β minimum. Ce choix de paramètre n'étant pas nécessairement optimal il est normal d'observer un décrochement de performances entre les régimes $K < M$ et $K > M$.
- Le maximum de vraisemblance (EMV-MM2) à des performances proche de la SCM mais s'avère plus précis en contexte de faible rapport fouillis à bruit et pour des fouillis plus hétérogènes.
- Le maximum de vraisemblance approché du sous-espace fouillis (AEMV) atteint des performances très proches de l'EMV-MM2. Ceci vient justifier l'approximation faite sur le modèle puisque négliger la différence des valeurs propres de la matrice de covariance du fouillis dans le modèle ne nuit pas grandement à l'estimation de son sous-espace propre. Ceci met de plus en avant un intérêt pratique de la méthode proposée. En effet, l'algorithme AEMV est plus simple à implémenter que les algorithmes développés chapitre 2 et atteint des performances similaires.
- L'estimateur dérivé de l'algorithme R-AEMV atteint les mêmes résultats que l'estimateur AEMV. Ceci est une propriété intéressante. En effet, la modification du modèle amont (supposant une contamination) conduit à un estimateur ayant des performances inchangées, et proches de l'optimum en contexte non contaminé. La robustesse de cet estimateur à de potentielles contaminations sera par la suite illustrée dans le chapitre 4.
- l'estimateur LR-FPE offre l'estimation la moins précise du projecteur sur le sous-espace fouillis. Ses performances semblent d'autant plus impactées que le fouillis est hétérogène. Cependant cet estimateur présente une grande robustesse à la contamination des données, ce que nous illustrerons dans le chapitre suivant.

Pour résumer, l'EMV atteint les meilleures performances en termes de précision d'estimation du sous-espace fouillis dans toutes les configurations. Cependant, les estimateurs AEMV et R-AEMV ont des performances quasi équivalentes ce qui suggère un fort intérêt car ces estimateurs sont plus simples à implémenter.

Néanmoins, comme cela a été soulevé chapitre 2, le NMSE est un critère illustratif qui n'est pas lié aux performances d'une application spécifique. Un estimateur de $\mathbf{\Pi}_c$ peut s'avérer précis en termes de NMSE sans nécessairement conduire au meilleur traitement adaptatif considéré. C'est pourquoi un critère lié à l'application du filtrage adaptatif rang faible sera étudié chapitre 4.

3.6 Synthèse du chapitre 3

Dans ce chapitre, nous avons présenté l'approximation rang faible, basée sur l'estimation du projecteur sur le sous-espace fouillis. De nombreux traitements de TS utilisent cette approximation car elle permet d'atteindre de bonnes performances avec moins d'échantillons, en comparaison avec les traitements basés sur un estimé de la covariance. En pratique, un estimateur du projecteur sur le sous-espace fouillis est obtenu à partir de la SVD d'un estimateur de la matrice de covariance.

Cependant, pour le contexte d'un fouillis CG plus bruit blanc gaussien, nous avons vu au chapitre 2 que l'EMV de la base du sous-espace fouillis n'est lié à aucun estimateur "classique". Cet EMV nécessitant des algorithmes relativement complexes pour être calculé, nous avons donc envisagé dans ce chapitre des relaxations du modèle pour obtenir des estimateurs simples et directs du sous-espace, adaptés au contexte considéré.

Dans un premier temps, nous avons considéré le modèle présenté à la section 2.5.1 (fort rapport fouillis à bruit) en relâchant la contrainte d'orthogonalité entre le sous-espace fouillis et son complémentaire. Ce modèle conduit à l'estimateur LR-FPE, obtenu par un algorithme de point fixe utilisant une pseudo-inverse du rang du fouillis.

Dans un second temps avons considéré le modèle simplifié où les valeurs propres de la matrice de covariance du fouillis sont égales. Ceci conduit à un EMV approché (AEMV) du sous-espace fouillis pouvant être obtenu par de simples itérations de SVD de la SCM des données pondérées par des facteurs adaptatif. Une interprétation des facteurs soulève le fait que l'EMV du sous-espace fouillis est généralement proche d'être contenu dans les vecteurs propres dominants de la SCM. Ce résultat est surprenant car la SCM s'avère être un mauvais estimateur de la matrice de covariance pour le contexte considéré. Ceci suggère de plus que l'EMV est potentiellement peu robuste à la corruption des données.

L'interprétation précédente, appuyée par un bref état de l'art, nous conduit à proposer un modèle modifié pour tenter de répondre à cette problématique de contaminations des données. En considérant une potentielle corruption (modélisée par un CG orthogonal au fouillis) dans le modèle AEMV, nous avons proposé un nouvel estimateur du sous-espace fouillis : R-AEMV. L'étude des facteurs adaptatifs définissant cet estimateur suggère qu'il puisse proposer un meilleur compromis performance-robustesse que les estimateurs EMV et AEMV.

Au cours de simulations de validation, nous avons observé que les estimateurs AEMV et R-AEMV atteignent des performances équivalentes à celles de l'EMV exact. Ceci suggère un fort intérêt car ces estimateurs sont plus simples à implémenter. L'estimateur LR-FPE s'est avéré peu précis, cependant, ses propriétés de robustesse seront illustrées dans le chapitre suivant.

Dans le chapitre suivant, nous testerons les performances des estimateurs développés au cours de cette thèse sur des critères liés à une application de radar STAP.

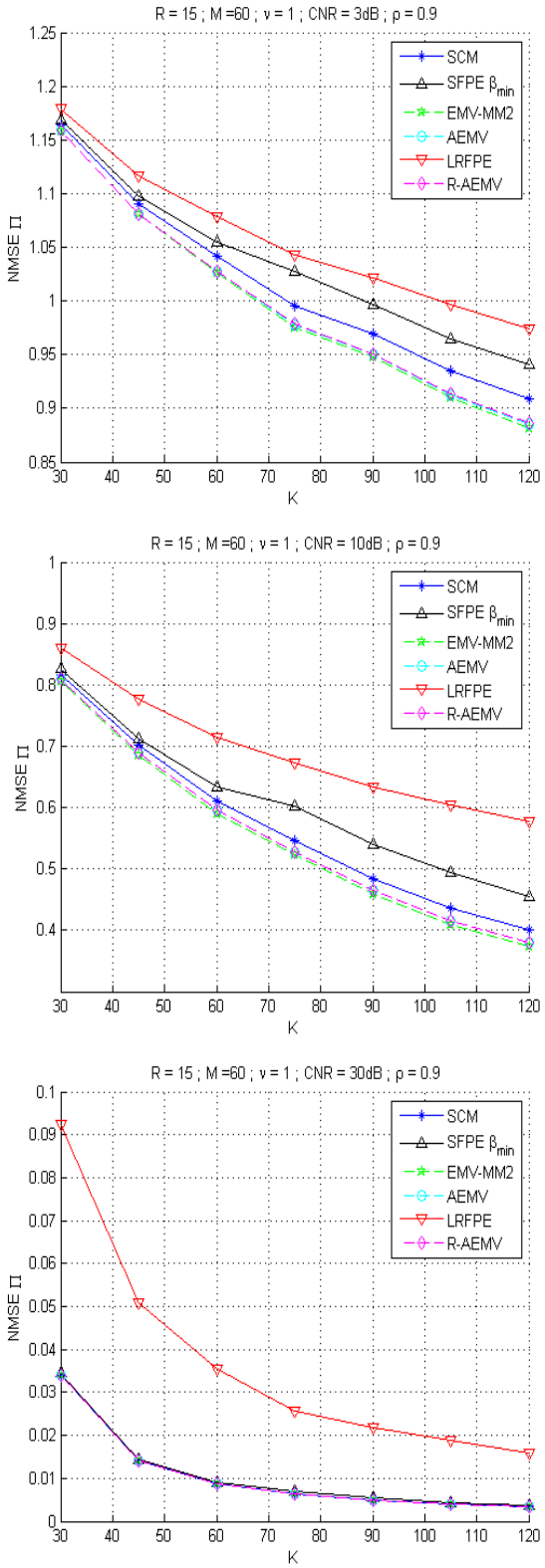

FIGURE 3.2 – NMSE sur $\mathbf{\Pi}_c$ des différents estimateurs considérés. $M = 60$, $R = 15$, $\nu = 1$. De haut en bas : CNR$= [3, 10, 30]$dB.

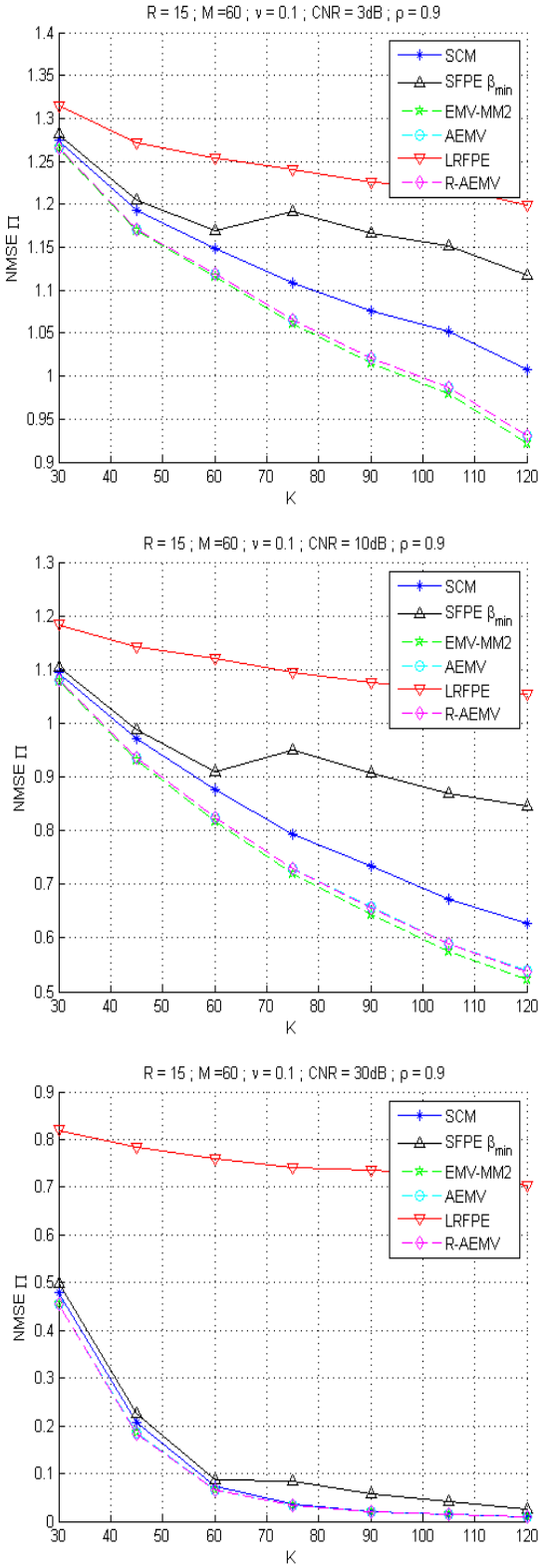

FIGURE 3.3 – NMSE sur $\mathbf{\Pi}_c$ des différents estimateurs considérés. $M = 60$, $R = 15$, $\nu = 0.1$. De haut en bas : CNR$= [3, 10, 30]$dB.

Annexe C

Preuves du chapitre 3

C.1 Preuve du théorème 3.3.1

Preuve C.1.1 *La log-vraisemblance*

$$\ln(f(\{\mathbf{z}_k\}|\mathbf{\Pi}_c,\{\tau_i\})) = \sum_{k=1}^{K}\sum_{r=1}^{R}\left[\frac{\tau_k}{1+\tau_k}\mathbf{z}_k^H\mathbf{v}_r\mathbf{v}_r^H\mathbf{z}_k - \ln(\tau_k+1)\right] + \gamma\,, \tag{C.1}$$

est dérivée par rapport à τ_k, pour un $k \in [\![1,K]\!]$ donné :

$$\frac{\partial \ln(f(\{\mathbf{z}_k\}|\tau_k))}{\partial \tau_k} = \frac{\mathbf{z}_k^H\mathbf{\Pi}_c\mathbf{z}_k}{(1+\tau_k)^2} - R\ln(\tau_k+1) \tag{C.2}$$

l'annulation de cette expression permet d'identifier l'EMV $\hat{\tau_k}$. Cependant, les textures sont des facteurs positifs, comme la vraisemblance est strictement décroissante après son maximum, l'EMV sous contrainte de positivité est obtenu par :

$$\hat{\tau_k} = \begin{cases} \frac{1}{R}\mathbf{z}_k^H\mathbf{\Pi}_c\mathbf{z}_k - 1 & \textit{si } ||\mathbf{\Pi}_c\mathbf{z}_k||^2 > R \\ 0 & \textit{sinon} \end{cases} \tag{C.3}$$

Supposant ces EMV non nuls (auquel cas, les données correspondantes sont simplement rejetées), leur expression est reportée dans (C.1) pour obtenir la log-vraisemblance réduite, exprimée avec $\mathbf{\Pi}_c = \sum_{r=1}^{R}\mathbf{v}_r\mathbf{v}_r^H$:

$$\ln(\hat{f}(\{\mathbf{z}_k\}|\{\mathbf{v}_k\})) = -\sum_{k=1}^{K}\mathbf{z}_k^H\mathbf{z}_k + \sum_{k=1}^{K}\sum_{r=1}^{R}\mathbf{z}_k^H\mathbf{v}_r\mathbf{v}_r^H\mathbf{z}_k - R\sum_{k=1}^{K}\ln\left(\frac{1}{r}\sum_{r=1}^{R}\mathbf{z}_k^H\mathbf{v}_r\mathbf{v}_r^H\mathbf{z}_k\right) + \gamma\,. \tag{C.4}$$

Pour contraindre l'orthonormalité des $\mathbf{v}_r$, on réutilise la méthode des multiplicateurs de Lagrange dérivée dans l'annexe A du chapitre 2. Notons $\mathcal{L}$ la fonction de Lagrange, la fonctionnelle à maximiser selon les $\mathbf{v}_k$ est alors :

$$g = \ln\hat{f} - \mathcal{L}\,, \tag{C.5}$$

qui dérivée par rapport à $\mathbf{v}_j^H$ pour un $j \in [\![1,R]\!]$ fixé, donne :

$$\frac{\partial g(\{\mathbf{z}_k\})}{\partial \mathbf{v}_j^H} = \frac{\partial \ln\hat{f}(\{\mathbf{z}_k\})}{\partial \mathbf{v}_j^H} + \frac{\partial \mathcal{L}}{\partial \mathbf{v}_j^H} \tag{C.6}$$

Dont le premier terme est donné par :

$$\frac{\partial \ln \hat{f}(\{\mathbf{z}_k\})}{\partial \mathbf{v}_j^H} = \sum_{k=1}^{K} \mathbf{z}_k \mathbf{z}_k^H \mathbf{v}_j - R \sum_{k=1}^{K} \frac{\mathbf{z}_k \mathbf{z}_k^H \mathbf{v}_j}{\mathbf{z}_k^H \left(\sum_{r=1}^{R} \mathbf{v}_r \mathbf{v}_r^H \right) \mathbf{z}_k} , \tag{C.7}$$

où l'on peut identifier les EMV $\hat{\tau}_k$ des paramètres τ_k :

$$\frac{\partial \hat{f}(\{\mathbf{z}_k\})}{\partial \mathbf{v}_j^H} = \left(\sum_{k=1}^{K} \frac{\hat{\tau}_k}{\hat{\tau}_k + 1} \mathbf{z}_k \mathbf{z}_k^H \right) \mathbf{v}_j = \hat{\mathbf{M}} \mathbf{v}_j , \tag{C.8}$$

avec $\hat{\mathbf{M}} = \sum_{k=1}^{K} \frac{\hat{\tau}_k}{\hat{\tau}_k + 1} \mathbf{z}_k \mathbf{z}_k^H$.
La dérivée de la fonction de Lagrange est obtenue du lemme A.1.1 (chapitre 2 Annexe A) :

$$\frac{\partial \mathcal{L}}{\partial \mathbf{v}_j^H} = \mathbf{V} \mathbf{\Lambda} \mathbf{e}_j , \tag{C.9}$$

avec $\mathbf{e}_i$ le i^{ieme} vecteur de la base canonique, $\mathbf{V}$ la matrice concaténant les vecteurs $\{\mathbf{v}_r\}$, et où $\mathbf{\Lambda}$ est la matrice hermitienne contenant les multiplicateurs de Lagrange. La dérivée de g s'annule pour :

$$\frac{\partial g(\mathbf{z}_1, ..., \mathbf{z}_K)}{\partial \mathbf{v}_j^H} = 0 \Leftrightarrow \hat{\mathbf{M}} \mathbf{v}_j = \mathbf{V} \mathbf{\Lambda} \mathbf{e}_j , \tag{C.10}$$

Ces expressions concaténées selon tous les $\mathbf{v}_j$ donnent :

$$\hat{\mathbf{M}} \mathbf{V} = \mathbf{V} \mathbf{\Lambda} , \tag{C.11}$$

ou, de façon équivalente, en utilisant $\mathbf{V}\mathbf{V}^H = \mathbf{\Pi}_c$:

$$\hat{\mathbf{M}} \mathbf{\Pi}_c = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^H . \tag{C.12}$$

Les matrices $\hat{\mathbf{M}}$ et $\mathbf{\Lambda}$ sont Hermitiennes, on a donc :

$$\mathbf{\Pi}_c \hat{\mathbf{M}} = \left(\hat{\mathbf{M}} \mathbf{\Pi}_c \right)^H = \left(\mathbf{V} \mathbf{\Lambda} \mathbf{V}^H \right)^H = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^H = \hat{\mathbf{M}} \mathbf{\Pi}_c \tag{C.13}$$

Ainsi, $\hat{\mathbf{M}}$ et $\mathbf{\Pi}_c$ commutent. Ce qui signifie que $\mathbf{\Pi}_c$ est un projecteur sur un sous-espace décrit par R vecteurs propres de $\hat{\mathbf{M}}$. En considérant cette solution avec un objectif de maximisation, on conclue la preuve de ce théorème : l'EMV de la base du sous-espace fouillis est défini par les $\{\hat{\mathbf{v}}_r\}$ étant les R vecteurs propres dominants de $\hat{\mathbf{M}}(\mathbf{\Pi}_c)$:

$$\hat{\mathbf{M}}(\mathbf{\Pi}_c) = \sum_{i=1}^{K} \frac{\hat{\tau}_i}{\hat{\tau}_i + 1} \mathbf{z}_i \mathbf{z}_i^H . \tag{C.14}$$

Chapitre 4

Application au radar STAP

Dans ce chapitre, nous étudions les performances des estimateurs proposés (et de l'état de l'art) sur une application de Space Time Adaptative Processing (STAP) pour radar aéroporté, au travers de simulations et de données réelles. Les données reçues dans cette application suivent en effet le modèle considéré au long de cette thèse car il est connu que la matrice de covariance du fouillis radar est de rang faible (ce rang peut être évalué grâce à la formule de Brennan). Nous présentons dans un premier temps le système étudié ainsi que le modèle de signaux observés correspondant. Nous illustrons ensuite les performances des estimateurs de matrice de covariance développés chapitre 2 sur une problématique de détection adaptative. Les performances des estimateurs sur le sous-espace fouillis développés chapitre 3 sont enfin étudiées au travers d'une application de filtrage adaptatif rang faible. Afin de compléter cette étude, nous étudions de plus la robustesse des méthodes développées à des erreurs de modèle (mauvaise estimation du rang, présence d'outliers...).

4.1 Présentation du système

Cette section présente une description succincte du système, inspirée de [12] et [47]. Pour de plus amples détails, nous renvoyons le lecteur au rapport technique [120].

Dans le cadre d'un radar aéroporté utilisé dans le but de détecter des cibles mobiles, les performances des détecteurs utilisant uniquement l'information temporelle (décalage *Doppler*) sont faibles à cause de l'étalement *Doppler* de la réponse du sol (appelé fouillis ou *clutter*) induit par la vitesse de la plate-forme. Par exemple, les cibles à faible vitesse sont alors indétectables. Une solution pour remédier à ce problème est d'ajouter une information spatiale en utilisant une antenne composée de plusieurs capteurs sur la plate-forme aéroportée. En utilisant conjointement les deux informations, temporelles et spatiales, il est alors possible de détecter des cibles mobiles même celles à basse vitesse [17] comme le montre la figure 4.1.

Nous présentons dans les sections suivantes le principe d'un radar STAP, les modèles des signaux de cible et de fouillis ainsi que la matrice de covariance du bruit.

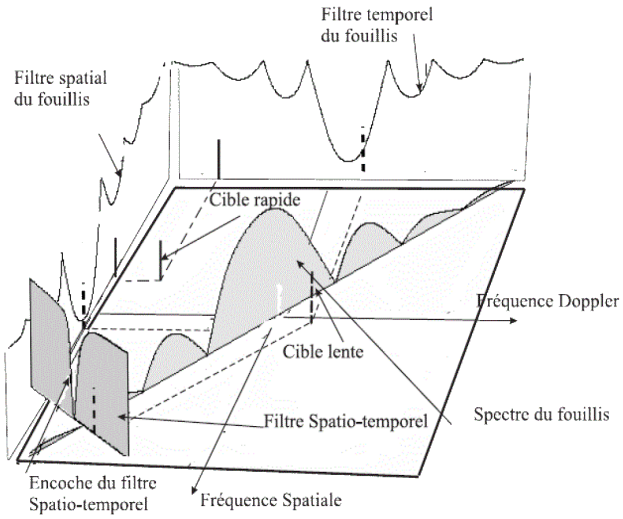

FIGURE 4.1 – Principe et intérêt du filtrage spatio-temporel.

4.1.1 Présentation du radar

La géométrie d'un système STAP classique est présentée sur la figure. 4.2. Nous considérons qu'il est composé d'un radar à impulsions *Doppler* émettant M impulsions cohérentes à une fréquence de répétition f_r et qu'il est embarqué sur une plate-forme aéroportée volant à une altitude h et à une vitesse uniforme v_p. L'antenne radar est constituée de N capteurs uniformément répartis (d est l'espacement inter-capteurs) orientés vers la direction du mouvement de la plate-forme (configuration *side-looking* qui est la plus simple). A la réception, les retours radar sur chaque élément d'antenne après transformation en bande de base et filtrage adapté sont échantillonnés à une fréquence f_e. La fréquence spatiale du radar est notée $\lambda = \frac{c}{f_0}$ avec c la vitesse de propagation des ondes électromagnétiques et f_0 la fréquence centrale des signaux émis. On se place dans le cadre d'un signal émis bande étroite.

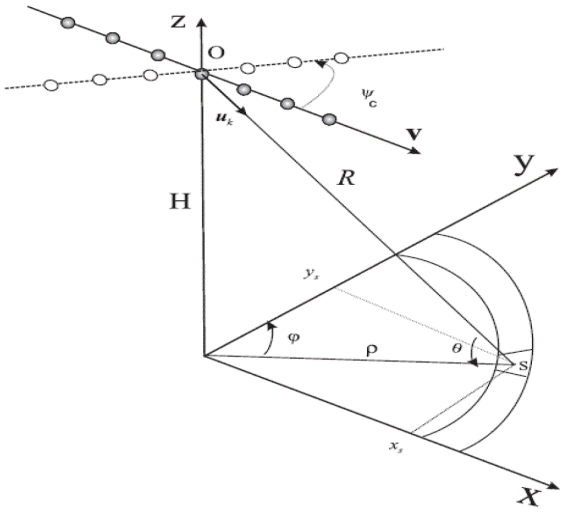

FIGURE 4.2 – Géométrie du STAP.

Le vecteur de position du n^{ieme} élément d'antenne est donné par :

$$\mathbf{d}_n = (n-1)d\mathbf{u}_x, \tag{4.1}$$

où $\mathbf{u}_x$ est un vecteur unitaire du système de coordonnées cartésiennes dans l'axe x. Le vecteur de vitesse de la plate-forme est donné par :

$$\mathbf{v}_p = v_p\mathbf{u}_x. \tag{4.2}$$

Dans la suite de ce chapitre, la position d'un point sera présentée par un vecteur unitaire $\mathbf{u}_k$ décrit par les angles azimut et élévation (ϕ, θ) :

$$\mathbf{u}_k(\phi,\theta) = cos\theta sin\phi\mathbf{u}_x + cos\theta cos\phi\mathbf{u}_y + sin\theta\mathbf{u}_z, \tag{4.3}$$

où $\mathbf{u}_y$ et $\mathbf{u}_z$ sont des vecteurs unitaires du système de coordonnées cartésiennes dans les axes y et z.

4.1.2 Modèle des signaux

Les données STAP forment ce qu'on appelle un data cube contenant les réponses spatio-temporelles des $K+1$ différentes cases distances de la scène illuminée par le radar. Pour chaque case distance k, les informations temporelles et spatiales sont concaténées dans un vecteur comme montré dans la figure 4.3 pour former un vecteur du signal reçu $\mathbf{z}_k$.

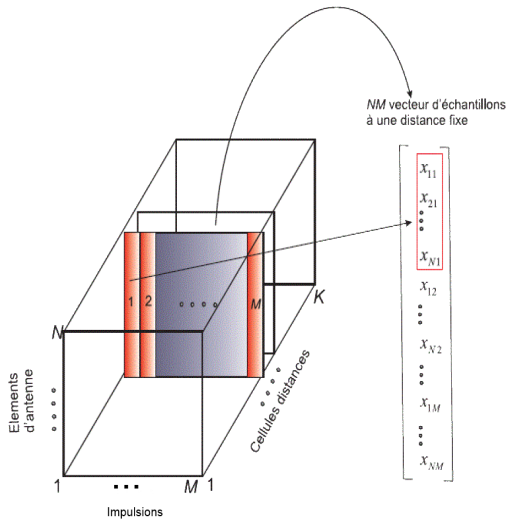

FIGURE 4.3 – Data Cube STAP et arrangement de données.

Ce signal reçu, $\mathbf{z}_k$, dans la cellule distance k est un mélange de différents signaux (on ne tient pas compte ici de possibles brouilleurs) :

- éventuellement le signal désiré provenant de la cible,
- le fouillis provenant de la réponse du sol,
- le bruit blanc causé par l'électronique du système.

Sauf spécifié dans l'étude sur la robustesse, nous supposerons dans la suite que la cible est présente dans une seule case distance, appelée la donnée sous test et notée $\mathbf{z}$. Sous l'hypothèse de présence de cible,

on peut donc écrire le signal reçu sous la forme suivante :

$$\mathbf{z}_0 = \mathbf{d} + \mathbf{c}_0 + \mathbf{n}_0, \tag{4.4}$$

où $\mathbf{d}$, $\mathbf{c}_0$ et $\mathbf{n}_0$ sont respectivement le signal provenant de la cible, du fouillis et du bruit blanc. On dispose de plus de K signaux contenant uniquement du fouillis et le bruit blanc :

$$\mathbf{z}_k = \mathbf{c}_k + \mathbf{n}_k \text{ avec } k = [\![1, K]\!] \tag{4.5}$$

Ces signaux sont communément appelés les données secondaires et servent à estimer les paramètres inconnus des interférences (bruit blanc, fouillis...).

Modèle du signal cible $\mathbf{d}$

Soit une cible définie comme étant un point s mobile à une vitesse v_t. Sa position est décrite dans l'équation (4.3) par le vecteur $\mathbf{u}_k(\phi_t, \theta_t)$ avec (ϕ_t, θ_t) l'azimut et l'élévation de la cible. À partir de ce vecteur de position, on peut calculer la fréquence *Doppler* de la cible :

$$f_t = \frac{2}{\lambda}\mathbf{u}_k(\phi_t, \theta_t)^T(\mathbf{v}_p - \mathbf{v}_t), \tag{4.6}$$

où $\mathbf{v}_p$ (définie dans l'équation (4.2)) et $\mathbf{v}_t$ sont les vecteurs vitesse de la plate-forme et de la cible. Nous utilisons le plus souvent la fréquence *Doppler* normalisée ω_t :

$$\omega_t = \frac{f_t}{f_r} \tag{4.7}$$

Le vecteur directionnel temporel $\mathbf{b}(\omega_t)$ est alors défini par :

$$\mathbf{b}(\omega_t) = \left[1\, e^{j2\pi\omega_t}\, \ldots\, e^{j2\pi(M-1)\omega_t}\right]^T . \tag{4.8}$$

Dans la dimension spatiale, nous définissons la fréquence spatiale de la cible ν_t :

$$\nu_t = \frac{d}{\lambda} cos\theta_t sin\phi_t \, , \tag{4.9}$$

et le vecteur directionnel spatial $\mathbf{a}(\nu_t)$:

$$\mathbf{a}(\nu_t) = \left[1\, e^{j2\pi\nu_t}\, \ldots\, e^{j2\pi(N-1)\nu_t}\right]^T \tag{4.10}$$

À partir de ces deux vecteurs, on définit alors le vecteur de taille $m = NM$ de direction spatio-temporelle $\mathbf{v}(\omega_t, \nu_t)$ comme étant la réponse d'une cible à une fréquence spatiale ν_t et à une fréquence *Doppler* normalisé ω_t :

$$\mathbf{v}(\omega_t, \nu_t) = \mathbf{b}(\omega_t) \otimes \mathbf{a}(\nu_t) \, . \tag{4.11}$$

La réponse de la cible peut donc s'écrire :

$$\mathbf{d} = \alpha_t \mathbf{v}(\omega_t, \nu_t) \, , \tag{4.12}$$

où α_t représente l'atténuation de l'onde après propagation et réflexion sur la cible.

Modèle du bruit total $\mathbf{c}_k + \mathbf{n}_k$

Nous supposons que le bruit (fouillis + bruit blanc) dans la cellule sous test et celui dans les données secondaires sont indépendants et partagent la même distribution statistique. Nous considérons le cas d'un fouillis hétérogène, qui sera modélisé par un processus CG (cf. section 1.2.3). Le modèle du bruit total correspond donc bien à celui considéré tout au long de ces travaux de thèse, présenté section 2.2, que nous justifions et rappelons tout de même brièvement ci dessous :

Bruit blanc

Le bruit blanc correspond ici au bruit causé par l'électronique du système et des capteurs (bruit thermique). Les réalisations de $\mathbf{n}_0$ (ou $\mathbf{n}_k$) sont modélisés par un vecteur aléatoire complexe gaussien $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_{NM})$.

Fouillis hétérogène

Le fouillis correspond ici uniquement à la réponse du sol. Le vecteur de fouillis $\mathbf{c}_0$ (ou $\mathbf{c}_k$) est la superposition d'un large nombre de points (appelés *clutter patchs*) distribués autour du radar à une distance fixe. Comme le sol est immobile, il existe une relation entre la fréquence spatiale et la fréquence *Doppler* normalisée :

$$\begin{array}{rcl} \nu_i & = & \frac{d}{\lambda} cos\theta_i sin\phi_i \ , \\ \omega_i & = & \beta \nu_i \ , \end{array} \tag{4.13}$$

où $\beta = \frac{2v_p}{df_r}$. Dans ce cadre, la réponse du fouillis peut s'écrire sous la forme suivante :

$$\mathbf{c}_k = \sum_{i=1}^{N_c} \gamma_{i,k} \mathbf{v}(\omega_i, \nu_i) \tag{4.14}$$

où N_c est le nombre de *clutter patch* et $\gamma_{i,k}$ est l'amplitude complexe aléatoire du *patch* i. Les $\gamma_{i,k}$ sont supposés décorrélés entre eux.

Comme dans les références [96, 50], nous considérons que le fouillis est hétérogène, c'est à dire que se puissance varie entre chaque cellule. Dans une telle situation, il est commun d'utiliser comme modèle les processus CG [127, 84]. Les processus $\mathbf{c}_0$ et $\mathbf{c}_k$ sont alors modélisés, conditionnellement à la texture, comme des vecteurs aléatoires complexes gaussiens $(\mathbf{c}_k|\tau_k) \sim \mathcal{CN}(\mathbf{0}, \tau_k \mathbf{\Sigma}_c)$ centrés de matrice de covariance $\tau_k \mathbf{\Sigma}_c$ avec $\mathbf{\Sigma}_c = E[\mathbf{c}_0 \mathbf{c}_0^H] = E[\mathbf{c}_k \mathbf{c}_k^H]$, et où la texture τ_k est une variable aléatoire positive. Dans le contexte du STAP, nous pouvons évaluer le rang du fouillis à l'aide de la formule de Brennan [18] qui assure une structure rang réduit pour la matrice de covariance du fouillis. Cette matrice de covariance $\mathbf{\Sigma}_c$ peut se décomposer à partir de sa SVD :

$$\mathbf{\Sigma}_c = \sum_{r=1}^{R} c_r \mathbf{v}_r \mathbf{v}_r^H \ , \tag{4.15}$$

avec l'hypothèse $R \ll M$.

Bruit total

Chaque échantillon est alors, conditionnellement à τ_k, distribué selon :

$$(\mathbf{z}_k|\tau_k) \sim \mathcal{CN}\left(\mathbf{0}, \tau_k\mathbf{\Sigma}_c + \mathbf{I}_{NM}\right). \tag{4.16}$$

La matrice de covariance totale du processus considéré est alors :

$$\mathbf{\Sigma} = \mathbb{E}(\tau)\mathbf{\Sigma}_c + \sigma^2\mathbf{I}_{NM}\ , \tag{4.17}$$

et est donc structurée de la forme rang faible plus identité.

4.2 Application basée sur l'estimation de la matrice de covariance : détection

Avant tout, notons que pour le STAP en conditions classiques, l'estimation du sous-espace fouillis est a priori suffisante : les traitements rang faible (étudiés dans la section suivante) s'avèrent généralement bien adaptés. Nous précisons donc que ni cette section, ni ce chapitre, ne sont destinés à comparer les traitements classiques et rang faibles entre eux. Le but de cette section est d'illustrer les performances des estimateurs robustes de matrices de covariances structurées développés dans cette thèse, au travers d'une application de détection. Par cette illustration nous voulons mettre en avant l'intérêt que représente ces estimateurs pour d'autres applications.

4.2.1 Problème considéré

Nous proposons d'illustrer les performances des estimateurs de matrice de covariance développés dans cette thèse au travers d'un problème de détection. Ce problème peut se formaliser à l'aide d'un test d'hypothèses binaires sur la donnée $\mathbf{z}_0$:

$$\begin{cases} H_0 : \mathbf{z}_0 = \mathbf{c}_0 + \mathbf{n}_0 & ; \quad \mathbf{z}_k = \mathbf{n}_k \ , \ \forall k \in [\![1, K]\!] \\ H_1 : \mathbf{z}_0 = \mathbf{d} + \mathbf{c}_0 + \mathbf{n}_0 & ; \quad \mathbf{z}_k = \mathbf{n}_k \ , \ \forall k \in [\![1, K]\!] \end{cases} \tag{4.18}$$

où la cible $\mathbf{d}$ et le bruit $\mathbf{c}_0 + \mathbf{n}_0$ suivent le modèle décrit section 4.1.

Nous étudierons les performances du détecteur Adaptive Normalized Matched Filter (ANMF) [67, 68] défini par :

$$\hat{\Lambda}(\hat{\mathbf{\Sigma}}) = \frac{|\mathbf{d}^H \hat{\mathbf{\Sigma}}^{-1} \mathbf{z}_0|^2}{|\mathbf{d}^H \hat{\mathbf{\Sigma}}^{-1} \mathbf{d}||\mathbf{z}_0^H \hat{\mathbf{\Sigma}}^{-1} \mathbf{z}_0|} \underset{H_0}{\overset{H_1}{\gtrless}} \delta_{\hat{\mathbf{\Sigma}}} \tag{4.19}$$

pour un estimateur de la matrice de covariance $\hat{\mathbf{\Sigma}}$ donné (calculé à l'aide des données secondaires $\{\mathbf{z}_k\}$). Nous étudierons les détecteurs construits à partir estimateurs suivants :

- RCML : l'estimateur décrit en section 1.4.3.
- SFPE : l'estimateur du point fixe régularisé, décrit en section 1.3.6. Comme il n'existe pas de règle de choix adaptatif "optimal" du paramètre de régularisation β pour le contexte que nous considérons, nous testerons différentes valeurs de ce paramètre entre 0.5 et 0.9 (valeurs ayant donné les meilleurs résultats dans ce contexte).
- EMV : le maximum de vraisemblance dans le contexte étudié. Pour le calculer, nous nous limiterons à l'algorithme EMV-2SR, car celui-ci présentait les meilleures performances à fort rapport fouillis à bruit (cf. chapitre 2).

4.2.2 Résultats de Simulations

Jeu de paramètres

Nous considérons la configuration STAP suivante : $M = 8$ capteurs, $N = 8$ impulsions. La fréquence centrale et la bande de fréquence sont respectivement $f_0 = 450$ MHz et $B = 4$ MHz. La vitesse du radar est $100\ m/s$. La distance entre les capteurs est $d = \frac{c}{2f_0}$, avec c la vitesse de la lumière. La fréquence de répétition des impulsions est $f_r = 600$ Hz. Le rang du fouillis est calculé selon la règle de Brennan [18] et vaut $r = 15 \ll 64$. La densité de probabilité de la texture est une loi Gamma

de paramètre d'intensité $\nu = 0.1$ et de paramètre d'échelle $\frac{1}{\nu}$, conduisant à un fouillis suivant une K-distribution. Le rapport fouillis à bruit est défini comme CNR $= \text{Tr}(\mathbf{\Sigma}_c)/R\sigma^2$ et le rapport signal à bruit comme SNR $= \text{norm}(\mathbf{d})/\sigma^2$. Nous fixons $\sigma^2 = 1$. La cible $\mathbf{d}$ a une célérité $V = 35\ m/s$ et se situe à $+10^\circ$ en azimut.

Méthodologie

Dans un premier temps, nous calculons la probabilité de fausse alarme (PFA) en fonction du seuil de détection pour 10^3 simulations Monte-Carlo sous l'hypothèse H_0. Ces résultats permettent de fixer le seuil de chaque détecteur (dépendant de l'estimateur de $\mathbf{\Sigma}$ utilisé) pour que celui-ci ait une PFA de 10^{-3}. Nous calculons ensuite la probabilité de détection (PD) des différents détecteurs en fonction du rapport signal à bruit de la cible, pour 10^3 simulations de Monte-Carlo sous l'hypothèse H_1. Cette méthodologie est appliquée pour différentes configurations des paramètres K, ν et CNR.

Résultats

Les figures 4.4 à 4.9 présentent les PFA en fonction du seuil du détecteur et les PD en fonction du SNR pour une PFA fixée à PFA$= 10^{-3}$. Les paramètres de chaque figure sont les suivants :

- Figures 4.4 et 4.5 : le nombre de données secondaires est $K = NM + 1$ (configuration sur-échantillonnée), le rapport fouillis à bruit est fort : CNR$=$ 30dB, le fouillis est modérément hétérogène ($\nu = 1$) pour la figure 4.4 et fortement hétérogène ($\nu = 0.1$) pour la figure 4.5.
- Figures 4.6 et 4.7 : le nombre de données secondaires est $K = 45$ (configuration sous-échantillonnée), le rapport fouillis à bruit est fort : CNR$=$ 30dB, le fouillis est modérément hétérogène ($\nu = 1$) pour la figure 4.6 et fortement hétérogène ($\nu = 0.1$) pour la figure 4.7.
- Figures 4.8 et 4.9 : le nombre de données secondaires est $K = 45$ (configuration sous-échantillonnée), le rapport fouillis à bruit est modéré : CNR$=$ 10dB, le fouillis est modérément hétérogène ($\nu = 1$) pour la figure 4.8 et fortement hétérogène ($\nu = 0.1$) pour la figure 4.9.

Les résultats montrent que pour presque toutes les configurations, l'estimateur développé au cours de cette thèse (EMV-2SR) conduit aux meilleures performances en termes de détection, c'est à dire que pour un SNR donné, il conduit à la plus grande probabilité de détection (sur les figures de droite). Dans certaines configuration ($e.g.$ figures 4.6 à 4.9) l'estimateur SFPE peut atteindre des performances proches de EMV-2SR et peut même le dépasser pour des faibles valeurs de rapport signal à bruit (comme en figure 4.7). Cependant, ces performances sont conditionelles au bon choix du paramètre de régularisation β. Nous soulignons qu'une méthode de choix adaptatif "optimal" de ce paramètre n'a pas encore été dérivée pour le modèle que nous considérons (fouillis hétérogène rang faible). On observe de plus que pour des configurations sur-échantillonnées, RCML offre de de bonnes performances en termes de détection, même si le bruit en présence est hétérogène (figures 4.4 et 4.5). Néanmoins, ces performances sont fortement dégradées quand le nombre de données est limité et que le bruit en présence est hétérogène (figures 4.6 à 4.9).

De plus, la figure 4.10 illustre la robustesse de la méthode EMV-2SR à une mauvaise évaluation du rang. Cette figure montre que la courbe de PFA-seuil associée au détecteur ne varie pas dramatiquement quand le rang est légèrement mal évalué. On remarque cependant que si le rang du fouillis est fortement

sous évalué (ici, 10 au lieu de 15) les performances du détecteur peuvent être fortement détériorées. En définitive, il semble préférable de sur-évaluer légèrement le rang.

4.2.3 Résultats sur données réelles

Les données STAP new-CELAR [16] sont fournies par un simulateur de l'agence DGA-MI qui nous permet de synthétiser, en configuration side looking, le datacube STAP à partir de mesures réelles SAR très haute résolution du radar RAMSES [34]. Le fouillis est donc issu de données réelles, cependant, les cibles ajoutées sont synthétiques.

Le nombre de capteurs est $N = 4$ et le nombre d'impulsions est $M = 64$, la taille des données est donc $QP = 256$. La fréquence centrale et la bande de fréquence sont respectivement $f_0 = 10$ GHz et $B = 5$ MHz. La vitesse du radar est 100 m/s. La distance entre les capteurs est $d = 0.3$. La fréquence de répétition des impulsions est $f_r = 1$ kHz. Le rang du fouillis est calculé selon la règle de Brennan [18] et vaut $r = 45 \ll 256$. Le rapport fouillis à bruit est évalué à 20dB et le fouillis est homogène, c'est à dire

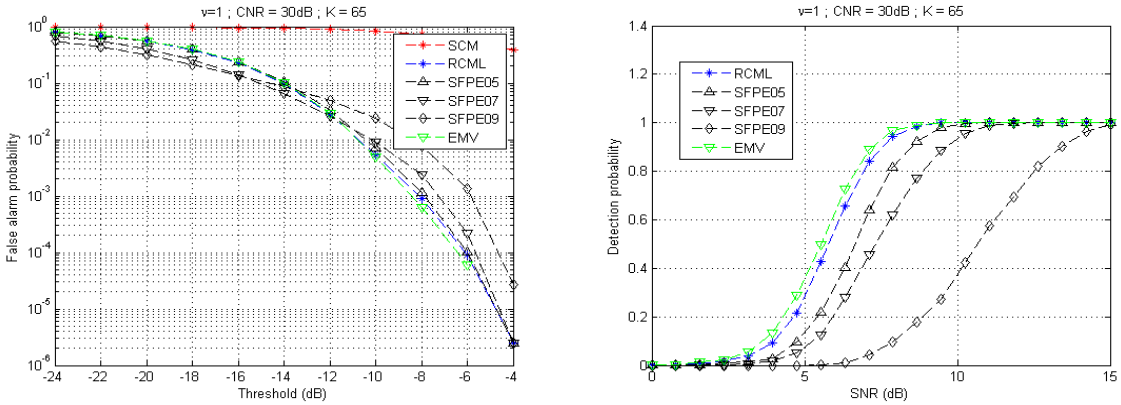

FIGURE 4.4 – Gauche : probabilité de fausse alarme en fonction du seuil. Droite : probabilité de détection en fonction du SNR pour PFA= 10^{-3}. $K = NM + 1$, $\nu = 1$, CNR= 30dB.

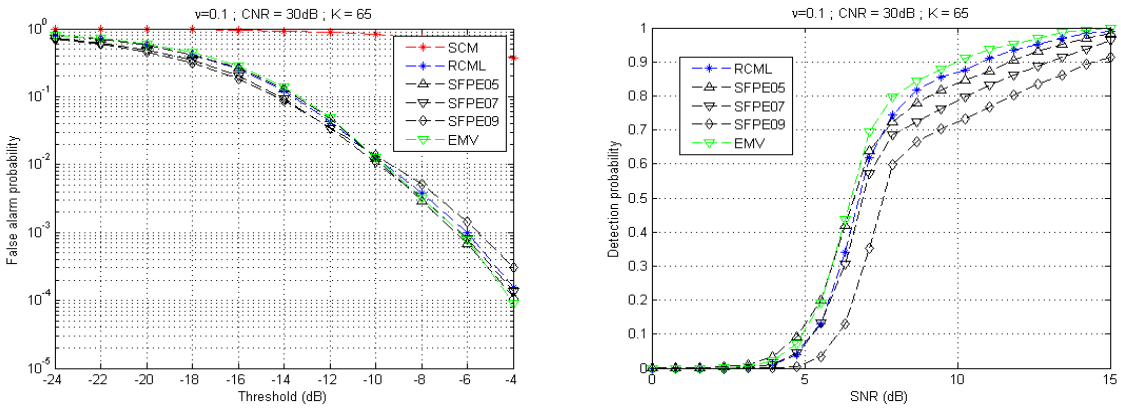

FIGURE 4.5 – Gauche : probabilité de fausse alarme en fonction du seuil. Droite : probabilité de détection en fonction du SNR pour PFA= 10^{-3}. $K = NM + 1$, $\nu = 0.1$, CNR= 30dB.

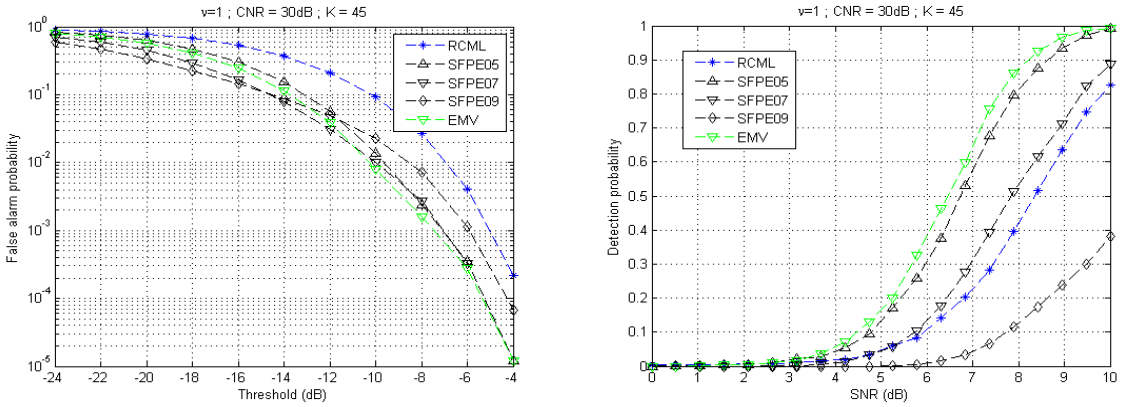

FIGURE 4.6 – Gauche : probabilité de fausse alarme en fonction du seuil. Droite : probabilité de détection en fonction du SNR pour PFA= 10^{-3}. $K = 3R$, $\nu = 1$, CNR= 30dB.

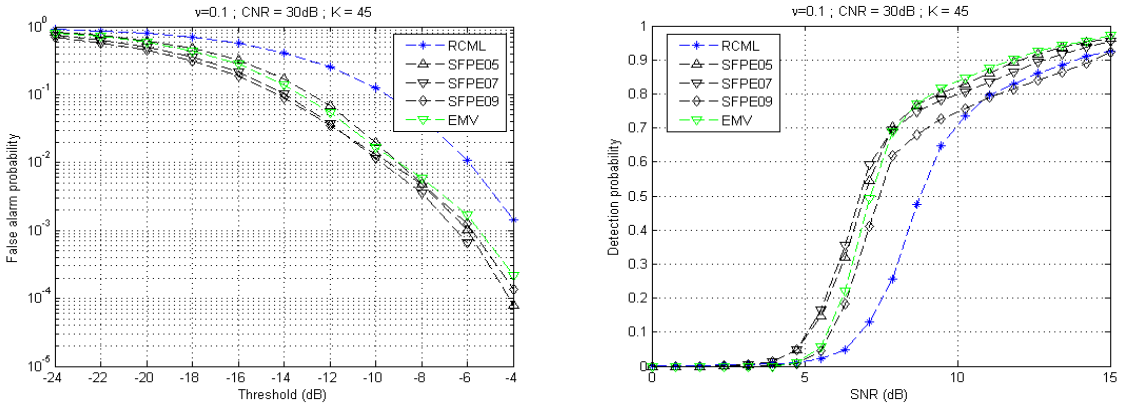

FIGURE 4.7 – Gauche : probabilité de fausse alarme en fonction du seuil. Droite : probabilité de détection en fonction du SNR pour PFA= 10^{-3}. $K = 3R$, $\nu = 0.1$, CNR= 30dB.

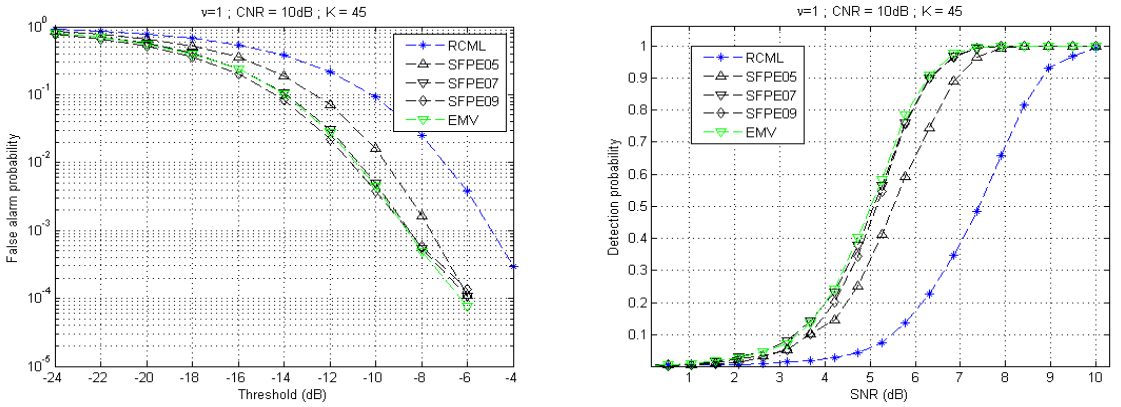

FIGURE 4.8 – Gauche : probabilité de fausse alarme en fonction du seuil. Droite : probabilité de détection en fonction du SNR pour PFA= 10^{-3}. $K = 3R$, $\nu = 1$, CNR= 10dB.

proche d'une distribution gaussienne. Dans la cellule sous test, une cible à (4 m/s, 0 deg) est présente avec un rapport signal à fouillis de -5dB. Le nombre maximum de données secondaires (taille du data-

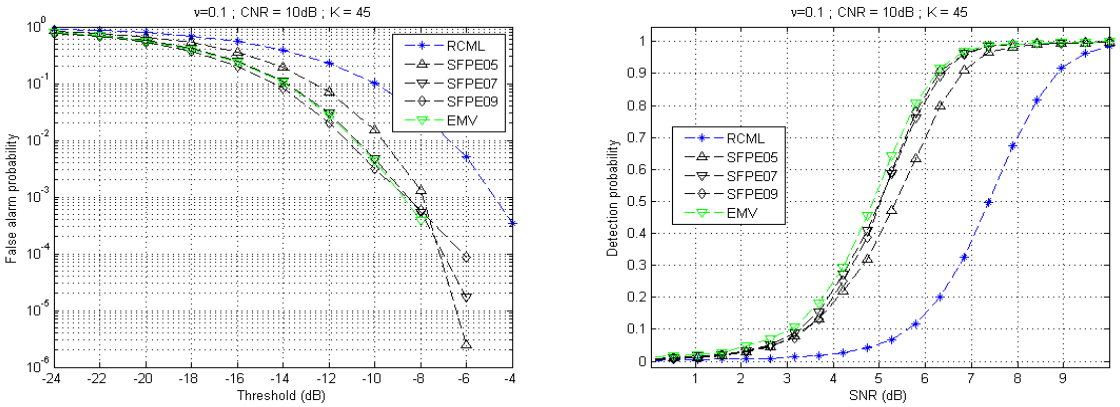

FIGURE 4.9 – Gauche : probabilité de fausse alarme en fonction du seuil. Droite : probabilité de détection en fonction du SNR pour PFA= 10^{-3}. $K = 3R$, $\nu = 0.1$, CNR= 10dB.

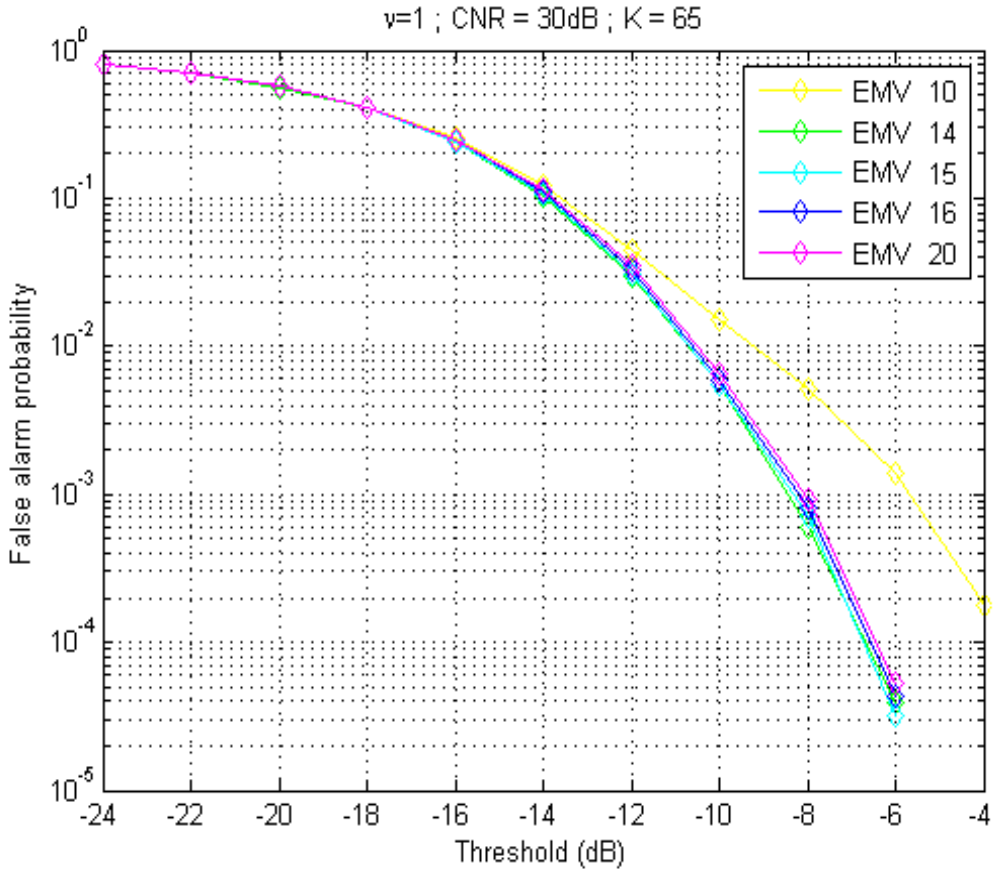

FIGURE 4.10 – Probabilité de fausse alarme en fonction du seuil pour l'estimateur EMV calculé avec différents rangs R autour de la vraie valeur $R = 15$. $K = NM + 1$, $\nu = 1$, CNR= 30dB.

cube) est $K = 408$ et (4 cellules de garde sont retirées autour de la cellule cible afin d'éviter d'inclure dans les données secondaires des réponses de la cible qui seraient étalées dans les cases voisines). Nous construisons les estimateurs de la matrice de covariance à partir de ces données secondaires, supposées $i.i.d$ et non contaminées par des cibles.

Pour traiter ces données, nous considérons les détecteurs suivants :
- $\hat{\Lambda}_{SCM}$: le détecteur construit avec la SCM. Pour les configurations sous-échantillonnées $K < M$, l'inverse de la SCM n'est pas calculable et le détecteur correspondant n'est pas défini. Dans ce cas, on remplace simplement $\hat{\Lambda}_{SCM}$ par $\hat{\Lambda}_{\mathbf{I}}$ (le détecteur construit avec la matrice identité) à titre d'illustration.
- $\hat{\Lambda}_{RCML}$: le détecteur construit avec l'estimateur RCML.
- $\hat{\Lambda}_{SFPE}$: le détecteur construit avec l'estimateur FPE régularisé. Notons que les performances de ce détecteur ont déjà été étudiés sur ces données dans la référence [90]. La valeur du paramètre de régularisation β donnant les meilleurs résultats se situe pour ces données entre 0.7 et 0.8 (variant selon K), nous la fixons donc arbitrairement ici à $\beta = 0.75$ pour tous les tests.
- $\hat{\Lambda}_{EMV}$: le détecteur basé sur l'EMV du modèle considéré, calculé avec l'algorithme EMV-2SR. Notons que les résultats sont sensiblement (visuellement) identiques avec les autres algorithmes.

Les figures que nous allons analyser présentent des sorties de détecteurs, c'est à dire la valeur du détecteur ANMF évaluée sur une grille de vecteurs cibles tests, dont les paramètres balayent différentes valeurs d'angles et de vitesse (le test de détection est donc effectué en chaque pixel sur une cible différente). Précisons que ces résultats présentent des réalisations uniques et ont donc principalement valeur d'illustration. Néanmoins, cette application des méthodes développées sur données réelles vient confirmer et renforcer les conclusions de l'étude quantitative faite en section précédente.

La figure 4.11 montre les sorties des détecteurs pour une configuration sous-échantillonnée, avec $K = 100$ (cependant $K > 2R$). Pour, $\hat{\Lambda}_I$ le bruit n'est pas blanchit : on observe en effet que les fortes valeurs du détecteur se situent sur la diagonale (réponse du fouillis) et que la cible n'est pas distinguable de ce fouillis. Les détecteurs $\hat{\Lambda}_{RCML}$ et $\hat{\Lambda}_{SFPE}$ permettent bien d'effectuer une détection de la cible, cependant $\hat{\Lambda}_{EMV}$ semble offrir la meilleure atténuation de l'interférence, ce qui se traduit par la disparition de la réponse du fouillis sur la diagonale de la sortie du détecteur. On peut aussi observer que les détecteurs construits à partir d'estimateurs structurés rang faible ($\hat{\Lambda}_{RCML}$ and $\hat{\Lambda}_{EMV}$) effectuent un meilleur rejet du fouillis, ce qui illustre l'intérêt de prendre en considération l'a priori de structure.

La figure 4.12 montre les sorties des détecteurs pour une configuration sur-échantillonnée, avec $K = 300$. Les mêmes conclusions que précédemment peuvent être tirées : le détecteur $\hat{\Lambda}_{SCM}$ ne semble pas permettre une détection de la cible. Les détecteurs $\hat{\Lambda}_{RCML}$ et $\hat{\Lambda}_{EMV}$ effectuent un meilleur rejet du fouillis. On remarque de plus que $\hat{\Lambda}_{RCML}$ et $\hat{\Lambda}_{EMV}$ on des performances similaires. Ceci pouvait être attendu car le fouillis présent dans ces données est pratiquement gaussien.

La figure 4.13 teste la robustesse des estimateurs à une potentielle contamination. Elle présente les sorties des différents détecteurs pour $K = 100 + 1$, où la donnée ajoutée contient aussi une cible ayant

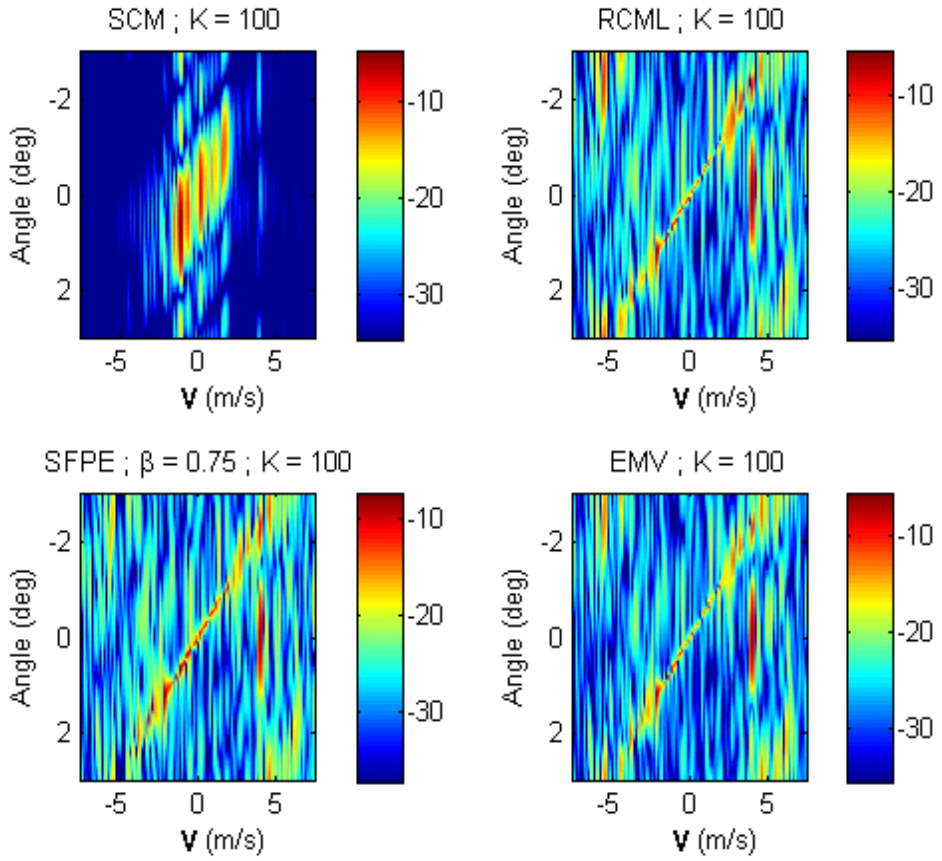

FIGURE 4.11 – Sortie des détecteurs, de gauche à droite : Λ_{SCM}, Λ_{RCML}, Λ_{SFPE} and Λ_{EMV}. $K = 100$.

les mêmes caractéristiques que celle présente dans la case sous test. On observe, en comparaison avec la figure 4.13, que cette contamination impacte les performances des détecteurs. Néanmoins, $\hat{\Lambda}_{EMV}$ semble mieux résister à cette contamination, ce qui illustre un potentiel intérêt de la méthode.

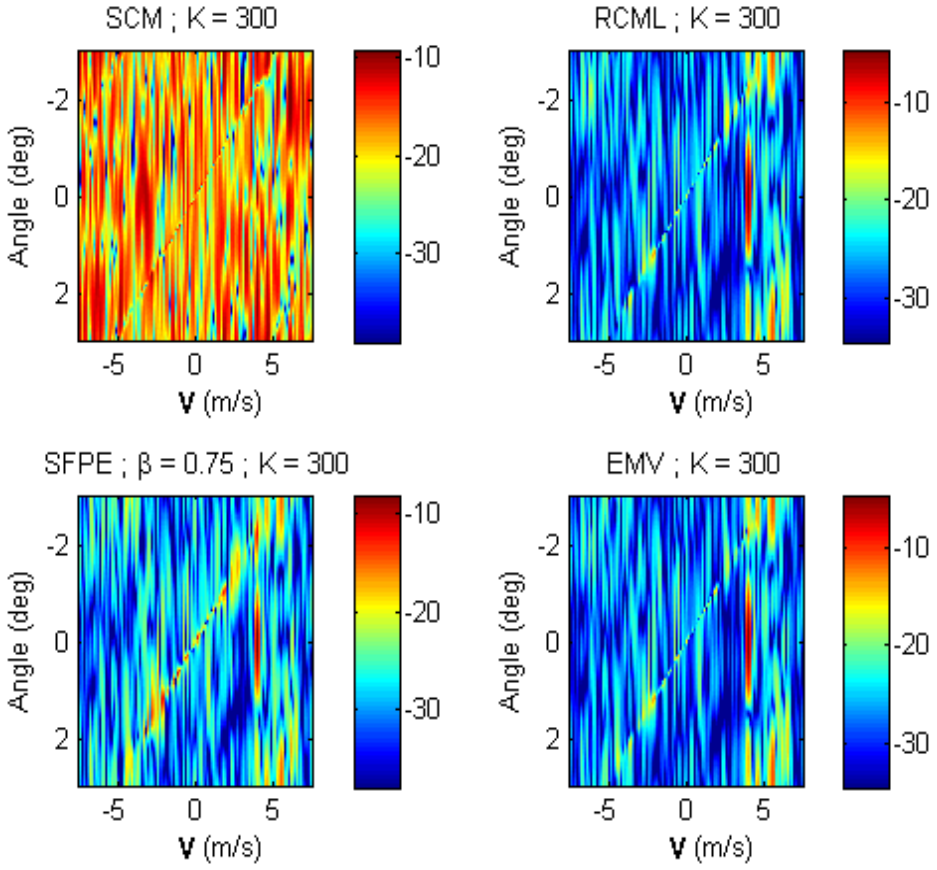

FIGURE 4.12 – Sortie des détecteurs, de gauche à droite : Λ_{SCM}, Λ_{RCML}, Λ_{SFPE} and Λ_{EMV}. $K = 300$.

4.3 Application basée sur l'estimation du sous-espace fouillis : filtrage rang faible

4.3.1 Problème considéré

Nous proposons maintenant d'illustrer les performances des estimateurs de sous-espace fouillis développés au travers d'un problème de filtrage STAP rang faible.

Le filtre STAP (non adaptatif) optimal en termes de rapport signal à bruit de sortie, noté $\mathbf{w}_{opt}$, est construit à l'aide de la matrice de covariance totale du bruit $\mathbf{\Sigma}_{tot}$ et le steering vecteur $\mathbf{d}$:

$$\mathbf{w}_{opt} = \mathbf{\Sigma}_{tot}^{-1}\mathbf{d} \, . \tag{4.20}$$

Classiquement, les filtres adaptatifs sont construits avec

$$\hat{\mathbf{w}} = \hat{\mathbf{\Sigma}}^{-1}\mathbf{d} \, , \tag{4.21}$$

où $\hat{\mathbf{\Sigma}}$ un estimateur de la matrice de covariance. Comme détaillé précédemment, pour un fouillis de rang

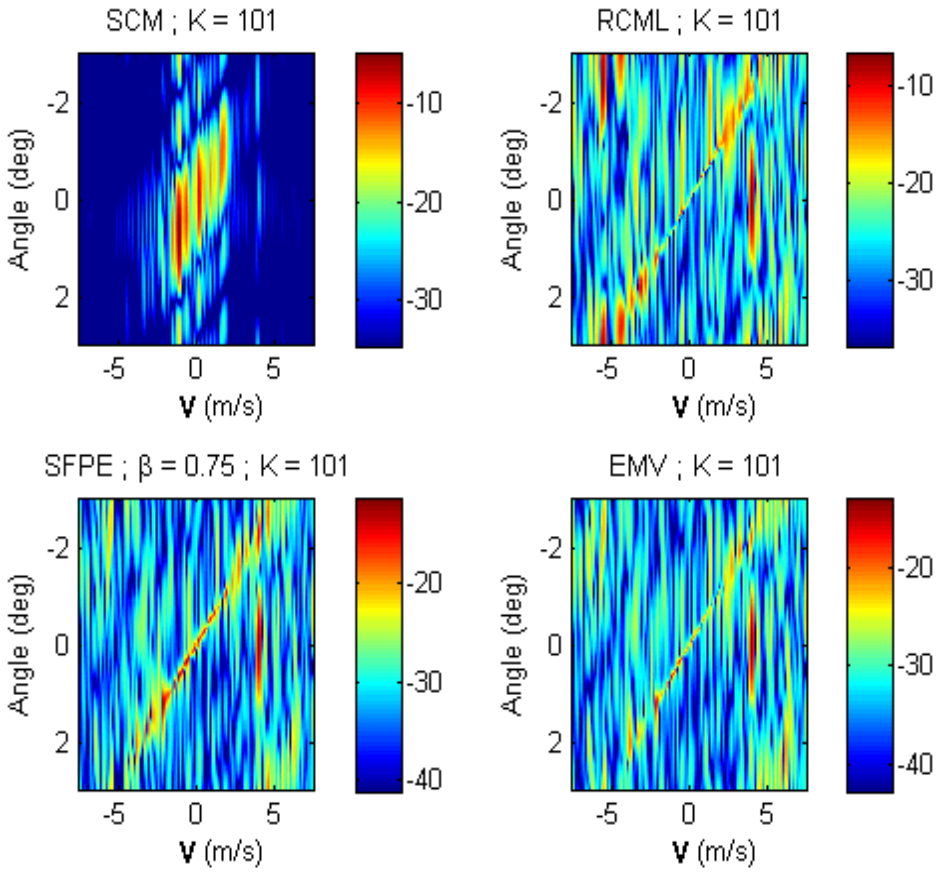

FIGURE 4.13 – Sortie des détecteurs, de gauche à droite : Λ_{SCM}, Λ_{RCML}, Λ_{SFPE} and Λ_{EMV}. $K = 100$. Une donnée corrompue par la cible.

faible, la matrice de covariance se décompose comme :

$$\mathbf{\Sigma}_{tot} = \sum_{r=1}^{R}(c_r + \sigma^2)\mathbf{v}_r\mathbf{v}_r^H + \sum_{r=R+1}^{M}\sigma^2\mathbf{v}_r\mathbf{v}_r^H \ . \tag{4.22}$$

Dans ce contexte, il possible de construire que le filtre adaptatif rang faible [66, 54] :

$$\mathbf{w}_{lr} = \left(\mathbf{I} - \hat{\mathbf{\Pi}}_c\right)\mathbf{d} \ , \tag{4.23}$$

où $\hat{\mathbf{\Pi}}_c$ est un estimateur du projecteur sur le sous-espace fouillis $\mathbf{\Pi}_c = \sum_{r=1}^{R}\mathbf{v}_r\mathbf{v}_r^H$ (cf. Chapitre 3). Ce filtre est asymptotiquement sous-optimal, mais offre de meilleurs performances à distance finie quand le nombre d'échantillons est faible (voire inférieur à la taille des données).

Dans cette section nous étudierons les performances des filtres rang faibles construits à partir de différents estimateurs du projecteur sur le sous-espace fouillis. Pour cela, nous utiliserons le critère SINR-Loss [120] : la perte en rapport signal à bruit de sortie du filtre adaptatif comparé au filtre optimal.

Le SINR-Loss est défini comme :

$$\rho_{\hat{\mathbf{\Pi}}_c} = \mathbb{E}\left(\frac{SINR_{out}}{SINR_{max}}\right) = \mathbb{E}\left(\frac{(\mathbf{d}^H\hat{\mathbf{\Pi}}_c^{\perp}\mathbf{d})^2}{\left(\mathbf{d}^H\hat{\mathbf{\Pi}}_c^{\perp}\mathbf{\Sigma}_{tot}\hat{\mathbf{\Pi}}_c^{\perp}\mathbf{d}\right)\left(\mathbf{d}^H\mathbf{\Sigma}_{tot}^{-1}\mathbf{d}\right)}\right) , \tag{4.24}$$

pour un estimateur $\hat{\mathbf{\Pi}}_c$ donné.

Remarque : Dans cette section, nous ne considérons pas le traitement de détection rang faible afin d'éviter la redondance avec l'étude précédente. Néanmoins, nous précisons que nos tests suggèrent que l'étude de filtres ou de détecteurs conduisent dans la grande majorité des cas à des conclusions similaires quant aux performances des estimateurs de sous-espaces.

Nous étudierons les filtres rang faibles construits à partir estimateurs suivants :

- SCM : le projecteur estimé au travers des R vecteurs propres dominants de la Sample Covariance Matrix, décrite en section 1.3.2.
- SFPE : le projecteur estimé au travers des R vecteurs propres dominants de l'estimateur FPE régularisé, décrit en section 1.3.6. Comme il n'existe pas de règle de choix adaptatif "optimal" du paramètre de régularisation β pour le contexte que nous considérons, nous testerons différentes valeurs de ce paramètre : $\beta_1 = \max(1 - K/M + \epsilon\ ,\ 0)$, à savoir le β minimum assurant l'existence de l'estimateur et donnant l'estimateur du point fixe si $K > M$, $\beta_2 = (\beta_1 + \beta_3)/2$ et $\beta_3 = 1 - \epsilon$. Nous fixons $\epsilon = 10^{-2}$.
- AEMV : le projecteur construit avec maximum de vraisemblance approché obtenu à la section 3.3.2. Nous nous limitons à cet estimateur car il a été montré chapitre 3 que ses performances étaient similaires à celles de l'EMV exact.
- LR-FPE : l'estimateur du projecteur dérivé via les itérations définies en section 3.2.
- R-AMEV : l'estimateur du maximum de vraisemblance du modèle modifié, prenant en compte une potentielle contamination orthogonale au fouillis (décrit à la section 3.4.2).

4.3.2 Résultats de simulations

Jeu de paramètres

Nous considérons la configuration STAP suivante : $N = 8$ capteurs, $M = 8$ impulsions. La fréquence centrale et la bande de fréquence sont respectivement $f_0 = 450$ MHz et $B = 4$ MHz. La vitesse du radar est $100\ m/s$. La distance entre les capteurs est $d = \frac{c}{2f_0}$, avec c la vitesse de la lumière. La fréquence de répétition des impulsions est $f_r = 600$ Hz. Le rang du fouillis est calculé selon la règle de Brennan [18] et vaut $r = 15 \ll 64$. La densité de probabilité de la texture est une loi Gamma de paramètre d'intensité $\nu = 0.1$ et de paramètre d'échelle $\frac{1}{\nu}$, conduisant à un fouillis suivant une K-distribution. Le rapport fouillis à bruit est défini comme CNR $= \mathrm{Tr}(\mathbf{\Sigma}_c)/R\sigma^2$ et le rapport signal à bruit comme SNR$=$ norm$(\mathbf{d})/\sigma^2$. Nous fixons $\sigma^2 = 1$. La cible $\mathbf{d}$ a une célérité $V = 35\ m/s$ et se situe à $+10^\circ$ en azimut.

Résultats

Les figures 4.14 et 4.15 présentent le SINR-Loss des différents filtres rang faibles en fonction de K. Les paramètres du fouillis sont les suivants :

- Pour la figure 4.14 le fouillis est modérément hétérogène ($\nu = 1$) et les résultats sont obtenus pour différents rapports fouillis à bruit 10dB, 20dB, et 30dB.
- Pour la figure 4.15 le fouillis est fortement hétérogène ($\nu = 0.1$) et les résultats sont obtenus pour différents rapports fouillis à bruit 10dB, 20dB, et 30dB.

Pour ces simulations, nous ne représentons pas les résultats de l'estimateur R-AEMV car ses performances sont strictement identiques à AEMV (comme cela a pu être observé chapitre 3). La différence entre ces deux estimateurs sera illustrée par la suite.

Tout d'abord, on remarque que pour toutes les configurations, l'estimateur AEMV conduit aux meilleures performances en termes de SINR-Loss (la valeur négative observée la plus haute), c'est à dire le moins de perte en SINR de sortie par comparaison au filtre non adaptatif optimal. Pour des conditions standards (rapport fouillis à bruit de 20dB et 30dB et $\nu = 1$), la SCM est un estimateur proche de l'EMV (comme cela à aussi pu être observé au chapitre 3), et atteint donc les mêmes performances en termes de filtrage. On observe que pour ces conditions, la perte classique de -3dB en SINR du filtre construit à partir de SCM est atteinte pour $K \simeq 2R$ échantillons, comme théoriquement prouvé dans [54, 50]. Les performances de SFPE dépendent fortement du paramètre de régularisation utilisé. On remarque que le SINR-Loss diminue si β augmente, ce qui est normal car un β plus grand implique que l'estimateur est plus fortement biaisé. En condition standards, SFPE avec le β minimum a des performances proches de celles de la SCM. En conditions non standard (rapports fouillis à bruit moyens ou faibles, fouillis fortement hétérogène), on observe que les performances de SFPE diminuent grandement par rapport aux autres estimateurs. Au contraire AEMV offre une meilleure résistance à ces conditions et surpasse même $\hat{\mathbf{\Pi}}_{SCM}$. Comme observé chapitre 3, l'estimateur LR-FPE conduit aux plus mauvaises performances, cependant son intérêt en termes de robustesse sera illustré dans les simulations suivantes.

Les figures 4.16 et 4.17 étudient la robustesse des filtres à la contamination des données par la présence de cible dans les données secondaires.

En figure 4.16, plusieurs données sont contaminées selon :

$$\mathbf{z}_k = \mathbf{c}_k + \mathbf{n}_k + \alpha\mathbf{d} \ , \ \ k \in 1, ...K_o \tag{4.25}$$

où $\mathbf{d}$ est la cible (normalisée) que l'on cherche à filtrer, ayant une puissance (notée ONR pour "Outlier to Noise Ratio") de ONR $= \alpha/\sigma^2$. On fixe ONR=10dB et on observe l'evolution du SINR-Loss des filtres adaptatifs en fonction du pourcentage de données contaminées. On constate que plus il y a de données contaminées, plus les performances des filtres décroissent. Tous les estimateurs sont impactés identiquement, à l'exception de R-AEMV qui montre ici une meilleure résistance à l'introduction de contamination multiples.

En figure 4.17, une donnée (arbitrairement $k = 1$) seulement est contaminée sous la forme :

$$\mathbf{z}_1 = \mathbf{c}_1 + \mathbf{n}_1 + \alpha\mathbf{d} \tag{4.26}$$

où $\mathbf{d}$ est la cible (normalisée) que l'on cherche à filtrer. On observe l'évolution du SINR-Loss des filtres adaptatifs en fonction de la puissance de la contamination : ONR $= \alpha/\sigma^2$. Dans cette figure on observe que les performances des filtres décroissent à partir d'une puissance "seuil". On remarque que pour l'estimateur R-AEMV, ce seuil est plus fort que pour SCM et AEMV, ce qui indique une

meilleure robustesse à l'introduction de cible. Les estimateurs SFPE et LR-FPE n'apparaissent pas (ou peu) impactés, quelle que soit la puissance de la contamination, ce qui semble être une propriété de robustesse intéressante.

4.3.3 Résultats sur données réelles

Dans cette partie proposons d'illustrer la robustesse de l'estimateur LR-FPE sur les données CELAR, présentées en section 4.2.3. La figure 4.18 montre les sorties des filtres rang faible adaptatifs construits à partir des estimateurs AEMV, SFPE (avec paramètre de régularisation minimal), R-AEMV et LR-FPE. On constate que ces sorties sont presque identiques pour tous les estimateurs considérés. La figure 4.19 présente ces résultats pour la même configuration avec deux données contaminées. On observe que les données contaminées viennent dégrader les performances des filtres basés sur AMV, R-AEMV et SFPE. En revanche, le filtre construit à partir de LR-FPE résiste à cette corruption et permet de distinguer la cible. Notons que ce type de résultat a aussi été observé sur de nombreuses sorties de filtres simulées, suggérant une bonne robustesse de cet estimateur malgré les mauvaises performances constatées en termes de SINR-Loss.

4.4 Synthèse du chapitre 4

Dans ce chapitre, nous avons décrit le système STAP pour un radar aéroporté. Nous avons présenté le modèle de signaux observés par ce système, qui correspond bien à celui étudié au cours des chapitres précédents de cette thèse. C'est pourquoi nous avons ensuite étudié les performances des estimateurs nouvellement développés au travers de deux traitements.

Dans un premier temps, nous nous sommes intéressés aux performances du détecteur ANMF construit à partir de divers estimateurs de la matrice de covariance. Les simulations sur données réelles ou de synthèse, ont montré que l'EMV du modèle considéré (calculé à l'aide des algorithmes développés au chapitre 2) conduisait aux meilleures performances en comparaison de l'état de l'art. Ceci montre l'intérêt de prendre en compte à la fois l'hétérogénéité du bruit et l'a priori de structure dans le processus d'estimation. Nous avons aussi illustré une certaine robustesse des méthodes proposées à un mauvais a priori de modèle. Le détecteur basé sur l'EMV ne se montre, en effet, pas dramatiquement impacté par une mauvaise évaluation du rang du fouillis (excepté si celui ci est fortement sous évalué).

Ensuite, nous avons étudié les performances des filtres rang faible construits à partir de différents estimateurs du projecteur sur le sous-espace fouillis. Nous avons observé que les estimateurs AEMV et R-AEMV (développés au chapitre 3) conduisent aux meilleures performances en termes de SINR-Loss. Par ailleurs, nous nous sommes de plus intéressés au problème de robustesse à la contamination des données par l'introduction de cibles. Dans ce cas de figure, l'estimateur LR-FPE a montré une grande robustesse malgré ses mauvaises performances en termes de SINR-Loss. Nous avons aussi observé que l'estimateur R-AEMV propose un bon compromis "performances-robustesse" puisqu'il résiste mieux à la contamination qu'AEMV (ou l'EMV exact), tout en atteignant les mêmes performances sans contamination.

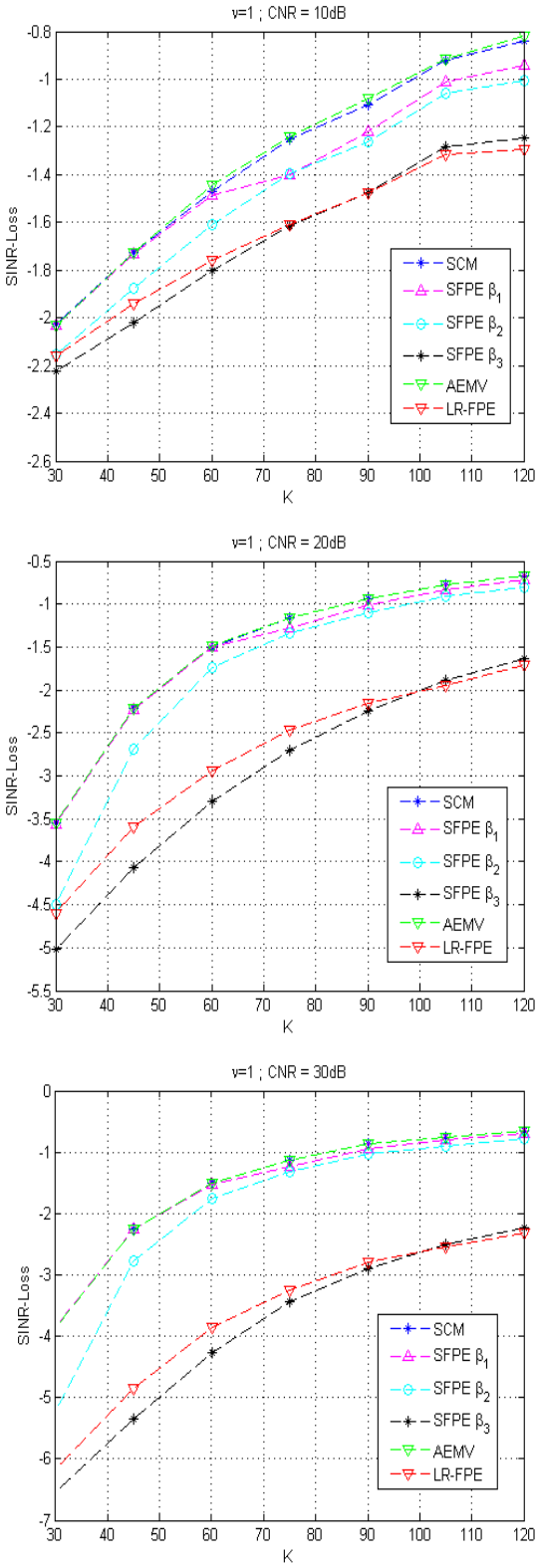

FIGURE 4.14 – SINR-Loss moyen des filtres rang faible construits à partir de différents estimateurs de projecteurs, en fonction de K. De haut en bas : CNR = 10dB, CNR = 20dB, CNR = 30dB. Fouillis modérément hétérogène $\nu = 1$.

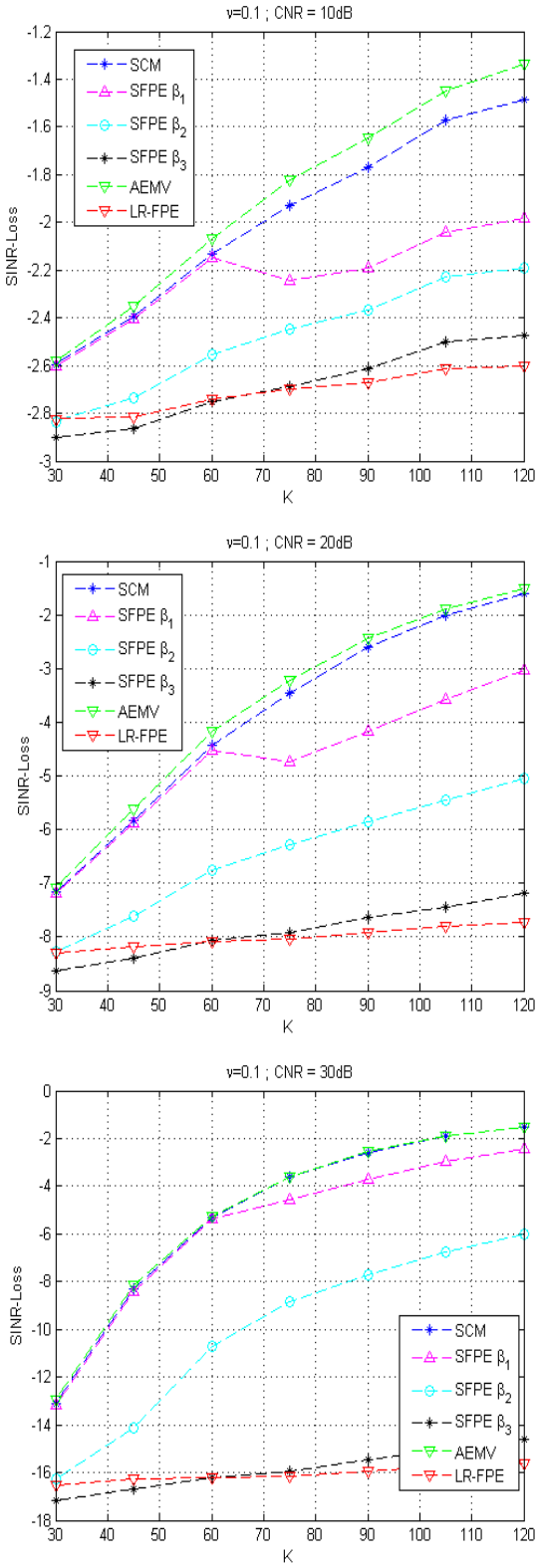

FIGURE 4.15 – SINR-Loss moyen des filtres rang faible construits à partir de différents estimateurs de projecteurs, en fonction de K. De haut en bas : CNR = 10dB, CNR = 20dB, CNR = 30dB. Fouillis fortement hétérogène $\nu = 0.1$.

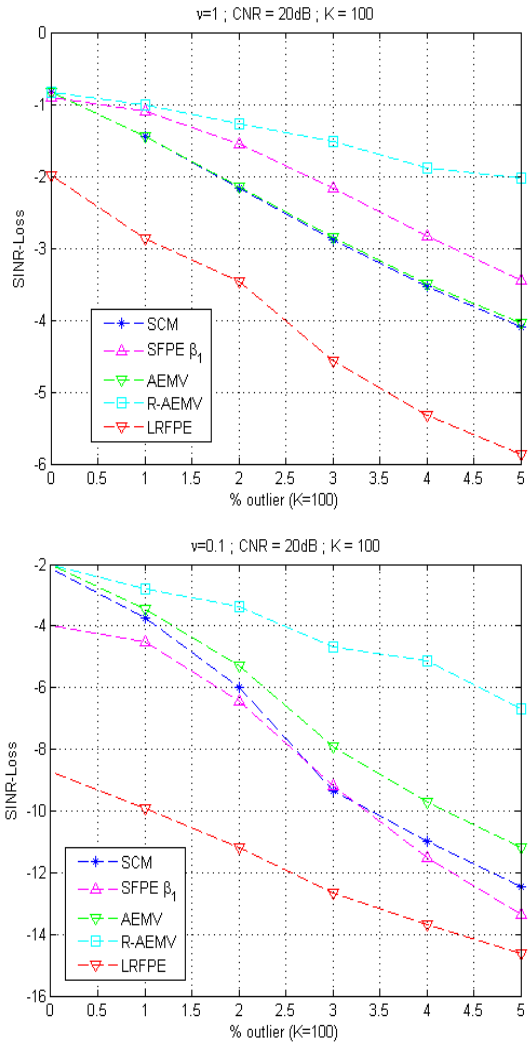

FIGURE 4.16 – SINR-Loss moyen des filtres rang faible construits à partir de différents estimateurs de projecteurs, en fonction du pourcentage de données corrompues $K = 100$ avec une cible de puissance ONR= 10dB et CNR= 20dB. Haut : $\nu = 1$, bas $\nu = 0.1$.

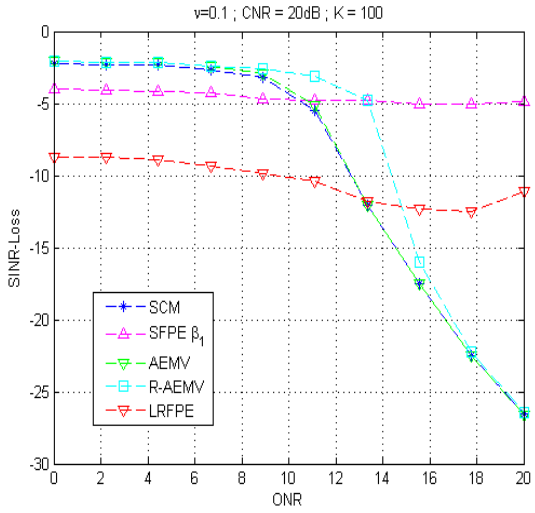

FIGURE 4.17 – SINR-Loss des filtres rang faible construits à partir de différents estimateurs de projecteurs, en fonction de de la puissance de l'outlier pour $K = 100$, $\nu = 0.1$ et CNR$=$ 20dB.

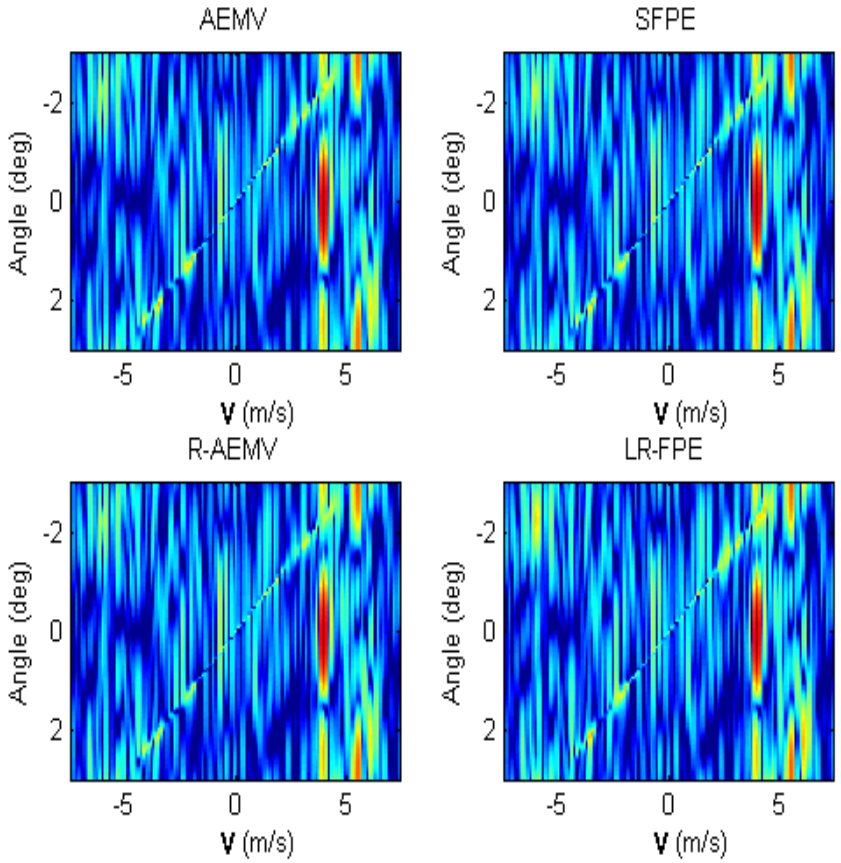

FIGURE 4.18 – Sortie des filtres rang faible adaptatifs sur les données STAP CELAR. $K = 400$ données.

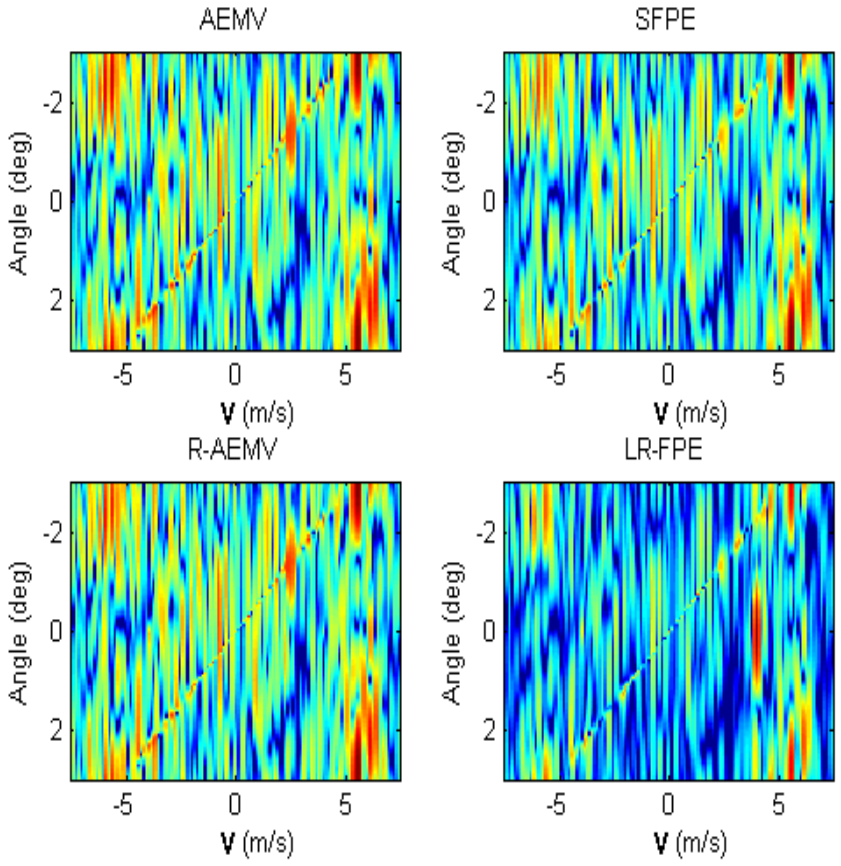

FIGURE 4.19 – Sortie des filtres rang faible adaptatifs sur les données STAP CELAR. $K = 400$ données, 2 données contaminées.

Conclusion

Synthèse

Dans ces travaux, nous nous sommes intéressés au problème d'estimation de la matrice de covariance (ou de son sous-espace propre dominant). Celle-ci joue, en effet, un rôle récurrent et prépondérant dans les traitements actuels, qui nécessitent la connaissance de la statistique d'ordre deux du milieu observé. Cette matrice étant en pratique inconnue, elle doit être estimée pour être ensuite utilisée dans des traitements dits adaptatifs. Les performances de ces traitements reposent donc directement sur la précision de l'estimation de la matrice de covariance.

La motivation à développer de nouveaux estimateurs vient de deux constatations liées à des problématiques actuelles. D'une part, les signaux mesurés par les systèmes modernes ne vérifient plus le modèle classiquement admis de données gaussiennes. Il est désormais nécessaire de tenir compte de l'hétérogénéité et parfois de la corruption des échantillons afin d'effectuer des traitements valides. D'autre part de nombreuses applications utilisent des données de grande dimension et font face au problème du nombre d'échantillons limité, rendant l'estimation de la matrice de covariance difficile.

Pour répondre à la première problématique, le modèle des distributions elliptiques, à attiré considérablement d'attention ces dernières années. Dans ce cadre, les M-estimateurs (Maximums de Vraisemblance généralisés) offrent une solution robuste et souvent appropriée. Une possible solution à la seconde problématique est d'effectuer des considérations en amont sur le système afin de dégager un a priori sur la structure de la matrice de covariance. L'estimation structurée permet en effet de dériver des traitements adaptatifs qui nécessitent moins de données secondaires que les traitements classiques pour des performances similaires. La problématique que nous considérons se situe à l'intersection de ces deux solutions : l'estimation robuste structurée est encore un domaine en cours d'exploration.

Dans ce travail, nous nous sommes tout particulièrement intéressés à des modèles où les signaux sont contenus dans un sous-espace de dimension inférieure aux données. Cette hypothèse est en effet vérifiée dans de nombreuses applications où le nombre de sources observées est restreint (estimation de direction d'arrivée, imagerie hyperspectrale…), ainsi que pour les mesures de fouillis (réponse de l'environnement) radar. La matrice de covariance de ce type de données présente alors une structure de rang faible. C'est pourquoi dans cette thèse, nous avons considéré le problème d'estimation robuste de matrices de covariances satisfaisant cette structure.

Le chapitre 1 a présenté un état de l'art sur les distributions elliptiques et les estimateurs robustes de la matrice de covariance. Il a ensuite décrit un bref état de l'art de l'estimation structurée de la matrice de covariance. Cette introduction a permis d'introduire la problématique considérée : celle de l'estimation de la matrice de covariance en contexte hétérogène rang faible.

Dans le chapitre 2, nous avons présenté le modèle statistique considéré au cours de la thèse : celui de sources hétérogènes ayant une matrice de covariance rang faible noyées dans un bruit blanc gaussien. Ce modèle est, par exemple, fortement justifié pour des applications de type radar. Nous avons dérivé l'expression de l'EMV de la matrice de covariance pour ce contexte. Cette expression n'ayant pas une forme explicite et analytique, nous avons développé différents algorithmes afin de l'obtenir efficacement.

Au cours du chapitre 3, nous nous sommes focalisés sur la problématique d'estimation du projecteur sur le sous-espace fouillis. En effet, pour le modèle considéré, il est possible de décomposer le signal sur deux sous-espaces orthogonaux à l'aide de la décomposition en valeurs singulières de la matrice de covariance. Les projecteurs orthogonaux sur chacun de ces sous-espaces peuvent alors être construits, permettant de développer des méthodes dites de rang faible (ou à sous-espace) telles que l'annulation d'interférence. Il a été montré que ces méthodes permettent d'obtenir des performances équivalentes à celles des traitements classiques tout en réduisant significativement le nombre d'échantillons nécessaire. Au moyen de diverses relaxations ou modifications du modèle initial, nous avons développé des estimateurs directs de projecteur. Ces estimateurs, algorithmiquement plus simples à obtenir, se sont montrés toutefois aussi performants que l'EMV exact.

Enfin, dans le chapitre 4, nous avons comparé les performances des estimateurs proposés à l'état de l'art au travers de simulations Monte Carlo et de données réelles sur une application STAP. Ces résultats ont illustré l'intérêt de ces travaux puisque les traitements adaptatifs construits autour des estimateurs proposés se sont montrés souvent optimaux pour le contexte considéré. Par ailleurs, nous nous sommes de plus intéressés au problème de robustesse à la contamination des données par l'introduction de cibles. Nous avons pu mettre en avant que certaines méthodes développées proposaient un meilleur rapport performance-robustesse par rapport à l'état de l'art.

Perspectives

Les estimateurs développés au cours de cette thèse peuvent s'utiliser dans de nombreuses applications de traitement signal nécessitant l'estimation de la matrice de covariance. La structure rang faible plus identité est en effet largement répandue car les sources d'intérêt sont souvent contenues dans un sous-espace de dimension réduit. L'exploitation de cette information a priori peut donc conduire à de meilleures performances pour d'autres traitements que la détection (principalement considérée dans ces travaux). C'est pourquoi l'extension des outils d'estimation développés dans cette thèse à d'autres contextes est une première piste que nous suggérons, on peut citer par exemple :

- la détection d'anomalies en imagerie hyperspectrale [39].
- en finance, les estimateurs robustes ont montré leur intérêt pour l'optimisation de portfolios [126] et l'approche rang faible semble être aussi une piste prometteuse.

– tout autre traitement basé sur un estimateur de la matrice de covariance, comme par exemple, la classification par méthode a noyaux...

Néanmoins, cette adaptation à d'autres contextes n'est pas triviale car elle nécessite de développer de nouveaux outils, incorporant l'estimation de nouveaux paramètres (supposés connus dans ces travaux).

Dans un premier temps, il serait intéressant de considérer le problème additionnel de l'estimation de la moyenne pour le modèle considéré. En effet, ce paramètre est parfois non nul et inconnu (notamment pour les applications liées au traitement d'images). Cette problématique peut se résoudre de différentes manières. On pourrait par exemple envisager d'utiliser d'un estimateur de type "plug-in" dans les algorithmes déjà développés. Un autre piste serait l'ajout d'un bloc "moyenne" à optimiser à l'aide des algorithmes MM développés au chapitre 2.

Dans un second temps, il semble nécessaire de s'affranchir de la connaissance a priori du rang afin de couvrir un plus large panel de situations. En effet, dans la plupart des cas, le rang de la matrice de covariance est inconnu et doit donc être estimé au préalable [7, 10, 109, 93, 42]. Il serait donc intéressant de considérer la problématique (complexe) d'estimation du rang pour le contexte que nous avons considéré dans ces travaux.

Une autre piste, permettant de promouvoir les estimateurs de sous-espace développés dans ces travaux, serait le développement de nouveaux traitements "rang faible". Il semblerait notamment utile de développer des détecteurs d'anomalie basés sur des méthodes à sous-espace pour l'application à l'imagerie hyperspectrale. En effet, les données hyperspectrales peuvent se modéliser comme une somme de contributions de peu de sources différentes. Ces données sont alors contenues dans un sous-espace de dimension très réduite comparée à la taille des observations. Afin de tirer profit de cet a priori, il serait intéressant de développer des rapports de vraisemblance permettant de vérifier si la donnée sous test appartient à l'espace engendré par les données environnantes, ou si elle contient une anomalie en dehors de ce sous-espace (en s'inspirant par exemple des travaux [100, 14]). Le développement de ces nouveaux outils ouvrirait de plus la possibilité d'une étude théorique de leur performances.

En termes d'étude théorique de traitements, les performances théoriques des filtres rang faible basés sur les projecteurs issus de la SCM et la NSCM ont été dérivées dans [48, 50], la loi du détecteur rang faible est étudiée dans [8] (non adaptatif) et [49] (adaptatif). De récents travaux ont aussi dérivé les performances théoriques des filtres [24] et détecteurs [23] adaptatifs pour les régimes de grande dimension. Il serait envisageable d'étendre ces résultats aux modèles et estimateurs considérés dans cette thèse afin de prendre en compte l'hétérogénéité du fouillis.

Concernant l'étude de performances, il reste de plus quelques points à éclaircir en termes d'estimation du sous-espace. Le projecteur sur le sous-espace fouillis n'est pas un paramètre apparaissant directement dans la vraisemblance, est-il néanmoins possible de dériver une borne inférieure (de type Cramér-Rao) sur l'erreur de son estimation ? De prime abord, il serait possible de dériver une borne sur la base du sous-espace, bien que techniquement difficile d'inclure les contraintes unitaires sur cette base [52]. Néanmoins, une borne sur l'erreur quadratique moyenne d'un projecteur ou d'une base a-t-elle réellement un sens quand on s'intéresse à l'estimation d'un sous-espace ? Une possible réponse à ces questions peut se trouver dans les travaux [106, 105], dans lesquels sont développées des bornes intrinsèques (indépendantes du système de coordonnées) sur le sous-espace d'intérêt. L'étude de ces bornes

pour le modèle considéré dans cette thèse est une piste de recherche qu'il serait intéressant d'explorer.

Enfin, nous souhaitons dans un futur proche, poursuivre le développement d'estimateurs robustes structurés. Les méthodes d'optimisation MM sont en effet adaptées cette problématique et il reste certaines pistes à explorer. Par exemple, la SVD est une paramétrisation omniprésente en Traitement du Signal. Néanmoins, à notre connaissance, il n'existe pas de méthodes de régularisation du spectre, ni des vecteurs propres, des estimateurs robuste de la matrice de covariance. Développer ces méthodes donnerait un outil permettant de prendre en compte une information a priori sur ces paramètres, applicable dans de de nombreux contextes.

Bibliographie

[1] H. Abeida and J.P. Delmas. Music-like estimation of direction of arrival for noncircular sources. *Signal Processing, IEEE Transactions on*, 54(7) :2678–2690, 2006.

[2] Y. Abramovich and O. Besson. Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach - part 1 : The over-sampled case. *Signal Processing, IEEE Transactions on*, PP(99) :1–1, 2013.

[3] Y.I. Abramovich and N.K. Spencer. Diagonally loaded normalised sample matrix inversion (lnsmi) for outlier-resistant adaptive filtering. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, volume 3, pages III–1105–III–1108, April 2007.

[4] T.E. Abrudan, J. Eriksson, and V. Koivunen. Steepest descent algorithms for optimization under unitary matrix constraint. *Signal Processing, IEEE Transactions on*, 56(3) :1134–1147, March 2008.

[5] T. Adali and S. Haykin. *Adaptive signal processing : next generation solutions*, volume 55. Wiley-IEEE Press, 2010.

[6] T. Adali, P.J. Schreier, and L.L. Scharf. Complex-valued signal processing : The proper way to deal with impropriety. *Signal Processing, IEEE Transactions on*, 59(11) :5101–5125, Nov 2011.

[7] H. Akaike. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6) :716–723, 1974.

[8] L. Anitori, R. Srinivasan, and M. Rangaswamy. Performance of low-rank STAP detectors. In *Radar Conference, 2008. RADAR '08. IEEE*, pages 1–6, 2008.

[9] A. Aubry, Antonio De Maio, L. Pallotta, and Alfonso Farina. Maximum likelihood estimation of a structured covariance matrix with a condition number constraint. *Signal Processing, IEEE Transactions on*, 60(6) :3004–3021, June 2012.

[10] Andrew Barron, Jorma Rissanen, and Bin Yu. The minimum description length principle in coding and modeling. *Information Theory, IEEE Transactions on*, 44(6) :2743–2760, 1998.

[11] S. Bausson, F. Pascal, P. Forster, J.P. Ovarlez, and P. Larzabal. First- and second-order moments of the normalized sample covariance matrix of spherically invariant random vectors. *IEEE Signal Processing Letters*, 14(6) :425 – 428, June 2007.

[12] H. Belkacemi. *Traitement adaptatif spatio-temporel pour les radars aeroportes monostatique/ bistatique*. PhD thesis, Universite Paris-sud, 2006.

[13] O. Besson and Y. Abramovich. Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach - part 2 : The under-sampled case. *Signal Processing, IEEE Transactions on*, PP(99) :1–1, 2013.

[14] O. Besson and Y. Abramovich. Adaptive detection in elliptically distributed noise and under-sampled scenario. *Signal Processing Letters, IEEE*, 21(12) :1531–1535, Dec 2014.

[15] O. Besson and S. Bidon. Adaptive processing with signal contaminated training samples. *Signal Processing, IEEE Transactions on*, 61(17) :4318–4329, Sept 2013.

[16] S. Bidon, M. Montecot, and L. Savy. Introduction au satp. partie iii : Les donnees du club stap. *Traitement du Signal*, 2011.

[17] L. E. Brennan and L. S. Reed. Theory of adaptive radar. *IEEE Trans. on Aero. and Elec. Syst.*, 9(2) :237 – 252, 1973.

[18] L. E. Brennan and F.M. Staudaher. Subclutter visibility demonstration. Technical report, RL-TR-92-21, Adaptive Sensors Incorporated, March 1992.

[19] J. P. Burg, D. G. Luenberger, and D. L. Wenger. Estimation of structured covariance matrices. *Proc. IEEE*, 70(9) :963–974, September 1982.

[20] S. Cambanis, S. Huang, and G. Simons. On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, 11(3) :368–385, 1981.

[21] P. Chargé, Y. Wang, and J. Saillard. A non-circular sources direction finding method using polynomial rooting. *Signal Processing*, 81(8) :1765–1770, 2001.

[22] Y. Chen, A. Wiesel, and A. O. Hero. Robust shrinkage estimation of high-dimensional covariance matrices. *Signal Processing, IEEE Transactions on*, 59(9) :4097–4107, 2011.

[23] A. Combernoux, F. Pascal, G. Ginolhac, and M. Lesturgie. Asymptotic performance of the low rank adaptive normalized matched filter in a large dimensional regime. *ICASSP*, Apr. 2015.

[24] A. Combernoux, F. Pascal, G. Ginolhac, and M. Lesturgie. Theoretical performance of low rank adaptive filters in gaussian context in the large dimensional regime. *IEEE Trans. on Signal Process.*, July 2015. submitted.

[25] E. Conte and M. Longo. Characterization of radar clutter as a spherically invariant random process. *IEE Proceeding, Part. F*, 134(2) :191–197, April 1987.

[26] E. Conte, M. Longo, and M. Lops. Modelling and simulation of non-rayleigh radar clutter. *IEE Proceeding, Part. F*, 138(2) :121–138, April 1991.

[27] E. Conte, M. Longo, M. Lops, and S.L. Ullo. Radar detection of signals with unknown parameters in K-distributed clutter. *Radar and Signal Processing, IEE Proceedings F*, 138(2) :131 –138, April 1991.

[28] E. Conte, M. Lops, and G. Ricci. Asymptotically optimum radar detection in compound-gaussian clutter. *IEEE Trans. on Aero. and Elec. Syst.*, 31(2) :617 – 625, April 1995.

[29] E. Conte, M. Lops, and G. Ricci. Adaptive matched filter detection in spherically invariant noise. *IEEE Sig. Proc. Letters*, (8), August 1996.

[30] E. Conte, M. Lops, and G. Ricci. Adaptive detection schemes in compound-gaussian clutter. *IEEE Trans. on Aero. and Elec. Syst.*, 34(4) :1058 – 1069, July 1998.

[31] E. Conte and A. De Maio. Exploiting persymmetry for cfar detection in compound-gaussian clutter. *IEEE Trans. on Aero. and Elec. Syst.*, 39(2) :719 – 724, April 2003.

[32] E. Conte, A. De Maio, and G. Ricci. Recursive estimation of the covariance matrix of a compound-Gaussian process and its application to adaptive CFAR detection. *IEEE Trans. on Sig. Proc.*, 50(8) :1908 – 1915, August 2002.

[33] Romain Couillet and Matthew McKay. Large dimensional analysis and optimization of robust shrinkage covariance matrix estimators. *Journal of Multivariate Analysis*, 131 :99–120, 2014.

[34] P. Dreuillet and H. Cantalloube and al. The onera ramses sar : latest significant results and future developments. In *Proc. of the IEEE Int. Radar Conf.*, 2006.

[35] A. Edelman, T. Arias, and S. Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications*, 20(2) :303–353, 1998.

[36] H.Y. Fang. Symmetric multivariate and related distributions (monographs on statistics and applied probability/36). *Recherche*, 67 :02, 1989.

[37] P. Forster. Generalized cross spectral matrices for array of arbitrary geometry. *IEEE Trans. on Sig. Proc.*, 49(5) :972–978, may 2001.

[38] G. Frahm. *Generalized elliptical distributions : theory and applications*. PhD thesis, Universität zu Köln, Germany, 2004.

[39] J. Frontera-Pons. *Robust target detection for Hyperspectral Imaging*. PhD thesis, STITS, Supelec, 2014.

[40] J. Frontera-Pons, M. Mahot, J.P. Ovarlez, F. Pascal, and J. Chanussot. Robust detection using M-estimators for hyperspectral imaging. In *Workshop on Hyperspectral Image and Signal Processing : Evolution in Remote Sensing*, 2012.

[41] J. Frontera-Pons, M. Mahot, J.P. Ovarlez, F. Pascal, S.K. Pang, and J. Chanussot. A class of robust estimates for detection in hyperspectral images using elliptical distribution background. In *IGARSS. Conference*. IEEE, 2012.

[42] M. Gavish and D.L. Donoho. The optimal hard threshold for singular values is. *Information Theory, IEEE Transactions on*, 60(8) :5040–5053, 2014.

[43] K. Gerlach. Outlier resistant adaptive matched filtering. *Aerospace and Electronic Systems, IEEE Transactions on*, 38(3) :885–901, Jul 2002.

[44] F. Gini, M. V. Greco, M. Diani, and L. Verrazzani. Performance analysis of two adaptive radar detectors against non-gaussian real sea clutter data. *IEEE Trans.-AES*, 36(4) :1429–1439, October 2000.

[45] F. Gini and M.V. Greco. Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter. *Signal Processing, special section on SP with Heavy Tailed Distributions*, 82(12) :1847–1859, December 2002.

[46] F. Gini and J.H. Michels. Performance analysis of two covariance matrix estimators in compound-gaussian clutter. *Radar, Sonar and Navigation, IEE Proceedings* -, 146(3) :133–140, Jun 1999.

[47] G. Ginolhac. *Detection / Estimation a l'aide de methodes algebriques - Application au domaine du RADAR*. Rapport HdR. Cachan, France, 2011.

[48] G. Ginolhac and P. Forster. Performance analysis of a robust low-rank stap filter in low-rank gaussian clutter. In *Proceedings of ICASSP*, Dallas, TX, USA, april 2010.

[49] G. Ginolhac and P. Forster. Approximate distribution of the low-rank adaptive normalized matched filter test statistic under the null hypothesis. *Submitted TO IEEE TRANS. ON AERO. AND ELEC. SYST., 2015.*, 2015.

[50] G. Ginolhac, P. Forster, F. Pascal, and J.-P. Ovarlez. Performance of two low-rank STAP filters in a heterogeneous noise. *Signal Processing, IEEE Transactions on*, 61(1) :57–61, 2013.

[51] J. Goldman. Detection in the presence of spherically symmetric random vectors. *IEEE Trans.-IT*, 22(1) :52–59, January 1976.

[52] J.D. Gorman and A.O. Hero. Lower bounds for parametric estimation with constraints. *Information Theory, IEEE Transactions on*, 36(6) :1285–1301, Nov 1990.

[53] M. Greco, S. Fortunati, and F. Gini. Maximum likelihood covariance matrix estimation for complex elliptically symmetric distributions under mismatched condition. *Signal Processing Elsevier*, 104 :381–386, Nov 2014.

[54] A. Haimovich. Asymptotic distribution of the conditional signal-to-noise ratio in an eigenanalysis-based adaptive array. *IEEE Trans. on Aero. and Elec. Syst.*, 33 :988 – 997, 1997.

[55] P. J. Huber. *Robust Statistics.* Wiley Series in Probability and Statistics. John Wiley & Sons, 1981.

[56] P. J. Huber and E. M. Ronchetti. *Robust statistics*. John Wiley & Sons Inc, 2009.

[57] M.K. Ibrahim and A.P. D'amico. On the generalised burg technique. *Communications, Radar and Signal Processing, IEE Proceedings F*, 134(2) :135–140, April 1987.

[58] E. Jakeman and P. N. Pusey. A model for non-rayleigh sea echo. *IEEE Trans.-AP*, 24(6) :806–814, November 1976.

[59] I.M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Annals of statistics*, pages 295–327, 2001.

[60] A. Kammoun, R. Couillet, F. Pascal, and Alouini M.-S. Convergence and fluctuations of regularized tyler estimators. *(submitted to) Signal Processing, IEEE Transactions on*, 2015.

[61] A. Kammoun, R. Couillet, F. Pascal, and Alouini M.-S. Optimal design of the adaptive normalized matched filter detector. *(submitted to) Signal Processing, IEEE Transactions on*, 2015.

[62] B. Kang, V. Monga, and M. Rangaswamy. Rank-constrained maximum likelihood estimation of structured covariance matrices. *Aerospace and Electronic Systems, IEEE Transactions on*, 50(1) :501–515, January 2014.

[63] S. M. Kay. *Fundamentals of Statistical Signal Processing - Estimation Theory*, volume 1. Prentice-Hall PTR, Englewood CliJs, NJ, 1993.

[64] D. Kelker. Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhya : The Indian Journal of Statistics, Series A*, 32(4) :419–430, 1970.

[65] J. T. Kent and D. E. Tyler. Redescending M-estimates of multivariate location and scatter. *Annals of Statistics*, 19(4) :2102–2119, December 1991.

[66] I. Kirsteins and D. Tufts. Adaptive detection using a low rank approximation to a data matrix. *IEEE Trans. on Aero. and Elec. Syst.*, 30 :55 – 67, 1994.

[67] S. Kraut and L.L. Scharf. The cfar adaptive subspace detector is a scale-invariant glrt. *Signal Processing, IEEE Transactions on*, 47(9) :2538–2541, 1999.

[68] S. Kraut, L.L. Scharf, and L.T. McWhorter. Adaptive subspace detectors. *IEEE Trans. on Sig. Proc.*, 49(1) :1–16, january 2001.

[69] K. Lange and H. Zhou. MM algorithms for geometric and signomial programming. *Mathematical programming*, 143(1-2) :339–356, 2014.

[70] Olivier Ledoit and Michael Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2) :365 – 411, 2004.

[71] M. Mahot. *Estimation robuste de la matrice de covariance en traitement du signal*. PhD thesis, ENS Cachan.

[72] M. Mahot, F. Pascal, P. Forster, and J. Ovarlez. Asymptotic properties of robust complex covariance matrix estimates. *Signal Processing, IEEE Transactions on*, 61(13) :3348–3356, July 2013.

[73] J.H. Manton. Optimization algorithms exploiting unitary constraints. *Signal Processing, IEEE Transactions on*, 50(3) :635–650, Mar 2002.

[74] R. A. Maronna. Robust M-estimators of multivariate location and scatter. *Annals of Statistics*, 4(1) :51–67, January 1976.

[75] R. A. Maronna, D. R. Martin, and J. V. Yohai. *Robust Statistics : Theory and Methods*. Wiley Series in Probability and Statistics. John Wiley & Sons, 2006.

[76] J. H. Michels, M. Rangaswamy, and B. Himed. Performance of parametric and covariance based STAP tests in compound-Gaussian clutter. *Digital Signal Processing*, 12 :307–328, April-July 2002.

[77] O. Besson, S. Bidon. Robust adaptive beamforming using a bayesian steering vector error model. *Signal Processing*, 93(12) :3290 – 3299, 2013. Special Issue on Advances in Sensor Array Processing in Memory of Alex B. Gershman.

[78] E. Ollila. On the circularity of a complex random variable. *Signal Processing Letters, IEEE*, 15 :841–844, 2008.

[79] E. Ollila, J. Eriksson, and V. Koivunen. Complex elliptically symmetric random variables-generation, characterization, and circularity tests. *Signal Processing, IEEE Transactions on*, 59(1) :58–69, 2011.

[80] E. Ollila and V. Koivunen. Robust antenna array processing using M-estimators of pseudo-covariance. In *Proc. 14th IEEE Int. Symp. Personal, Indoor, Mobile Radio Commun.(PIMRC)*, pages 7–10, 2003.

[81] E. Ollila and V. Koivunen. Adjusting the generalized likelihood ratio test of circularity robust to non-normality. In *Signal Processing Advances in Wireless Communications, 2009. SPAWC'09. IEEE 10th Workshop on*, pages 558–562. IEEE, 2009.

[82] E. Ollila and D.E. Tyler. Distribution-free detection under complex elliptically symmetric clutter distribution. In *Sensor Array and Multichannel Signal Processing Workshop (SAM), 2012 IEEE 7th*, pages 413–416, June 2012.

[83] E. Ollila and D.E. Tyler. Regularized M -estimators of scatter matrix. *Signal Processing, IEEE Transactions on*, 62(22) :6059–6070, Nov 2014.

[84] E. Ollila, D.E. Tyler, V. Koivunen, and H.V. Poor. Complex elliptically symmetric distributions : Survey, new results and applications. *Signal Processing, IEEE Transactions on*, 60(11) :5597–5625, 2012.

[85] E. Ollila, D.E. Tyler, V. Koivunen, and H.V. Poor. Compound-Gaussian clutter modeling with an inverse gaussian texture distribution. *Signal Processing Letters, IEEE*, 19(12) :876–879, Dec 2012.

[86] B Ottersten, P Stoica, and R Roy. Covariance matching estimation techniques for array signal processing applications. *Digital Signal Processing*, 8(3) :185 – 210, 1998.

[87] G. Pailloux, P. Forster, J.P. Ovarlez, and F. Pascal. Persymmetric adaptive radar detectors. *IEEE Trans. on Aero. and Elec. Syst.*, 47(4) :2376–2390, October 2011.

[88] F. Pascal, L. Bombrun, J.-Y. Tourneret, and Y. Berthoumieu. Parameter estimation for multivariate generalized gaussian distributions. *Signal Processing, IEEE Transactions on*, 61(23) :5960–5971, Dec 2013.

[89] F. Pascal, Y. Chitour, J.P. Ovarlez, P. Forster, and P. Larzabal. Existence and characterization of the covariance matrix maximum likelihood estimate in Spherically Invariant Random Processes. *IEEE Trans on Sig. Proc.*, 56(1) :34 – 48, January 2008.

[90] F. Pascal, Y. Chitour, and Y. Quek. Generalized robust shrinkage estimator and its application to stap detection problem. *Signal Processing, IEEE Transactions on*, 62(21) :5640–5651, Nov 2014.

[91] F. Pascal, P. Forster, J.P. Ovarlez, and P. Larzabal. Performance analysis of covariance matrix estimates in impulsive noise. *IEEE Trans on Sig. Proc.*, 56(6) :2206 – 2217, June 2008.

[92] F. Pascal, H. Harari-Kermadec, and P. Larzabal. The empirical likelihood method applied to covariance matrix estimation. *Signal Processing*, 90(2) :566–578, 2010.

[93] P.O. Perry and P.J. Wolfe. Minimax rank estimation for subspace tracking. *Selected Topics in Signal Processing, IEEE Journal of*, 4(3) :504–513, June 2010.

[94] R.S. Raghavan. Statistical interpretation of a data adaptive clutter subspace estimation algorithm. *IEEE Trans. on Aero. and Elec. Syst.*, 48(2) :1370 – 1384, 2012.

[95] M. Rangaswamy, I.P. Kirsteins, B.E. Freburger, and D.W. Tufts. Signal detection in strong low rank Compound-Gaussian interference. In *Sensor Array and Multichannel Signal Processing Workshop. 2000. Proceedings of the 2000 IEEE*, pages 144–148, 2000.

[96] M. Rangaswamy, F.C. Lin, and K.R. Gerlach. Robust adaptive signal processing methods for heterogeneous radar clutter scenarios. *Signal Processing*, 84 :1653 – 1665, 2004.

[97] M. Rangaswamy, D. D. Weiner, and A. Ozturk. Non-Gaussian vector identification using spherically invariant random processes. *IEEE Trans.-AES*, 29(1) :111–124, January 1993.

[98] M. Razaviyayn, M. Hong, and Z. Luo. A unified convergence analysis of block successive minimization methods for nonsmooth optimization. *SIAM Journal on Optimization*, 23(2) :1126–1153, 2013.

[99] I.S. Reed, J.D. Mallett, and L.E. Brennan. Rapid convergence rate in adaptive arrays. *IEEE Trans. on Aero. and Elec. Syst.*, AES-10(6) :853 – 863, November 1974.

[100] L.L. Scharf and B. Friedlander. Matched subspace detectors. *IEEE Trans. on Sig. Proc.*, 42(8) :2146–2157, august 1994.

[101] R. O. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Trans.-ASSP*, 34(3) :276–280, March 1986.

[102] P.H. Schönemann. A generalized solution of the orthogonal procrustes problem. *Psychometrika*, 31(1) :1–10, 1966.

[103] P.J. Schreier and L.L. Scharf. *Statistical signal processing of complex-valued data : the theory of improper and noncircular signals*. Cambridge University Press, 2010.

[104] P. Shah and V. Chandrasekaran. Group symmetry and covariance regularization. *Electron. J. Statist.*, 6 :1600–1640, 2012.

[105] S.T. Smith. Intrinsic cramer-rao bounds and subspace estimation accuracy. In *Sensor Array and Multichannel Signal Processing Workshop. 2000. Proceedings of the 2000 IEEE*, pages 489–493, 2000.

[106] S.T. Smith. Covariance, subspace, and intrinsic crame acute ;r-rao bounds. *Signal Processing, IEEE Transactions on*, 53(5) :1610–1630, May 2005.

[107] I. Soloveychik and A. Wiesel. Group symmetry and non-gaussian covariance estimation. In *Global Conference on Signal and Information Processing (GlobalSIP), 2013 IEEE*, pages 1105–1108, Dec 2013.

[108] I. Soloveychik and A. Wiesel. Tyler's covariance matrix estimator in elliptical models with convex structure. *Signal Processing, IEEE Transactions on*, 62(20) :5251–5259, Oct 2014.

[109] Petre Stoica and Y. Selen. Model-order selection : a review of information criterion rules. *Signal Processing Magazine, IEEE*, 21(4) :36–47, July 2004.

[110] Y. Sun, P. Babu, and D.P. Palomar. Robust estimation of structured covariance matrix for heavy-tailed elliptical distributions. *Signal Processing, IEEE Transactions on, submited*.

[111] Y. Sun, P. Babu, and D.P. Palomar. Regularized tyler's scatter estimator : Existence, uniqueness, and algorithms. *Signal Processing, IEEE Transactions on*, 62(19) :5143–5156, Oct 2014.

[112] J.K. Thomas, L.L. Scharf, and D.W. Tufts. The probability of a subspace swap in the svd. *Signal Processing, IEEE Transactions on*, 43(3) :730–736, Mar 1995.

[113] G. V. Trunk and S. F. George. Detection of targets in non-gaussian sea clutter. *IEEE Trans.-AES*, 6(8) :620–628, September 1970.

[114] D. Tyler. A distribution-free M-estimator of multivariate scatter. *The Annals of Statistics*, 15(1) :234–251, 1987.

[115] D.E. Tyler. Radial estimates and the test for sphericity. *Biometrika*, 69(2) :429, 1982.

[116] D.E. Tyler. Robustness and efficiency properties of scatter matrices. *Biometrika*, 70(2) :411, 1983.

[117] D.E. Tyler. A distribution-free M-estimator of multivariate scatter. *The Annals of Statistics*, 15(1) :234–251, 1987.

[118] 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 :354, 1988.

[119] M.L. Vis and L.L. Scharf. A note on recursive maximum likelihood for autoregressive modeling. *Signal Processing, IEEE Transactions on*, 42(10) :2881–2883, Oct 1994.

[120] J. Ward. Space-time adaptive processing for airborne radar. Technical report, Lincoln Lab., MIT, Lexington, Mass., USA, December 1994.

[121] K. D. Ward. Compound representation of high resolution sea clutter. *Electronics Letters*, 17(16) :561–563, August 1981.

[122] K. D Ward, S. Watts, and R. Tough. *Sea clutter : scattering, the K distribution and radar performance*, volume 20. IET, 2006.

[123] S. Watts. Radar detection prediction in sea clutter using the compound k-distribution model. *IEE Proceeding, Part. F*, 132(7) :613–620, December 1985.

[124] A. Wiesel. Geodesic convexity and covariance estimation. *Signal Processing, IEEE Transactions on*, 60(12) :6182–6189, Dec 2012.

[125] A. Wiesel. Unified framework to regularized covariance estimation in scaled Gaussian models. *Signal Processing, IEEE Transactions on*, 60(1) :29–38, 2012.

[126] L. Yang, R. Couillet, and M. McKay. A robust statistics approach to minimum variance portfolio optimization. *Transactions on Signal Processing, IEEE (submitted)*, 21(12) :1531–1535, 2014.

[127] K. Yao. A representation theorem and its applications to spherically invariant random processes. *IEE Trans. on Inf. Th.*, 19(5) :600 – 608, September 1973.

[128] T. Zhang, A. Wiesel, and M.S. Greco. Multivariate generalized gaussian distribution : Convexity and graphical models. *Signal Processing, IEEE Transactions on*, 61(16) :4141–4148, Aug 2013.

Publications et Communications

Articles de Revues Internationales

[J1] A. Breloy, G. Ginolhac, F. Pascal, P. Forster, "Clutter Subspace Estimation in Low Rank Heterogeneous Noise Context", Signal Processing, IEEE Transactions on , vol.63, no.9, pp.2173-2182, May1, 2015.

[J2] A. Breloy, G. Ginolhac, F. Pascal, P. Forster, "Robust Covariance Matrix estimation in Low-Rank Heterogeneous Context", Signal Processing, IEEE Transactions on, soumis juin 2015.

[J3] Y. Sun, A. Breloy, P. Babu, D. Palomar, F. Pascal, G. Ginolhac, "Low-Complexity Algorithms for Low Rank Clutter Parameters Estimation in Radar Systems", Signal Processing, IEEE Transactions on, soumis août 2015.

Articles de Conférences Internationales

[C1] A. Breloy, L. Le Magoarou, G. Ginolhac, F. Pascal, P. Forster, "Maximum likelihood estimation of clutter subspace in non homogeneous noise context", proceedings of EUSIPCO 2013.

[C2] A. Breloy, G. Ginolhac, F. Pascal, P. Forster, "Robust estimation of the clutter subspace for a low rank heterogeneous noise under high clutter to noise ratio assumption", proceedings of ICASSP 2014.

[C3] A. Breloy, G. Ginolhac, F. Pascal, P. Forster, "CFAR property and robustness of the low rank adaptive normalized matched fiters detectors in low rank compound Gaussian context", proceedings of SAM 2014.

[C4] A. Breloy, L. Le Magoarou, G. Ginolhac, F. Pascal, P. Forster, "Numerical performances of low rank STAP based on different heterogeneous clutter subspace estimators", proceedings of International RADAR Conference 2014.

Articles de Conférences Nationales

[FC1] A. Breloy, L. Le Magoarou, G. Ginolhac, F. Pascal, P. Forster, "Estimation par maximum de vraisemblance du sous espace clutter dans un bruit hétérogène rang faible avec application au STAP", GRETSI 2013.

[FC2] A. Breloy, G. Ginolhac, F. Pascal, P. Forster, "Estimation Robuste de la Matrice de Covariance en contexte Hétérogène Rang Faible", GRETSI 2015.

Séminaires / Workshops

[SW1] A. Breloy, L. Le Magoarou, G. Ginolhac, F. Pascal, P. Forster, "Maximum likelihood estimation of clutter subspace in Heterogeneous noise context", SONDRA Workshop 2013.