Statistical signal processing exploiting low-rank priors with applications to detection in Heterogeneous Environment

Rayen Ben Abdallah

Université Paris Nanterre

200 av. de la République
92001 Nanterre Cedex
www.parisnanterre.fr

Ecole doctorale « Connaissance, Langage, Modélisation»
Laboratoire Energétique Mécanique Electromagnétisme
50 rue de sèvres, 92410, Ville d'Avray

Rayen BEN ABDALLAH

Statistical signal processing exploiting low-rank priors with applications to detection in Heterogeneous Environment

Traitements statistiques exploitant des a priori rang faible pour la détection en contexte hétérogène

Thèse présentée et soutenue publiquement le 04 novembre 2019
en vue de l'obtention du doctorat en traitement de signal de l'Université Paris Nanterre
sous la direction de David Lautru (Université Paris Nanterre)
et de Mohammed Nabil El Korso et Arnaud Breloy (co-encadrants, Université Paris Nanterre)

Jury * **:**

Rapporteur :	M. Karim Abed-Meraim	Professeur, Polytechnique Orléans
Rapporteur :	M. Olivier Besson	Professeur, Institut supérieur de l'aéronautique et l'espace (ISAE-SUPAERO) à Toulouse
Membre du jury :	Mme Audrey Giremus	Maître de conférences, Université de Bordeaux
Membre du jury :	M. Jean -Philippe Ovarlez	Directeur de recherche, ONERA & Université Paris Saclay
Membre du jury :	M. Thomas Rodet	Professeur, ENS Paris Saclay

To the memory of my grandmother Jamila
To my aunt Rafiaa
To my dear parents, Mohamed and Farouza
To my husband Yassine
To my sister and my two brothers
To my little daughter Ella

« The important thing is not to stop questioning. Curiosity has its own reason for existence. One cannot help but be in awe when he contemplates the mysteries of eternity, of life, of the marvelous structure of reality. It is enough if one tries merely to comprehend a little of this mystery each day. »

Albert Einstein

Acknowledgment

I would like to offer my special thanks to the members of my dissertation committee not only for their time and extreme patience, but for their intellectual contributions to my development as a scientist.

I am deeply grateful to all members of the jury for agreeing to read the manuscript and to participate in the defense of this thesis. I am grateful to Olivier Besson, professor at ISAE-SUPAERO, and Karim Abed Meraim, professor at Polytechnique Orléans, for the careful reading of the text. Their corrections and suggestions greatly improved the readability of the thesis. I would also like to warmly thank Audrey Giremus, associate professor at university Bordeaux, Jean philippe ovarlez, director of research at ONERA and university Paris Saclay, and Thomas Rodet, professor at ENS Paris-Saclay, for their participation to the jury and their constructive comments.

I would like to express my sincere gratitude to my advisors Arnaud Breloy, associate teacher at university Nanterre, and Mohamed Nabil El Korso, associate teacher at university Nanterre, for the continuous support of my Ph.D study and related research, for their patience, motivation, and immense knowledge. Their guidance helped me in all the time of research and writing of this thesis. Besides, I would like to thank my supervisor David Lautru, professor at university Nanterre, for his insightful comments and encouragement.

My sincere thanks also goes to Najet Neji, associate teacher at university Paris-Saclay, and Hichem Aroui, director of research at university Paris-Saclay, who provided me an opportunity to join their team as temporary assistant teacher at university Evry. I would like to express my deepest appreciation to Ammar Mian, doctor at Aalto university, and Abigael Taylor, a research engineer at ONERA, for their helpful contribution. An appreciation goes to all my colleagues and friends from university Nanterre and univesrity Evry: Majda, Rayen, Fatma, Sarra and Manel.

I am grateful to my parents Mohamed and Farouza, who have provided me through moral and emotional support in my life. I would like to offer my special thanks to my aunt Rafiaa for being a second mom, and for always being there for me. I am also grateful to my husband Yassine who has supported me along the way. A very special gratitude goes out to my sister Fida and my brothers Houssem and Ghassen.

Abstract

In this thesis, we consider first the problem of low dimensional signal subspace estimation in a Bayesian context. We focus on compound Gaussian signals embedded in white Gaussian noise, which is a realistic modeling for various array processing applications. Following the Bayesian framework, we derive algorithms to compute both the maximum a posteriori and the so-called minimum mean square distance estimator, which minimizes the average natural distance between the true range space of interest and its estimate. Such approaches have shown their interests for signal subspace estimation in the small sample support and/or low signal to noise ratio contexts. As a byproduct, we also introduce a generalized version of the complex Bingham Langevin distribution in order to model the prior on the subspace orthonormal basis. Numerical simulations illustrate the performance of the proposed algorithms. Then, a practical example of Bayesian prior design is presented for the purpose of radar detection.

Second, we aim to test common properties between low rank structured covariance matrices. Indeed, this hypothesis testing has been shown to be a relevant approach for change and/or anomaly detection in synthetic aperture radar images. While the term "similarity" usually refers to equality or proportionality, we explore the testing of shared properties in the structure of low rank plus identity covariance matrices, which are appropriate for radar processing. Specifically, we derive generalized likelihood ratio tests to infer *i*) on the equality/proportionality of the low rank signal component of covariance matrices, and *ii*) on the equality of the signal subspace component of covariance matrices. The formulation of the second test involves non-trivial optimization problems for which we tailor efficient Majorization-Minimization algorithms. Eventually, the proposed detection methods enjoy interesting properties, that are illustrated on simulations and on an application to real data for change detection.

Résumé

Dans un premier lieu, nous considérons le problème de l'estimation de sous-espace d'un signal d'intérêt à partir d'un jeu de données bruité. Pour ce faire, nous adoptons une approche Bayésienne afin d'obtenir un estimateur minimisant la distance moyenne entre la vraie matrice de projection $\mathbf{U}\mathbf{U}^H$ et son estimée $\widehat{\mathbf{U}}\widehat{\mathbf{U}}^H$. Plus particuliérement, nous étendons les estimateurs au contexte Gaussien composé pour les sources où l'a priori sur la base $\mathbf{U}$ sera une loi *complex generalized Bingham Langevin*. Enfin, nous étudions numériquement les performances de l'estimateur proposé sur une application de type *space time adaptive processing* pour un radar aéroporté au travers de données réelles.

Dans un second lieu, nous nous intéressons au test de propriété communes entre les matrices de covariance. Nous proposons de nouveaux tests statistiques dans le contexte de matrices de covariance structurées. Plus précisément, nous considérons un signal de rang faible corrompu par un bruit blanc Gaussien additif. Notre objectif est de tester la similarité des composantes principales à rang faible communes à un ensemble de matrices de covariance. Dans un premier temps, une statistique de décision est dérivée en utilisant le rapport de vraisemblance généralisée. Le maximum de vraisemblance n'ayant pas d'expression analytique dans ce cas, nous proposons un algorithme d'estimation itératif de type majoration-minimisation pour pouvoir évaluer les tests proposés. Enfin, nous étudions les propriétés des détecteurs proposés à l'aide de simulations numériques.

Contents

List of Figures

List of Tables

Acronyms

ACG	Angular Central Gaussian
AFE	Average Fraction of Energy
CES	Complex Elliptically Symmetric
CG	Compound Gaussian
CGB	Complex Generalized Bingham
CGBL	Complex Generalized Bingham Langevin
CIB	Complex Invariant Bingham
CL	Complex Langevin
CSP	Clutter Subspace Projector
DOA	Direction Of Arrival
EBMM	Eigenspace Block Maximization Minimization
EVD	Eigenvalue Decomposition
FPE	Fixed Point Estimator
GFPE	Generalized Fixed Point Estimator
GLRT	Generalized Likelihood Ratio Test
LR-ANMF	Low Rank Adaptive Normalized Matched Filter
MAP	Maximum A Posterior
MCMC	Monte Carlo Markov Chain
MLE	Maximum Likelihood Estimator
MM	Maximization Minimization
MMSD	Minimum Mean Square Distance
MUSIC	MUltiple SIgnal Classification
ONR	Outlier to Noise Ratio
PCA	Principal Component Analysis
PD	Probability of Detection
p.d.f.	probability density function
PFA	Probability of False Alarm
RBL	Real Bingham Langevin
ROC	Receiver Operating characteristic Curve
SAR	Synthetic Aperture Radar
SCM	Sample Covariance Matrix
SFPE	Shrinkage Fixed Point Estimator
SNR	Signal to Noise Ratio
STAP	Space Time Adaptive Processing
SVD	Singular Value Decomposition
vCBL	vector Complex Bingham Langevin
w.r.t.	with respect to

Mathematical Symbols

a	Scalar quantity
$\mathbf{a}$	Complex vector quantity
$\tilde{\mathbf{a}}$	Real vector quantity
$\mathbf{A}$	Complex matrix
$\tilde{\mathbf{A}}$	Real matrix
$\mathbb{C}$	The set of complex variables
$\mathbb{R}$	The set of real variables
$\mathcal{H}_N^+$	The set of $N \times N$ positive semi-definite Hermitian matrices
$\mathcal{H}_N^{++}$	The set of $N \times N$ positive definite Hermitian matrices
$\mathcal{S}_N^+$	The set of $N \times N$ positive semi-definite symmetric matrices.
$\mathcal{U}_N^R$	The set of $N \times R$ semi-unitary matrices
$\mathcal{U}_N^R(\mathbb{R})$	The set of $N \times R$ unitary matrices
$\mathcal{H}_R^{\mathrm{LR}}$	The set of low rank plus identity matrices
$\mathcal{H}_N^R$	The set of square matrices of size N and rank R
$\{w_n\}$	The set of elements w_n, with $n \in [\![1, N]\!]$
$\mathcal{G}_N^R$	The set of orthogonal projection matrix
$\overset{\mathrm{EVD}}{=}$	The eigenvalue decomposition of a given matrix
$\overset{\mathrm{SVD}}{=}$	The singular value decomposition of a given matrix
diag(.)	Diagonal matrix built from a set of elements (or a vector)
^{H}	The transpose conjugate operator
^{T}	The transpose operator
$\mathbb{E}\{.\}$	The expectation operator
$\|.\|$	The determinant operator
Tr{.}	The trace operator
Re{.}	The real part of a given complex variable
etr{.}	The exponential trace of a given matrix
$\overset{d}{=}$	Has the same distribution as
$\propto$	Proportional to
$\sim$	Distributed as
$\mathbf{I}_N$	The identity matrix of dimension $N \times N$
i	Complex number whose square equals to -1
$\delta_{i,j}$	The Kronecker symbol $\delta_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$
$\mathcal{CN}$	Complex Gaussian distribution
$\mathcal{N}$	The real multivariate normal distribution
$\mathcal{CG}$	Complex compound Gaussian distribution.
Gamma	Gamma distribution
$\mathbb{E}_{\mathbf{U},\mathbf{Z}}\{.\}$	The expectation operator applied on both $\mathbf{U}$ and $\mathbf{Z}$
Unif	The continuous uniform distribution

Introduction

0.1 Motivations

Signal estimation/detection systems must frequently operate in environments that are difficult to characterize with precise statistical models. Thus, it is of interest to develop signal detection procedures that are insensitive to moderate deviations from the assumed signals and noise. During the past two decades, the estimation of the parameters of interest from a high dimensional data set in the presence of noise, clutter or interference is an active research topic that raises many difficulties:

• First, one of the most challenging issues is estimating the parameters of interest in high dimensional settings (i.e., in the case where the data dimension is close to or larger than the sample size). As a remedy, the low-rank approach can robustly and efficiently handle high-dimensional data. Indeed, in many large data sets, relevant information often lies in a subspace of much lower dimension than the ambient space. Thus, the aim of many algorithms can be broadly interpreted as trying to find or exploit this underlying "structure" that is present in the data. As an example, a structure that is particularly useful both due to its wide-ranging applicability and efficient computation is the low-rank structured covariance matrix where the signal of interest lies in low-rank subspace.

• Second, a selection of an appropriate data model is favorable for the robustness of signal detection procedure. One of the simplest and probably most extensively used model is the Gaussian one which is not always realistic and can lead to poor estimation performance due to the presence of possible disruptive outliers or non-Gaussian observations. Therefore, the widely used hypothesis of a Gaussian model may not be valid, in some cases, e.g. as in most high resolution array processing applications. For that purpose, non-Gaussian models for the data have to be considered. To this end, we propose the use of the so-called compound Gaussian that is generally chosen for its statistical properties and for its good fit to empirical distributions. This modeling encompasses many classical distributions as for example the Gaussian, the Student-t or the K-distributions.

0.2 Thesis purpose and manuscript organization

0.2.1 Thesis purpose

The contributions of this thesis are centered around two axes:

• First, we consider the problem of low dimensional signal subspace estimation in a Bayesian context. Indeed, adding a prior information to the signal subspace can improve the estimation process in some critical cases, e.g., low signal to noise ratio or/and small sample support. We

focus on compound Gaussian signals embedded in white Gaussian noise. This modeling has been widely used in several array processing applications for its robustnesses to the presence of outliers, possibly caused by heavy tailed impulsive noise. Following the Bayesian framework, we derive two algorithms to compute the maximum a posteriori estimator and the so-called minimum mean square distance estimator. As a byproduct, we also introduce a generalized version of the complex Bingham Langevin distribution in order to model the prior on the subspace orthonormal basis.

• Second, we study the problem of testing whether the set under test shares common properties with the secondary sets. To this aim, we propose novel detectors based on testing the similarity between the low-rank principal components for the following reasons: *i*) Incorporating the low-rank structure in covariance matrices equality testing offers an improvement of the detection performance (especially for small local windows) such as ensuring a low probability false alarm in local anomaly detection in the context of change detection. *ii*) The low-rank based approaches allow to increase the spatial resolution of the detection process, as they can operate in lower sample support. In this context, we make use of the generalized likelihood ratio test based on the binary hypothesis test in the context of low-rank structured covariance matrix.

0.2.2 Manuscript organization

The presented document is divided in three parts:

- Part I gives an overview of some commonly used distribution and presents a particular structure of the covariance matrix. Specifically, in Chapter 1, we present briefly some commonly used statistical models such as Gaussian, compound Gaussian and mixture model (compound Gaussian plus white Gaussian noise) distributions. In order to incorporate a prior knowledge into the estimation of the basis of interest, Chapter 2 gives an overview on the main directional distributions for orthonormal basis, such as real Bingham Langevin, real Bingham and real Langevin. Then, we introduce a generalization of the aforementioned distributions to the case of random matrix variables with complex entries, i.e., the complex generalized Bingham Langevin distribution, and we propose a practical sampling method adapted to the proposed distribution. In Chapter 3, we present a specific model named low-rank structured covariance matrix which is common for radar processing and is used in the following chapters as our parametric data model.

- In Part II, we are particularly interested in signals being contained in a low-rank subspace. *i*) First, we propose new Bayesian subspace estimators (maximum a posterior and minimum mean square distance based estimators) in the context of compound Gaussian distributed sources embedded in white Gaussian noise. To this aim, we assign to the signal subspace the complex generalized Bingham Langevin distribution. *ii*) Second, in order to be robust to the possible presence of outliers, we design our estimators taking into account the compound Gaussian distributed sources embedded in heterogeneous noise (compound Gaussian outliers plus white additive Gaussian noise). Specifically, we develop algorithms to compute the maximum a posterior estimator for the aforementioned models based on the majorization minimization algorithm. Then, we derive Gibbs-sampler based algorithm to compute the minimum mean square distance estimator for each data model. The interest of the proposed approach is illustrated by comparing the detection performance of the newly proposed low-rank adaptive normalized matched filter on a real data set for a space

time adaptive processing application in challenging configurations, i.e., for small sample support and/or in presence of outliers.

- In Part III, we derive three new statistical tests in the context of structured covariance matrices. Specifically, we consider low-rank signal component plus white Gaussian noise structure. Our aim is to test the similarity between the principal components of a group of covariance matrices where the information usually lies in. The proposed detection schemes are derived using the generalized likelihood ratio test. As the formulation of the proposed tests implies a non-trivial optimization problem, we derive appropriate majorization minimization algorithms in order to evaluate them. Finally, the benefits of the proposed methods are illustrated for a change detection application on a UAVSAR dataset.

0.3 List of publications

Peer-reviewed journals
[J2] R. Ben Abdallah, A. Breloy, M. N. El Korso and D. Lautru, "Bayesian robust subspace estimation in presence of compound Gaussian sources", Elsevier Signal Processing Journal, accepted.
[J1] R. Ben Abdallah, A. Mian, A. Breloy, A. Taylor, M. N. El Korso and D. Lautru, "Detection Methods Based on Structured Covariance Matrices for Multivariate SAR Images Processing", IEEE Geoscience and Remote Sensing Letters, vol. 16, no. 7, pp. 1160-1164, 2019.

International conferences
[IC3] R. Ben Abdallah, A. Breloy, A. Taylor, M. N. El Korso and D. Lautru, "Signal subspace change detection in structured covariance matrices", 27th European Signal Processing Conference EUSIPCO'19, A Coruna, Spain, 2019.
[IC2] R. Ben Abdallah, A. Breloy, M. N. El Korso and D. Lautru, "Bayesian Robust Signal Subspace Estimation in Non-Gaussian Environment", 27th European Signal Processing Conference EUSIPCO'19, A Coruna, Spain, 2019.
[IC1] R. Ben Abdallah, A. Breloy, M. N. El Korso, D. Lautru and H. Ouslimani, "Minimum Mean Square Distance Estimation of Subspaces in presence of Gaussian sources with application to STAP detection", 7th International Conference on New Computational Methods for Inverse Problems, IOP Conf. Series: Journal of Physics: Conf. Series 904, Paris, 2017.

National conferences
[NC2] R. Ben Abdallah, A. Breloy, A. Taylor, M. N. El Korso and D. Lautru, "Détection de changement de sous-espace signal de matrices de covariance structurées", Lille, France, GRETSI 2019, August 2019.
[NC1] R. Ben Abdallah, A. Breloy, M. N. El Korso, D. Lautru and H. Ouslimani, "Estimation de sous-espaces en présence de sources gaussiennes avec application à la détection STAP", Proc. GRETSI 2017, Juan-les-pins, France, September 2017.

Presentations and workshops
[C4] Adaptive processing for radar based on Bayesian Subspace methods, Oral presentation, Seminar MARGARITA-Centrale Supelec, January 2019.
[C3] Detection Methods Based on Structured Covariance Matrices for Multivariate SAR Images Processing, Oral presentation , GDR ISIS seminar "Extraction d'attributs et apprentissage pour

l'analyse des images de télédétection", October 2018.
[C2] Bayesian Robust Signal Subspace Estimation in Non-Gaussian Environment, Poster session, Journée porte ouverte -University Nanterre, September 2018.
[C1] Minimum Mean Square Distance Estimation of Subspaces in presence of Gaussian sources with application to STAP detection, Oral presentation, Summer school GRETSI 2017-Peyresq, June 2017.

Part I

Distributions and matrix structures

Chapter 1

Distributions on vectors

1.1 General settings

In signal processing applications, the received data is generally composed of multiple contributions, that can be modeled by random processes. Typically, $\mathbf{z} \in \mathbb{C}^N$ denotes a sample vector of probability density function (p.d.f.) $f(\mathbf{z}|\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is an appropriate vector of parameters, involved in the parametric model of $\mathbf{z}$. The estimation of the parameter $\boldsymbol{\theta}$ is at the cornerstone of a plethora of applications and algorithms. To name a few, here are some examples where an estimation step is required:

- In direction of arrival (DOA) estimation, the parameter to be estimated $\boldsymbol{\theta}$ is the set of angle of arrival of the sources [1].
- In principal component analysis (PCA), $\boldsymbol{\theta}$ denotes the eigenvectors of the covariance matrix [2].
- In adaptive filtering or adaptive detection, $\boldsymbol{\theta}$ corresponds generally to the covariance matrix of the samples [3]. An estimate of this parameter can then be used as a plug-in in the formulation of standard filters/detectors [4].
- In the context of change detection, several sample sets (e.g., observations at different times) are gathered and indexed by i. The problem then consists in deciding if a change occurred in the unknown parameter of interest $\boldsymbol{\theta}_i$ for the corresponding set i [5].

In order to derive estimation processes, the distribution of the vector $\mathbf{z}$ has to be set depending on the application. An accurate representation of the underlying physics will obviously condition the performance (e.g., estimation accuracy) that will be eventually achieved. Therefore, the choice of the data model and its p.d.f. is crucial. In the following, we present some standard distributions of the data vector $\mathbf{z}$ that are commonly used in the literature. Specifically, we consider the Gaussian, compound Gaussian (CG) and mixture model (CG plus white Gaussian noise (WGN)) distributions.

1.2 Complex multivariate circular Gaussian distribution

In statistical signal processing communities, the complex multivariate Gaussian distribution [6] (also referred to as Normal distribution) is probably the most commonly invoked distribution for random measurements. Indeed, it has a practical formulation, it has a strong theoretical

justification given by the central limit theorem [7] and it has a good empirical fit with numerous empirically measured data distribution.

Definition 1.2.1 ***The complex multivariate Gaussian distribution***
A given random vector $\mathbf{z}$ *follows complex multivariate Gaussian (normal) distribution if its p.d.f. reads as*

$$f(\mathbf{z}) = \frac{1}{\pi^N |\boldsymbol{\Sigma}|} \exp\left\{-(\mathbf{z}-\boldsymbol{\mu})^{\mathrm{H}} \boldsymbol{\Sigma}^{-1} (\mathbf{z}-\boldsymbol{\mu})\right\} \tag{1.1}$$

where

i) $\boldsymbol{\mu}$ *is the mean of the random vector* $\mathbf{z}$*, i.e.,* $\mathbb{E}\{\mathbf{z}\} = \boldsymbol{\mu}$*.*

ii) $\boldsymbol{\Sigma} \in \mathcal{H}_N^+$ *is the covariance matrix, which presents the correlations and the variances between entries of* $\mathbf{z}$*, also referred to as the second order moment, i.e.,*

$$\boldsymbol{\Sigma} = \mathbb{E}\{(\mathbf{z}-\boldsymbol{\mu})(\mathbf{z}-\boldsymbol{\mu})^H\} \tag{1.2}$$

We denote:

$$\mathbf{z} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{1.3}$$

The majority of signal processing algorithms derived in the literature assume that the additive noise has a Gaussian distribution. As an example, this provides a good empirical fit to thermal noise in electronic systems. However, various studies [8, 9] have shown that signals in most high resolution array processing applications are non-Gaussian and exhibit impulsive characteristics. Unfortunately, conventional signal processing algorithms developed for Gaussian noise are known to perform poorly in the presence of non-Gaussian noise [10].

A possible solution is to develop statistical signal processing algorithms adapted for non-Gaussian distributed observations. Specifically, we will consider the CG distributions.

1.3 Compound Gaussian (CG) distribution

The CG distribution [11, 12] has been widely employed in the statistical signal processing literature e.g., in image processing [13–15], and for modeling radar clutter [16, 17]. A given random variable having a CG distribution is also referred to as spherically invariant random vector [11]. Such family of distributions provides a good fit to non-Gaussian real data, and allows to develop robust estimation procedures [18–20].

Remark 1.3.1 *We note that the CG is a special case of the complex elliptically symmetric (CES) [12]. The CES distributions constitute a wider class of distributions that include, among others, the class of CG distributions as special cases. We will focus on CG because it has a simple and practical stochastic representation and it still covers a large panel of standard non-Gaussian distributions. Furthermore, the interest of this modeling is reinforced by a good correspondence with empirical distributions of real data [16, 17].*

Definition 1.3.1 ***Compound Gaussian distribution***
A N-dimensional CG observation is represented as a product of two statistically independent components, i.e., if $\mathbf{z} \in \mathbb{C}^N$ *follows a CG distribution, denoted* $\mathbf{z} \sim \mathcal{CG}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, f_\tau)$*, it has the following stochastic representation*

$$\mathbf{z} \overset{d}{=} \boldsymbol{\mu} + \sqrt{\tau}\mathbf{d}, \tag{1.4}$$

where

i) $\boldsymbol{\mu}$ *is the mean of the random vector* $\mathbf{z}$*, i.e.,* $\mathbb{E}\{\mathbf{z}\} = \boldsymbol{\mu}$.

ii) τ *is a positive random scalar, called texture, of p.d.f.* f_τ*. This parameter is statistically independent of* $\mathbf{d}$.

iii) $\mathbf{d}$ *follows a zero-mean multivariate complex Gaussian distribution of covariance matrix* $\boldsymbol{\Sigma}$*, denoted,* $\mathbf{d} \sim \mathcal{CN}(0, \boldsymbol{\Sigma})$*. The parameter* $\boldsymbol{\Sigma} \in \mathcal{H}_N^+$ *is referred to as the scatter matrix.*

Notice that if $\mathbb{E}\{\tau\} < \infty$*, the covariance matrix of* $\mathbf{z}$ *exists and is proportional to the scatter matrix, i.e.,* $\mathbb{E}\{(\mathbf{z}-\boldsymbol{\mu})(\mathbf{z}-\boldsymbol{\mu})^H\} = \mathbb{E}\{\tau\}\boldsymbol{\Sigma}$.

Remark 1.3.2 *We note that CG distributions are not uniquely defined. Indeed, let us consider* $\mathbf{z} \sim \mathcal{CG}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, f_\tau)$ *and* $\mathbf{z}' \sim \mathcal{CG}(\boldsymbol{\mu}, c\boldsymbol{\Sigma}, f_{\tau'})$ *with* $\tau' = \frac{\tau}{c}$ *and* $c \in \mathbb{R}_+^*$*, then,* $\mathbf{z} \overset{d}{=} \mathbf{z}'$*. Consequently, for a fixed mean* $\boldsymbol{\mu}$*, the couples* $(\boldsymbol{\Sigma}, \tau)$ *and* $(c\boldsymbol{\Sigma}, \frac{1}{c}\tau)$ *define the same distribution. In order to avoid any ambiguity, we may impose an arbitrary constraint on the scatter covariance matrix* $\boldsymbol{\Sigma}$ *such as* $Tr\{\boldsymbol{\Sigma}\} = 1$*, or on the texture parameter* τ *for instance* $\mathbb{E}\{\tau\} = 1$ *(when* $\mathbb{E}\{\tau\} < \infty$*). In this thesis, we will use the texture constraint in our simulations, i.e.,* $\mathbb{E}\{\tau\} = 1$.

The p.d.f. of a random vector $\mathbf{z} \sim \mathcal{CG}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, f_\tau)$ is defined as

$$f(\mathbf{z}) = \pi^{-M}|\boldsymbol{\Sigma}|^{-1} \int_0^\infty \tau^{-M} \exp\left\{\frac{-(\mathbf{z}-\boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu})}{\tau}\right\} f_\tau(\tau) d\tau \tag{1.5}$$

Conditionally to the texture, the random vector $\mathbf{z}$ has thus the following distribution:

$$\mathbf{z}|\tau \sim \mathcal{CN}(\boldsymbol{\mu}, \tau\boldsymbol{\Sigma}) \tag{1.6}$$

The CG encloses some usual multivariate distributions [12] such as:

Example 1.3.1 ***The complex multivariate Gaussian distribution***
A random vector $\mathbf{z}$ *is said to have the complex multivariate Gaussian distribution, if* $\mathbf{z} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*, we have* $\mathbf{z} \sim \mathcal{CG}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, f_\tau)$ *where*

$$f_\tau(\tau) = \delta_{\tau,1} \tag{1.7}$$

Example 1.3.2 ***The K-distribution***
A random vector $\mathbf{z}$ *is said to have the multivariate K-distributed, if the texture follows a Gamma distribution of p.d.f.,*

$$f(\tau, \nu) = \frac{\tau^{\nu-1}\nu^\nu \exp\{-\nu\tau\}}{\Gamma(\nu)} \tag{1.8}$$

denoted $\tau \sim \Gamma(\nu, 1/\nu)$ *with* $\nu > 0$ *represents the shape parameter.*

Remark 1.3.3 *The Gaussian distribution is a limit case of the K-distribution when* $\nu \to \infty$*, If* $\nu \to 0$*, the resulting K-distribution is heavy tailed.*

Example 1.3.3 ***The student t-distribution***
A random vector $\mathbf{z}$ *is said to have the complex multivariate student t-distribution with degrees of freedom* ν*, if it has the CG representation* (1.4) *where the texture* $\tau \overset{d}{=} \nu/x$ *with* $x \sim \Gamma(\nu/2, 2)$ *or equivalently stated* $\tau^{-1} \sim \Gamma(\nu/2, 2/\nu)$.

Depending on f_τ, the CG representation can lead to various standard multivariate distributions for $\mathbf{z}$ (as given in aforementioned examples). In order to design algorithms that are robust to this whole family of distributions, we will generally consider that the texture τ is unknown and deterministic [21–25]. Indeed, in most applications, a strong prior information on the texture distribution is not available. Thus, this assumption has been widely used in several array processing applications [26–29] since it offers an interesting robustness-performance trade off [18,30,31]. This model will be denoted as below

Example 1.3.4 ***CG distribution with unknown deterministic texture***
Consider a set of secondary data $\{\mathbf{z}_k\}$ following CG distribution while assuming a deterministic texture τ_k for each sample $\mathbf{z}_k$. We denote this model by

$$\mathbf{z}_k \sim \mathcal{CG}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \delta_{\tau,\tau_k}), \text{ or equivantely, } \mathbf{z}_k|\tau_k \sim \mathcal{CN}(\boldsymbol{\mu}, \tau_k\boldsymbol{\Sigma}),\ \forall k \tag{1.9}$$

This model is also referred to as mixture of scaled Gaussian.

1.4 Mixture Model (CG plus white Gaussian noise (WGN))

In this thesis, an other focus will be on the additive mixtures of CG and WGN. This modeling has been considered in many modern robust signal processing applications, as it can account for local power fluctuations of the sources, as well as the thermal noise. As an example, this formulation is useful to model clutter plus thermal noise. In practice, this modeling has been used for detection in heterogeneous environment [32,33] and for robust structured covariance matrix estimation in [30,34].

Definition 1.4.1 ***CG plus WGN mixture model***
For this mixture model, the sample vector $\mathbf{z}$ is drawn as

$$\mathbf{z} = \mathbf{s} + \mathbf{n} \tag{1.10}$$

where

- $\mathbf{s}$ *follows the CG distribution, i.e., $\mathbf{s} \sim \mathcal{CG}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, f_\tau)$. For instance, such signal may model the ground response when considering correlation and/or power fluctuations.*
- $\mathbf{n} \sim \mathcal{CN}(0, \sigma^2\mathbf{I}_N)$ *is an additive WGN which represents the contribution of the thermal noise.*

We note that the signal $\mathbf{s}$ conditionally to the texture parameter τ has a multivariate Gaussian distribution

$$\mathbf{s}|\tau \sim \mathcal{CN}(\boldsymbol{\mu}, \tau\boldsymbol{\Sigma}) \tag{1.11}$$

Eventually, the data vector conditionally to τ is drawn as $\mathbf{z}|\tau \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_\tau)$ where

$$\boldsymbol{\Sigma}_\tau = \tau\boldsymbol{\Sigma} + \sigma^2\mathbf{I}_N \tag{1.12}$$

Its p.d.f. reads as

$$f(\mathbf{z}|\tau) = \frac{1}{\pi^N|\boldsymbol{\Sigma}_\tau|}\exp\left\{-(\mathbf{z}-\boldsymbol{\mu})^{\mathrm{H}}\boldsymbol{\Sigma}_\tau^{-1}(\mathbf{z}-\boldsymbol{\mu})\right\} \tag{1.13}$$

Then, the p.d.f. of the data vector $\mathbf{z}$ is obtained as

$$f(\mathbf{z}) = \int f(\mathbf{z}|\tau)f_\tau(\tau)d\tau \tag{1.14}$$

Remark 1.4.1 *This distribution is not a particular case neither of CG nor of CES except when the CG distribution is a Gaussian one.*

Remark 1.4.2 *The derivation of the above p.d.f. is not obvious and may introduce non closed form expressions. Furthermore, in order to develop a robust algorithm to various multivariate distributions, the texture τ will be assumed unknown and deterministic in our derivations.*

Definition 1.4.2 ***Mixture model with unknown deterministic textures***
Let us consider independent $\{\mathbf{z}_k\}$ following a CG plus WGN see definition 1.4.1. For assumed unknown deterministic texture of the CG component (as in example 1.3.4), the likelihood of $\{\mathbf{z}_k\}$ reads

$$\prod_{k=1}^{K} f(\mathbf{z}_k|\tau_k) = \prod_{k=1}^{K} \frac{1}{\pi^N |\mathbf{\Sigma}_{\tau_k}|} \exp\left\{-(\mathbf{z}_\mathrm{k}-\boldsymbol{\mu})^\mathrm{H} \mathbf{\Sigma}_{\tau_\mathrm{k}}^{-1} (\mathbf{z}_\mathrm{k}-\boldsymbol{\mu})\right\} \tag{1.15}$$

To conclude, this chapter presented several distributions that can account for various physical phenomenon such as impulsiveness. The distribution presented in definition 1.4.1 offers an interesting and realistic choice for modeling the samples in array processing applications, as it accounts for a mixture of (possibly non-Gaussian) contributions. Moreover, without additional information, choosing to model the textures as unknown deterministic offers an interesting performance-robustness trade off on the whole class of CG distributions.

The next chapter will present distributions on orthonormal basis, which will play a central role in this thesis.

Chapter 2

Distributions on orthonormal basis

In this chapter, we present distributions on orthonormal basis as well as their corresponding sampling techniques. First, let us introduce some distributions from the field of directional data analysis.

The subscript $\tilde{.}$ will denote a real quantity (matrix or vector). Furthermore, we define the following spaces:

Definition 2.0.1 ***Real Stiefel manifold*** $\mathcal{U}_N^R(\mathbb{R})$
$\mathcal{U}_N^R(\mathbb{R})$ *denotes the set of* $N \times R$ *semi-unitary matrices, i.e., tall matrices whose columns form an orthonormal basis,*

$$\mathcal{U}_N^R(\mathbb{R}) = \left\{\mathbf{U} \in \mathbb{R}^{N\times R} | \mathbf{U}^T\mathbf{U} = \mathbf{I}_R\right\}$$

Definition 2.0.2 ***Complex Stiefel manifold*** $\mathcal{U}_N^R$
$\mathcal{U}_N^R$ *denotes the set of* $N \times R$ *complex semi-unitary matrices, i.e., tall matrices whose columns form an orthonormal basis,*

$$\mathcal{U}_N^R = \left\{\mathbf{U} \in \mathbb{C}^{N\times R} | \mathbf{U}^H\mathbf{U} = \mathbf{I}_R\right\}$$

Probability distributions and statistical inference for data in $\mathcal{U}_N^R(\mathbb{R})$ and $\mathcal{U}_N^R$ have been developed primarily in the spatial statistics literature [35], starting with the circle $\mathcal{U}_2^1(\mathbb{R})$ and the sphere $\mathcal{U}_3^1(\mathbb{R})$, then extended to higher dimensions. Overviews of directional probability distributions can be found in [35,36]. In the following, we present some distributions that will be used in this thesis.

2.1 Real Bingham Langevin distribution

Definition 2.1.1 ***Real Bingham Langevin (RBL) distribution***
The RBL distribution (also called Bingham von Mises Fisher distribution) [37] is parametrized by the matrices $\tilde{\mathbf{A}} \in \mathcal{S}_N^+$, $\tilde{\mathbf{B}} \in \mathbb{R}^{R\times R}$ *and* $\tilde{\mathbf{C}} \in \mathbb{R}^{N\times R}$ *where* $\tilde{\mathbf{B}}$ *is a diagonal matrix. The matrix* $\tilde{\mathbf{U}} \in \mathcal{U}_N^R(\mathbb{R})$ *is said to follow a RBL distribution, denoted* $\tilde{\mathbf{U}} \sim \mathrm{RBL}(\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}})$*, if its p.d.f. is given as*

$$p_{\mathrm{RBL}}(\tilde{\mathbf{U}}) = c_{\mathrm{RBL}}(\tilde{\mathbf{C}}, \tilde{\mathbf{A}}, \tilde{\mathbf{B}})\mathrm{etr}\left\{\tilde{\mathbf{C}}^T\tilde{\mathbf{U}} + \tilde{\mathbf{B}}\tilde{\mathbf{U}}^T\tilde{\mathbf{A}}\tilde{\mathbf{U}}\right\} \tag{2.1}$$

where $c_{\mathrm{RBL}}(\tilde{\mathbf{C}}, \tilde{\mathbf{A}}, \tilde{\mathbf{B}})$ *is the normalizing constant so that* p_{RBL} *defines a p.d.f. [36].*

Remark 2.1.1 *In general case, the normalizing constant of the RBL does not have a closed form expression. The evaluation of* $c_{\mathrm{RBL}}(\tilde{\mathbf{C}}, \tilde{\mathbf{A}}, \tilde{\mathbf{B}})$ *is theoretically investigated in [38, 39] and references therein. Nevertheless, this constant will not be needed in the derivation of this thesis.*

The RBL distribution encompasses several special cases, such as :

Example 2.1.1 ***Real Bingham distribution***
The real Bingham is a special case of the RBL distribution for which $\tilde{\mathbf{C}} = \mathbf{0}$ *and* $\tilde{\mathbf{B}} = \mathbf{I}_R$. *The matrix* $\tilde{\mathbf{U}} \in \mathcal{U}_N^R(\mathbb{R})$ *is said to have a real Bingham distribution with parameter* $\tilde{\mathbf{A}}$, *denoted* $\tilde{\mathbf{U}} \sim B(\tilde{\mathbf{A}})$, *if its p.d.f. reads*

$$p_B(\tilde{\mathbf{U}}) = {}_1F_1(\frac{1}{2}R, \frac{1}{2}N, \tilde{\mathbf{A}})^{-1}\text{etr}\{\tilde{\mathbf{U}}^T\tilde{\mathbf{A}}\tilde{\mathbf{U}}\} \tag{2.2}$$

where $\tilde{\mathbf{A}} \in \mathcal{S}_N^+$ *characterizes the center of distribution and* ${}_1F_1(\frac{1}{2}R, \frac{1}{2}N, \tilde{\mathbf{A}})$ *is an hypergeometric functions of matrix argument defined in [36].*

Example 2.1.2 ***Real Langevin distribution***
The real Langevin is a special case of the RBL distribution for which $\tilde{\mathbf{A}} = \mathbf{0}$ *and* $\tilde{\mathbf{B}} = \mathbf{0}$. *The matrix* $\mathbf{U} \in \mathcal{U}_N^R(\mathbb{R})$ *is said to have real Langevin distribution (also called von Mises Fisher distribution) with parameter* $\tilde{\mathbf{C}}$, *denoted* $\mathbf{U} \sim L(\tilde{\mathbf{C}})$, *if its p.d.f. reads as*

$$p_{\text{L}}(\tilde{\mathbf{U}}) = {}_0F_1(\frac{1}{2}N, \frac{1}{4}\tilde{\mathbf{C}}^T\tilde{\mathbf{C}})^{-1}\text{etr}\{\tilde{\mathbf{C}}^T\tilde{\mathbf{U}}\} \tag{2.3}$$

where $\tilde{\mathbf{C}}$ *characterizes the center of distribution and* ${}_0F_1(\frac{1}{2}N, \frac{1}{4}\tilde{\mathbf{C}}^T\tilde{\mathbf{C}})$ *is the hypergeometric function of matrix argument defined in [36].*

Remark 2.1.2 *Notice that as defined in* (2.2) *and* (2.3), p_{B} *is invariant by rotation* $\tilde{\mathbf{U}}' = \tilde{\mathbf{U}}\mathbf{Q}$, $\forall \mathbf{Q} \in \mathcal{U}_R^R$ *while,* p_{L} *is not. Hence, the Bingham distribution* p_{B} *characterizes a distribution for the range space of* $\tilde{\mathbf{U}}$ *(cf. Section 3.2), whereas the Langevin distribution* p_{L} *characterizes the distribution of the orthogonal basis* $\tilde{\mathbf{U}}$.

To illustrate these distributions, Table 2.1 displays the p.d.f. of the real Bingham distribution

$$\mathbf{u} \sim \text{B}(\kappa\bar{\mathbf{u}}\bar{\mathbf{u}}^T) \tag{2.4}$$

and the real Langevin distributions

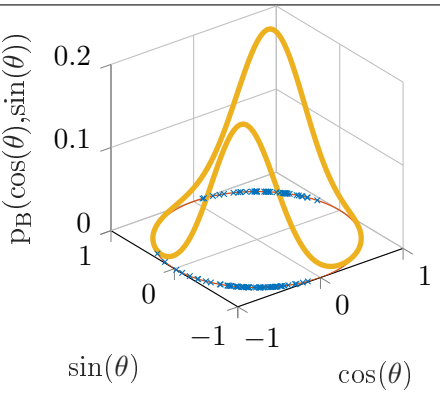
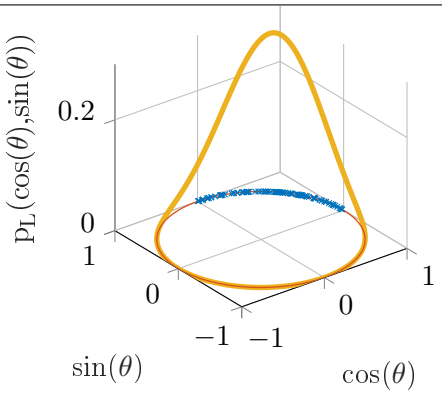
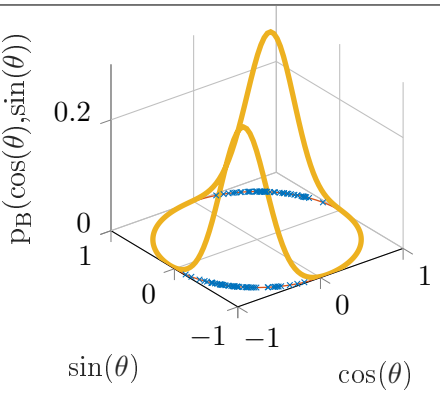
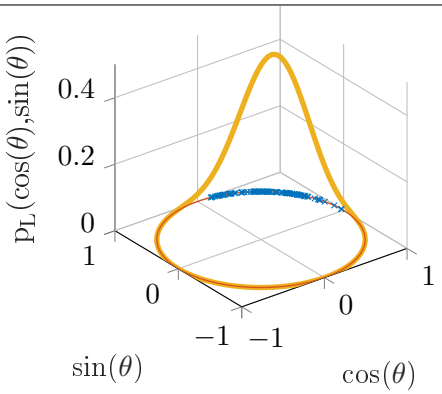
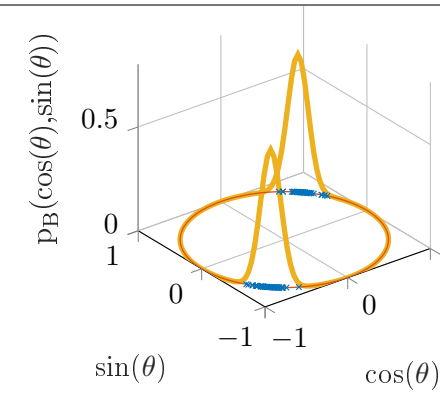
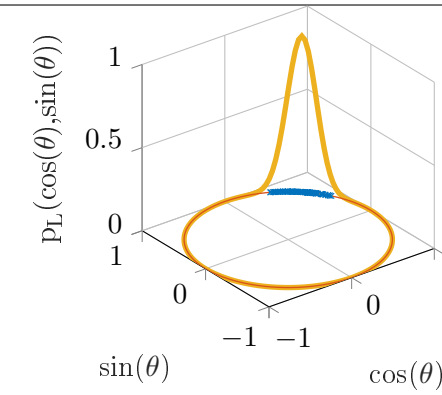
$$\mathbf{u} \sim \text{L}(\kappa\bar{\mathbf{u}}) \tag{2.5}$$

on a unit circle $\mathcal{U}_2^1(\mathbb{R})$ where $\bar{\mathbf{u}} = [1/\sqrt{2}, 1/\sqrt{2}]$ defines the center of distribution and κ is a concentration parameter. We note that for high value of $\kappa \in \mathbb{R}_+^*$, i.e., $\kappa = 50$, the generated samples $\mathbf{u} \sim \text{L}(\kappa\bar{\mathbf{u}})$ are more concentrated around the center $\bar{\mathbf{u}}$. For real Bingham distribution, the samples are gathered on both sides $\bar{\mathbf{u}}$ and $-\bar{\mathbf{u}}$, since, this distribution characterizes the quantity $\mathbf{u}\mathbf{u}^T$ (cf. previous remark).

2.2 Complex generalized Bingham Langevin (CGBL) distribution

In this section, we present the CGBL distribution as a generalization of the aforementioned usual directional statistics to the case of matrix variables with complex entries: the CGBL is a probability distribution on the set of the Stiefel manifold $\mathcal{U}_N^R$ which combines linear and quadratic terms.

Table 2.1: The p.d.f. of Bingham and Langevin distributions for various values of concentration parameter κ and for a prior center $\bar{\mathbf{u}} = [1/\sqrt{2}, 1/\sqrt{2}]$ on a unit circle and 100 samples generated according to these distributions

	Bingham distribution	Langevin distribution
$\kappa = 5$		
$\kappa = 10$		
$\kappa = 50$		

Definition 2.2.1 ***Complex generalized Bingham Langevin (CGBL) distribution*** *The CGBL is parametrized by the set of Hermitian matrices* $\{\mathbf{A}_r\} \subset \mathcal{H}_N^+$ *and the matrix* $\mathbf{C} =$

$[\mathbf{c}_1, \ldots, \mathbf{c}_R]$. *We denote* $\mathbf{U} \sim \mathrm{CGBL}(\mathbf{C}, \{\mathbf{A}_r\})$ *when the p.d.f. of* $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_R]$ *on* $\mathcal{U}_N^R$ *reads*

$$p_{CGBL}(\mathbf{U}) = c_{\mathrm{CGBL}}(\mathbf{C}, \{\mathbf{A}_r\}) \exp\left\{ \sum_{r=1}^{R} \mathrm{Re}\{\mathbf{c}_r^H \mathbf{u}_r\} + \mathbf{u}_r^H \mathbf{A}_r \mathbf{u}_r \right\} \tag{2.6}$$

where $c_{\mathrm{CGBL}}(\mathbf{C}, \{\mathbf{A}_r\})$ *is a normalizing constant.*

Remark 2.2.1 *From* (2.6), p_{CGBL} *promotes the concentration of each vector* $\mathbf{u}_r$ *around* $\mathbf{c}_r$ *and each range space* $\mathbf{u}_r\mathbf{u}_r^H$ *around the subspace associated to the strongest eigenvalue of the Hermitian matrix* $\mathbf{A}_r$. *Typically, if* $\mathbf{A}_r = \mathbf{A}$, $\forall r \in [\![1, R]\!]$, *the range space* $\mathbf{U}\mathbf{U}^H$ *tends to be close to the dominant space of* $\mathbf{A}$.

Remark 2.2.2 *Notice that for the quadratic forms involved in the CGBL definition, we used the formulation* $\sum_{r=1}^{R} \mathbf{u}_r^H \mathbf{A}_r \mathbf{u}_r$ *rather than the immediate counterpart of the real distribution* $\mathrm{Tr}\{\mathbf{B}\mathbf{U}^H\mathbf{A}\mathbf{U}\}$. *This offers a more general parameterization that will be necessary in the derivations of Part II.*

In the following, we list some special cases of the CGBL distribution [35]:

Example 2.2.1 ***Complex invariant Bingham (CIB) distribution***
The CIB is a special case of the CGBL where $\mathbf{C} = \mathbf{0}$ *and* $\mathbf{A}_r = \kappa\bar{\mathbf{U}}\bar{\mathbf{U}}^H$, $\forall r \in [\![1, R]\!]$ *where* $\bar{\mathbf{U}} \in \mathcal{U}_N^R$ *represents the center of the distribution and* κ *denotes the concentration parameter. We denote* $\mathbf{U} \sim CIB(\kappa, \bar{\mathbf{U}}\bar{\mathbf{U}}^H)$ *when* $\mathbf{U}$ *has as p.d.f. of the form*

$$p_{CIB}(\mathbf{U}) = c_{CIB}(\kappa, \bar{\mathbf{U}}) \mathrm{etr}\left\{\kappa \mathbf{U}^H \bar{\mathbf{U}}\bar{\mathbf{U}}^H \mathbf{U}\right\} \tag{2.7}$$

in which $c_{CIB}(\kappa, \bar{\mathbf{U}})$ *denotes the normalizing constant.*

Remark 2.2.3 *As for the real case, note that the* $p_{CIB}(\mathbf{U})$ *is invariant by rotation* $\mathbf{U}' = \mathbf{U}\mathbf{Q}$, $\forall \mathbf{Q} \in \mathcal{U}_R^R$. *This means that* p_{CIB} *characterizes a distribution for the subspace represented by the orthogonal projector* $\mathbf{U}\mathbf{U}^H$, *which will be its main interest.*

Example 2.2.2 ***Complex Langevin (CL) distribution***
The CL is a special case of the CGBL for which $\{\mathbf{A}_r = \mathbf{0}\}$ *and* $\mathbf{C} = \kappa\bar{\mathbf{U}}$ *where* $\bar{\mathbf{U}}$ *is the center of distribution and* κ *is the concentration parameter. We denote* $\mathbf{U} \sim \mathrm{CL}(\kappa, \bar{\mathbf{U}})$ *when* $\mathbf{U}$ *has as p.d.f. of the following form*

$$p_{\mathrm{CL}}(\mathbf{U}) = {}_0F_1(\frac{1}{2}N, \frac{1}{4}\bar{\mathbf{U}}^H\bar{\mathbf{U}})^{-1} \mathrm{etr}\left\{\kappa \mathrm{Re}\{\bar{\mathbf{U}}^H\mathbf{U}\}\right\} \tag{2.8}$$

with ${}_0F_1(\frac{1}{2}N, \frac{1}{4}\bar{\mathbf{U}}^H\bar{\mathbf{U}})^{-1}$ *is the normalizing constant for this distribution.*

Example 2.2.3 ***Complex generalized Bingham (CGB) distribution***
The CGB is a special case of the CGBL distribution for which $\mathbf{C} = \mathbf{0}$. *We denote* $\mathbf{U} \sim \mathrm{CGB}(\{\mathbf{A}_r\})$ *when its p.d.f. is defined as*

$$p_{\mathrm{CGB}}(\mathbf{U}) = c_{\mathrm{CGB}}(\{\mathbf{A}_r\}) \exp\left\{ \sum_{r=1}^{R} \mathbf{u}_r^H \mathbf{A}_r \mathbf{u}_r \right\} \tag{2.9}$$

in which $c_{\mathrm{CGB}}(\{\mathbf{A}_r\})$ *denotes the normalizing constant of* p_{CGB}.

2.3 Sampling of the CGBL distribution

In the literature, several sampling methods are proposed in order to simulate random matrices drawn from the aforementioned distribution. Such approaches are based on Markov Chain Monte Carlo (MCMC) methods [40] and/or acceptance rejection schemes [41]. In Appendix A, we present a general method to draw samples as $\mathbf{U} \sim \text{CGBL}(\mathbf{C}, \{\mathbf{A}_r\})$. The proposed sampling technique is based on the results of [40] and [41] and summed as follows:

i) The generation of $\mathbf{U}$ is obtained as a Markov chain on the columns $\{\mathbf{u}_r\}$, as proposed in [40, 42].

ii) The generation of each $\mathbf{u}_r$ is obtained by using the results of [41]. In [41], the authors rely on an acceptance rejection method to sample $\{\mathbf{u}_r\}$. Notice that for sampling a vector following a RBL distribution, [40] proposed a Markov chain on the entries of the vectors. Both methods allow to sample the desired distribution, however, the acceptance rejection scheme from [41] allows to reduce significantly the computation time.

iii) Notice that [40] and [41] proposed methods adapted to real distributions. In order to generalize these sampling techniques to the case of complex distributions, we resort to the change of variables proposed in [35] (refer to Appendix A for more details).

Chapter 3

Low rank structures

This chapter presents a particular structure of covariance matrix and related parameterizations. Notably, we introduce the so called low rank structured matrices, which are common for radar processing. These structures have been successfully studied and applied to the context of radar detection [43–45].

3.1 Preliminary definitions

In the following, we list some useful preliminary definitions:

Definition 3.1.1 ***Singular value decomposition (SVD)***
The SVD of a given matrix $\mathbf{A} \in \mathbb{C}^{M \times N}$ *is given as*

$$\mathbf{A} \overset{\text{SVD}}{=} \mathbf{U}\mathbf{S}\mathbf{V}^H \tag{3.1}$$

with $\mathbf{U} \in \mathcal{U}_M^M$, $\mathbf{S}$ *is a diagonal* $M \times N$ *matrix where its diagonal entries are known as the singular values of* $\mathbf{A}$ *and* $\mathbf{V} \in \mathcal{U}_N^N$.

Definition 3.1.2 ***Eigenvalue decomposition (EVD)***
For an hermitian matrix $\mathbf{A} \in \mathbb{C}^{N \times N}$, *we can group the eigenvalues of* $\mathbf{A}$ *in an* $N \times N$ *diagonal matrix* $\mathbf{\Lambda} = diag(\{\lambda_n\})$ *and their eigenvectors in a* $N \times N$ *unitary matrix* $\mathbf{E} = [\mathbf{e}_1, \ldots, \mathbf{e}_N]$, *then,*

$$\mathbf{A}\mathbf{E} = \mathbf{E}\mathbf{\Lambda} \tag{3.2}$$

Furthermore, $\mathbf{A}$ *can be factorized as*

$$\mathbf{A} \overset{\text{EVD}}{=} \mathbf{E}\mathbf{\Lambda}\mathbf{E}^H \tag{3.3}$$

In order to avoid any ambiguity in this definition, we assume ordered eigenvalues as $\lambda_1 \geq \cdots \geq \lambda_N > 0$, *and an arbitrary element of each* $\mathbf{e}_j$ *(e.g., the first entry) can be assumed to be real positive.*

Definition 3.1.3 ***Rank of a matrix***
Let $\mathbf{A}$ *be a matrix. The rank of* $\mathbf{A}$ *is the maximal number of linearly independent column vectors in* $\mathbf{A}$. *Equivalently, it corresponds to the number of non zero singular values in the SVD of* $\mathbf{A}$.

3.2 Subspaces

In order to represent a subspace in this thesis, we need to define the following notions:

Definition 3.2.1 ***Range space of matrix***
Let $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_R]$ *be a* $N \times R$ *complex matrix where rank(*$\mathbf{A}$*)=*R*, then the range of* $\mathbf{A}$ *is the linear space generated by the column vectors* $\{\mathbf{a}_r\}$*, denoted*

$$span(\mathbf{A}) = \left\{ \mathbf{x} \in \mathbb{C}^N | \mathbf{x} = \sum_{r=1}^{R} \alpha_r \mathbf{a}_r; \quad \forall \alpha_r \in \mathbb{C} \right\} \tag{3.4}$$

Remark 3.2.1 *Notice that for a given space full rank matrix* $\mathbf{A} \in \mathbb{C}^{N \times R}$*, there exists an orthonormal basis, i.e.,* $\mathbf{U} \in \mathcal{U}_N^R$ *such that*

$$span(\mathbf{A}) = span(\mathbf{U}) \tag{3.5}$$

in which $\mathbf{U}$ *is an orthonormal basis of span(*$\mathbf{A}$*). However, the existence of* $\mathbf{U}$ *is not unique, i.e., span(*$\mathbf{U}$*)=span(*$\mathbf{UQ}$*),* $\forall \mathbf{Q} \in \mathcal{U}_R^R$*.*

Definition 3.2.2 ***The set of orthogonal projection matrices*** $\mathcal{G}_N^R$
The set of orthogonal projection matrix $\mathcal{G}_N^R$ *from* N*-dimensional space onto* R*-dimensional space* $(R < N)$ *is defined as*

$$\mathcal{G}_N^R = \left\{ \mathbf{U}\mathbf{U}^H | \mathbf{U} \in \mathcal{U}_N^R \right\} \tag{3.6}$$

Remark 3.2.2 *Let* $\mathbf{U} \in \mathcal{U}_N^R$ *be an orthonormal basis that spans a linear space of dimension* R*. Notice that the existence of* $\mathbf{U}$ *is not unique as shown in remark 3.2.1. Nevertheless, its corresponding projection matrix* $\mathbf{\Pi} = \mathbf{U}\mathbf{U}^H$ *is unique, i.e.,* $\mathbf{U}\mathbf{U}^H = \mathbf{U}\mathbf{Q}\mathbf{Q}^H\mathbf{U}^H$*, since* $\mathbf{Q}\mathbf{Q}^H = \mathbf{I}_R$*,* $\forall \mathbf{Q} \in \mathcal{U}_R^R$*.*

Consider $\mathbf{U} \in \mathcal{U}_N^N$ that can be splited as $\mathbf{U} = [\mathbf{U}_R | \mathbf{U}_R^\perp]$, where $\mathbf{U}_R \in \mathcal{U}_N^R$ and $\mathbf{U}_R^\perp \in \mathcal{U}_N^{N-R}$ is the orthogonal complement of the orthonormal basis $\mathbf{U}$. Denote $\mathbf{\Pi} = \mathbf{U}_R\mathbf{U}_R^H \in \mathcal{G}_N^R$ and its orthogonal complement $\mathbf{\Pi}^\perp = \mathbf{U}_R^\perp \mathbf{U}_R^{\perp H} \in \mathcal{G}_N^{N-R}$. We have the following properties:

$$\begin{aligned} \mathbf{\Pi}^\perp &= \mathbf{I}_N - \mathbf{\Pi} \\ \mathbf{\Pi}\mathbf{U}_R &= \mathbf{U}_R \\ \mathbf{\Pi}\mathbf{\Pi} &= \mathbf{\Pi} \\ \mathbf{\Pi}\mathbf{U}_R^\perp &= \mathbf{0} \\ \mathbf{\Pi}\mathbf{\Pi}^\perp &= \mathbf{0} \end{aligned} \tag{3.7}$$

3.3 Structured covariance models

Let us start with the definitions of the following sets:

Definition 3.3.1 ***The set of*** N ***dimensional Hermitian matrices of rank*** R
$\mathcal{H}_N^R$ *denotes the set of Hermitian semi definite matrix of size* N *and rank* R*, i.e.,*

$$\mathcal{H}_N^R = \left\{ \mathbf{\Sigma}_R \in \mathcal{H}_N^+ | rank(\mathbf{\Sigma}_R) = R \right\} \tag{3.8}$$

Thus, each $\mathbf{\Sigma}_R \in \mathcal{H}_N^R$ *can be represented by its low rank SVD as*

$$\begin{aligned}\mathbf{\Sigma}_R &\overset{\text{SVD}}{=} [\mathbf{U}_R|\mathbf{U}_R^\perp] \begin{bmatrix}\mathbf{\Lambda}_R & 0\\ 0 & 0\end{bmatrix} [\mathbf{U}_R|\mathbf{U}_R^\perp]^H \\ &= \mathbf{U}_R\mathbf{\Lambda}_R\mathbf{U}_R^H \end{aligned} \tag{3.9}$$

where $\mathbf{U}_R \in \mathcal{U}_N^R$ *and* $\mathbf{\Lambda} \in \mathbb{R}^{R\times R}$ *is a diagonal matrix with positive entries.*

Definition 3.3.2 ***Low rank plus identity structured matrices***
$\mathcal{H}_R^{LR}$ *denotes the set of low rank component plus scaled identity matrix given as*

$$\mathcal{H}_R^{LR} = \left\{\mathbf{\Sigma}_R + \sigma^2\mathbf{I}_N \in \mathbb{C}^{N\times N} | \mathbf{\Sigma}_R \in \mathcal{H}_N^R\right\} \tag{3.10}$$

where the quantities N, R *and* σ^2 *allow to define various sets (and are left implicit for the ease of notation).*

The low rank plus identity structured matrices is common for radar processing. Indeed, this structure appears naturally for lows dimensional signals embedded in thermal noise. For example, if we consider the model of CG plus WGN in Definition 1.4.1, the covariance matrix of the signal belongs to $\mathcal{H}_R^{LR}$ if the covariance of the CG component belongs to $\mathcal{H}_N^R$.

3.4 List of operators

We define, in this section, some operators which will be needed in our derivations:

Definition 3.4.1 $\mathcal{P}_R\{.\}$
The operator $\mathcal{P}_R\{.\}$ *extracts the first* R *eigenvectors from a given matrix in* $\mathcal{H}_N^+$*, is defined by*

$$\begin{array}{rlcl}\mathcal{P}_R: & \mathcal{H}_N^+ & \longrightarrow & \mathcal{U}_N^R \\ & \mathbf{M} \overset{\text{SVD}}{=} [\mathbf{U}_R|\mathbf{U}_R^\perp]\ \mathbf{D}\ [\mathbf{U}_R|\mathbf{U}_R^\perp]^H & \longmapsto & \mathcal{P}_R\{\mathbf{M}\} = \mathbf{U}_R.\end{array} \tag{3.11}$$

Definition 3.4.2 $\mathcal{R}_R\{.\}$
$\mathcal{R}_R\{.\}$ *is the range space spanned by the* R *principal eigenvectors of a given Hermitian semi definite matrix defined as*

$$\begin{array}{rlcl}\mathcal{R}_\text{R}: & \mathcal{H}_N^+ & \longrightarrow & \mathcal{G}_N^R \\ & \mathbf{\Sigma} \overset{\text{SVD}}{=} [\mathbf{U}_R|\mathbf{U}_R^\perp]\mathbf{D}[\mathbf{U}_R|\mathbf{U}_R^\perp]^H & \longmapsto & \mathbf{U}_R\mathbf{U}_R^H\end{array} \tag{3.12}$$

Definition 3.4.3 $\mathcal{P}_\text{Proc}\{.\}$
The operator $\mathcal{P}_\text{Proc}$ *is the projection onto the complex Stiefel manifold see (cf. definition 2.0.2), defined as*

$$\begin{array}{rlcl}\mathcal{P}_\text{Proc}: & \mathbb{C}^{N\times R} & \longrightarrow & \mathcal{U}_N^R \\ & \mathbf{Z} \overset{\text{TSVD}}{=} \mathbf{U}\mathbf{D}\mathbf{V}^H & \longmapsto & \mathcal{P}_\text{Proc}\{\mathbf{Z}\} = \mathbf{U}\mathbf{V}^H\end{array} \tag{3.13}$$

where $\overset{\text{TSVD}}{=}$ *defines the thin-singular value decomposition of a given matrix, i.e., the SVD that only keeps non-zero eigenvalues.*

Definition 3.4.4 $\mathcal{T}_R\{.\}$
This operator associates to any Hermitian matrix $\mathbf{\Sigma} \in \mathcal{H}_N^+$ *its projection on* $\mathcal{H}_R^{LR}$ *where*

$$\begin{array}{rlcl}\mathcal{T}_R: & \mathcal{H}_N^+ & \longrightarrow & \mathcal{H}_N^+ \\ & \mathbf{\Sigma} \overset{\text{SVD}}{=} [\mathbf{U}|\mathbf{U}^\perp] \begin{bmatrix}\mathbf{\Lambda}_R & 0\\ 0 & \mathbf{\Lambda}_{N-R}\end{bmatrix} [\mathbf{U}|\mathbf{U}^\perp]^H & \longmapsto & \mathcal{T}_R\{\mathbf{\Sigma}\} \overset{\text{SVD}}{=} [\mathbf{U}|\mathbf{U}^\perp] \begin{bmatrix}\tilde{\mathbf{\Lambda}}_R & 0\\ 0 & \psi\mathbf{I}_{N-R}\end{bmatrix} [\mathbf{U}|\mathbf{U}^\perp]^H\end{array} \tag{3.14}$$

where $\psi \in \mathbb{R}_+^*$ *is assumed known and*

$$[\tilde{\mathbf{\Lambda}}]_{i,i} = \begin{cases} \max([\mathbf{\Lambda}]_{i,i}, \psi), & i \leq R \\ \psi, & i > R \end{cases} \tag{3.15}$$

Chapter 4

Conclusion

In this first part, we presented the definitions of the distributions and matrix structures that are used in this thesis. Chapter 1 presented several distributions for vectors observations including Gaussian, CG and and their mixture. Chapter 2 introduced several distributions on the Stiefel manifold (set of semi-unitary matrices). In Chapter 3, we focused on Hermitian matrices modeled as a sum of low rank component (where the signal of interest lies in) and we presented several matrix structures. The following of this thesis is divided into two parts that propose adaptive signal processing methods:

In Part II, first, we consider the problem of low dimensional signal subspace estimation in a Bayesian context. We focus on CG signals embedded in WGN, which is a realistic modeling for various array processing applications. We propose algorithms to compute various Bayesian subspace estimators in this context. In order to introduce the prior information on the subspace basis (required for deriving Bayesian estimators), we make use of the CGBL distribution. Such approaches have shown their interests for signal subspace estimation in the small sample support and/or low signal to noise ratio (SNR) contexts. Numerical simulations illustrate the performance of the proposed algorithms. Finally, some practical examples of Bayesian prior design are presented for the purpose of radar detection.

Part III deals with testing the similarity of covariance matrices from groups of observations in the context of low rank structured covariance matrix. Such approach has been shown to be a relevant for change and/or anomaly detection in synthetic aperture radar (SAR) images. While the term "similarity" usually refers to equality or proportionality, we explore the testing of shared properties in the structure of covariance matrices as specified in Chapter 3 in Part I, i.e., low rank component plus scaled identity matrix. Specifically, we derive generalized likelihood ratio tests (GLRTs) to infer:

i) on the equality of the low rank signal component of covariance matrices,

ii) on the proportionality of the low rank signal component of covariance matrices,

iii) on the equality of the principal subspace component

The formulation of these tests involves non-trivial optimization problems for which we tailor efficient maximization minimization (MM) algorithms. Eventually, the proposed detection methods enjoy interesting properties, that are illustrated on simulations and on an application to real data for change detection of SAR images.

Part II

Bayesian subspace estimation

Chapter 1

Subspace estimation

Subspace estimation is an ubiquitous problem in signal processing, as it is often required to infer the low-dimensional space where information lies in. It is considered as the cornerstone of a plethora of applications and algorithms such as PCA [2], DOA estimation [1, 46, 47] and for adaptive filtering/detection and interference cancellation [3, 28, 48–54]. The performances of the aforementioned applications depend on the accuracy of subspace estimation. Nevertheless, the latter becomes a challenging problem in the presence of non-standard conditions such as low sample support, low SNR, non-Gaussian observations or presence of outliers in the training set, which motivated the present work.

Along this Part, N denotes the size of the data, K represents the number of samples and R is the rank of the signal subspace ($R < N$). We denote by $\{\mathbf{z}_k\}$ (columns of $\mathbf{Z} \in \mathbb{C}^{N\times K}$) the sample vectors, $\mathbf{U} \in \mathcal{U}_N^R$ an unknown orthonormal basis of the signal subspace, $\mathbf{S} \in \mathbb{C}^{R\times K}$ the matrix containing the signal of interest and $\mathbf{N} \in \mathbb{C}^{N\times K}$ the additive noise.

1.1 The subspace estimation problem

1.1.1 General formulation and motivation

In signal processing application, the information in samples $\{\mathbf{z}_k\}$ often lies in low dimensional structures, i.e., $\{\mathbf{z}_k\}$ can be well represented by a low dimensional signal embedded in a high dimensional ambient space. For this model, $\mathbf{z}_k \in \mathbb{C}^N$ are drawn as:

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

where $\mathbf{s}_k$ denotes the signal of interest, which lies in low rank subspace of dimension $R \ll N$, spanned by the semi unitary matrix $\mathbf{U} \in \mathcal{U}_N^R$ and $\mathbf{n}_k$ is an additive noise.

Our main interest is to estimate the subspace from the data set $\{\mathbf{z}_k\}$. Several approaches exist to respond to this problem. From a geometric perspective, subspace estimators can generally be seen as minimizers of a distance d, defined between the samples and $\mathbf{U}$, i.e.,

$$\widehat{\mathbf{U}} = \arg\min_{\mathbf{U}} \sum_{k=1}^{K} d(\mathbf{U}, \mathbf{z}_k) \tag{1.2}$$

1.1.2 Notes on rank estimation

The problem of determining the dimension of the subspace R from the observations will be referred to as rank estimation. It can be casted into a classical model order selection problem

that arises in a variety of important statistical signal and array processing systems. In [55,56], the problem of optimal rank estimation is investigated in the context of the standard array observation model with additive WGN. The latter studies demonstrated the existence of a phase transition threshold, below which eigenvalues and associated eigenvectors of the sample covariance matrix (SCM) fail to provide any information on population eigenvalues. As such, the majority of algorithms for rank estimation are based on analysis of the eigenvalues of the SCM. For instance, before performing a given processing (as for the multiple signal classification (MUSIC) algorithm [1]), the number of sources is assumed known. To do so, methods to detect the number of sources have also been developed both in the statistics and signal processing communities [57].

In other cases, the determination of R may be directly obtained by integrating physical prior knowledge on this parameter [58] (e.g., this is the case for space time adaptive processing (STAP)). Indeed, the disturbance in STAP [59] is generally known to be composed of a low rank clutter plus an additive WGN where the rank can be evaluated from [60].

In this thesis, we will always assume that the rank R is known or pre-established (e.g. by one of the cited references).

1.2 State of the art of subspace estimators

1.2.1 Probabilistic PCA and sample covariance matrix (SCM)

Most commonly, the signal subspace is estimated through the strongest eigenvectors of the SCM. This corresponds to the maximum likelihood estimator (MLE) for the classical model with additive WGN [61]. The SCM estimate is given as

$$\widehat{\mathbf{U}}_{\text{SCM}} = \mathcal{P}_R \left\{ \frac{1}{K} \sum_{k=1}^{K} \mathbf{z}_k \mathbf{z}_k^H \right\} \tag{1.3}$$

where the definition of the operator $\mathcal{P}_R\{.\}$ is given in Definition 3.4.1 in Part I.

The SCM provides an accurate estimator for high SNR and/or for large number of samples. Nevertheless, it shows its limits outside these asymptotic regimes. The latter is also known to be sensitive to miss-modeling, e.g., presence of outliers or non-Gaussian observations.

1.2.2 Fixed point estimator (FPE)

To overcome the robustness issue, a possible alternative is to consider the CG modeling [16]. This approach leads to new subspace estimator, built from the SVD of a robust covariance matrix estimator [18,26]. For example, one can use Tyler's estimator, defined by

$$\widehat{\mathbf{\Sigma}}_{\text{FPE}} = \frac{N}{K} \sum_{k=1}^{K} \frac{\mathbf{z}_k \mathbf{z}_k^H}{\mathbf{z}_k^H \widehat{\mathbf{\Sigma}}_{\text{FPE}}^{-1} \mathbf{z}_k} \tag{1.4}$$

This estimator is an approached MLE of the covariance matrix for full rank CG component, i.e., $\mathbf{z} \sim \mathcal{CG}(0, \mathbf{\Sigma}, \tau)$, its corresponding subspace estimator is

$$\widehat{\mathbf{U}}_{\text{FPE}} = \mathcal{P}_R\{\widehat{\mathbf{\Sigma}}_{\text{FPE}}\} \tag{1.5}$$

The latter is popular thanks to its robustness and it is employed in several array processing applications such as radar detection (see [12] and references therein). Nevertheless, this estimation process requires a number of samples $K \geq N$. For under-sampled configurations, i.e.,

$K < N$, [62–64] proposed to regularize this covariance matrix estimator by a diagonal-loading technique. This estimator is referred to as shrinkage fixed point estimator (SFPE) reads,

$$\widehat{\mathbf{\Sigma}}_{\text{SFPE}}(\beta) = (1-\beta)\frac{N}{K}\sum_{k=1}^{K}\frac{\mathbf{z}_k\mathbf{z}_k^H}{\mathbf{z}_k^H\widehat{\mathbf{\Sigma}}_{\text{SFPE}}^{-1}\mathbf{z}_k} + \beta\mathbf{I}_N \tag{1.6}$$

where $\beta \in [\max(0, 1-K/N), 1]$ and its corresponding subspace estimator is

$$\widehat{\mathbf{U}}_{\text{SFPE}} = \mathcal{P}_R\{\widehat{\mathbf{\Sigma}}_{\text{SFPE}}\} \tag{1.7}$$

We note that for small number of samples ($K \ll N$), existence conditions [65] impose the introduced bias to be inherently significant.

Remark 1.2.1 *The Tyler estimator* (1.5) *may not be the most appropriate for the considered model of low-rank CG component plus WGN for two reasons: First, its corresponding model does not take into account the low-rank structure of the CG component. Therefore, it may lead to a loss of performance. Secondly, this estimator requires $K > N$ secondary data to be computed, which is a problem for high-dimensional data. Moreover, this requirement does not allow to take full advantage of the low-rank assumption in the cases where $2R \ll N$.*

1.2.3 MLE for CG sources embedded in WGN

In [19,45], the MLE for the signal subspace is derived for the CG plus WGN of definition 1.4.1 of Part I, i.e., $\mathbf{s}_k \sim \mathcal{CG}(0, \mathbf{\Sigma}, \tau_k)$ and $\mathbf{n}_k \sim \mathcal{CN}(0, \sigma^2)$. In these works, the CG component is assumed to lie in a low rank subspace spanned by $\mathbf{U} \in \mathcal{U}_N^R$ (see Section 1.1.1), so its scatter matrix $\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$ where $\mathbf{\Lambda} \in \mathbb{R}^{R\times R}$ is a diagonal matrix. Algorithms to compute this MLE are proposed in [19,45].

1.2.4 MLE for CG sources embedded in heterogeneous noise

In [30], a MLE based subspace estimator is designed for data modeled as a CG distributed signal of interest embedded in heterogeneous noise (CG outliers plus WGN) where each sample vector $\mathbf{z}_k$ is modeled as

$$\mathbf{z}_k = \mathbf{s}_k + \mathbf{n}_k + \mathbf{o}_k \tag{1.8}$$

with $\mathbf{s}_k + \mathbf{n}_k$ follows the same mixture model as described in the previous Section 1.2.3 for which $\mathbf{\Sigma} = \mathbf{U}\mathbf{U}^H$. However, [30] considered the potential presence of outlier $\mathbf{o}_k$, modeled as $\mathbf{o}_k \sim \mathcal{CG}(0, \mathbf{U}^\perp\mathbf{U}^{\perp H}, \beta_k)$ where $\mathbf{U}^\perp \in \mathcal{U}_N^{N-R}$ denotes the orthogonal complement of the orthonormal basis $\mathbf{U}$. This modeling allows to develop robust estimation procedure to various noise distributions and to outliers and its MLE reads

$$\widehat{\mathbf{U}}_{\text{MLE}} = \mathcal{P}_R\left\{\sum_{k=1}^{K}\phi(\widehat{\mathbf{U}}_{\text{MLE}}, \mathbf{z}_k)\mathbf{z}_k\mathbf{z}_k^H\right\} \tag{1.9}$$

where the function ϕ is defined in [30].

1.2.5 Bayesian subspace estimation methods: maximum a posteriori probability (MAP) and Minimum mean square distance (MMSD)

The aforementioned estimators were not derived in a Bayesian setting, in the sense that they did not assume a prior on the subspace of interest. In this context, a prior distribution of the

subspace orthonormal basis can be assumed to integrate a prior knowledge on the basis of interest $\mathbf{U}$ denoted by $p(\mathbf{U})$. Indeed, assuming p.d.f. $p(\mathbf{Z}|\mathbf{U},\boldsymbol{\theta})$ and $p(\mathbf{U})$ yields various estimators such as the MAP or the MMSD (where $\boldsymbol{\theta}$ denotes side/nuisance parameters in the model of $\mathbf{Z}$).

The MAP estimation has been adopted as a standard approach for solving many high dimensional inverse problems where a prior knowledge is available. This estimator has been used in subspace estimation context in e.g. [66] and [67], where the MAP is proposed assuming Bingham/Langevin as a prior distributions on $\mathbf{U}$. Its generic definition is the following:

Definition 1.2.1 ***Maximum a posteriori probability (MAP) for*** $\mathbf{U}$
Assuming a parametric model a distribution $p(\mathbf{Z}|\mathbf{U},\boldsymbol{\theta})$ *and a prior distribution* $\mathbf{U} \sim p(\mathbf{U})$*, the MAP is defined as*

$$\widehat{\mathbf{U}}_{\text{MAP}} = \arg\max_{\mathbf{U}} \; p\left(\mathbf{U}|\mathbf{Z},\boldsymbol{\theta}\right) = \arg\max_{\mathbf{U}} \; p\left(\mathbf{Z}|\mathbf{U},\boldsymbol{\theta}\right) p(\mathbf{U}) \tag{1.10}$$

The MMSD minimizes the expected distance between the true projection matrix and its estimate [66,68]. The latter is an intuitively appealing method, as it is based on a natural metric in the complex Grassmann manifold [36]. In the context of subspace estimation, the MMSD has been introduced in [66,68]. More specifically, [66] derives a practical formulation of the MMSD estimators when the subspace of interest is parameterized by its orthonormal basis.

Definition 1.2.2 ***Minimum mean square distance (MMSD)***
Assuming a parametric model a distribution $p(\mathbf{Z}|\mathbf{U},\boldsymbol{\theta})$ *and a prior distribution* $\mathbf{U} \sim p(\mathbf{U})$*, the MMSD is defined as*

$$\begin{aligned}
\widehat{\mathbf{U}}_{MMSD} &= \arg\min_{\widehat{\mathbf{U}}} \; \mathbb{E}_{\mathbf{U},\mathbf{Z}} \left\{ \| \, \widehat{\mathbf{U}}\widehat{\mathbf{U}}^H - \mathbf{U}\mathbf{U}^H \, \|_F^2 \right\} \\
&= \arg\max_{\widehat{\mathbf{U}}} \; \mathbb{E}_{\mathbf{U},\mathbf{Z}} \{ Tr\{ \widehat{\mathbf{U}}^H \mathbf{U}\mathbf{U}^H \widehat{\mathbf{U}} \} \} \\
&= \arg\max_{\widehat{\mathbf{U}}} \; \int \left[\int Tr\{ \widehat{\mathbf{U}}^H \mathbf{U}\mathbf{U}^H \widehat{\mathbf{U}} \} p(\mathbf{U}|\mathbf{Z}) d\mathbf{U} \right] p(\mathbf{Z}) d\mathbf{Z} \\
&= \arg\max_{\widehat{\mathbf{U}}} \; \int Tr\{ \widehat{\mathbf{U}}^H \mathbf{U}\mathbf{U}^H \widehat{\mathbf{U}} \} p(\mathbf{U}|\mathbf{Z}) d\mathbf{U} \\
&= \arg\max_{\widehat{\mathbf{U}}} \; Tr \left\{ \widehat{\mathbf{U}}^H \left[\int \mathbf{U}\mathbf{U}^H p(\mathbf{U}|\mathbf{Z}) d\mathbf{U} \right] \widehat{\mathbf{U}} \right\}
\end{aligned} \tag{1.11}$$

which can be obtained as

$$\widehat{\mathbf{U}}_{MMSD} = \mathcal{P}_R \left\{ \int \mathbf{U}\mathbf{U}^H p(\mathbf{U}|\mathbf{Z}) d\mathbf{U} \right\} = \mathcal{P}_R \left\{ \mathbf{M}(p(\mathbf{U}|\mathbf{Z})) \right\} \tag{1.12}$$

in which

$$\mathbf{M}(p(\mathbf{U}|\mathbf{Z})) = \int \mathbf{U}\mathbf{U}^H p(\mathbf{U}|\mathbf{Z}) d\mathbf{U} \tag{1.13}$$

Remark 1.2.2 *The expression of the MMSD depends on* $p(\mathbf{U}|\mathbf{Z})$*, which is specified based on both the data model and the prior distribution assigned to the parameters. Usually, there is no closed form solutions to compute* $\mathbf{M}(p(\mathbf{U}|\mathbf{Z}))$*. However,* (1.12) *can still be evaluated using the so-called induced arithmetic mean (IAM) [66] of the semi-unitary matrix, as*

$$\widehat{\mathbf{U}} \approx \mathcal{P}_R \left\{ \frac{1}{N_r} \sum_{n=N_{bi}+1}^{N_{bi}+N_r} \mathbf{U}_{(n)} {\mathbf{U}_{(n)}}^H \right\} \tag{1.14}$$

where $\mathbf{U}_{(n)}$ *are sampled from* $p(\mathbf{U}|\mathbf{Z})$*,* N_{bi} *stands for the burn-in samples (number of thrown samples from the Markov chain), and* N_r *is the number of samples used to evaluate the integral.*

Example 1.2.1 ***MMSD estimator***
We present an example of the MMSD estimator for linear model described in [66] involving the uniform prior for the sources distribution with real entries, i.e., $\mathbf{z}_k \sim \mathcal{N}(0, \tilde{\mathbf{U}}\tilde{\mathbf{U}}^T + \sigma^2\mathbf{I}_N)$, $\tilde{\mathbf{U}} \in \mathcal{U}_N^R(\mathbb{R})$ *and real Bingham prior for the basis of interest, i.e.,* $\tilde{\mathbf{U}} \sim B(\mathbf{A})$. *This MMSD has the following closed form expression*

$$\widehat{\mathbf{U}}_{\text{MMSD-LM-B}} = \mathcal{P}_R\left\{\mathbf{A} + \frac{1}{2\sigma^2}\mathbf{Z}\mathbf{Z}^T\right\} \tag{1.15}$$

where σ^2 *is the variance of an additive WGN.*

1.3 List of following contributions

- In Chapter 2, we consider the low rank CG sources embedded in WGN. The novelty in our statistical model, is to incorporate a prior knowledge on $\mathbf{U}$, i.e., $\mathbf{U} \sim \text{CGBL}(\mathbf{C}, \{\mathbf{A}_r\})$. We develop an algorithm to compute the MAP estimator for the proposed model based on the MM algorithm [69]. Then, we derive a Gibbs-sampler based algorithm to compute the MMSD estimator, which follows the framework of [66].

- In Chapter 3, we consider the same model as Chapter 2 with an added outlier component. This modeling has been successfully used in [30] in order to develop robust estimators in a non Bayesian context. We extend this work to the Bayesian framework and propose the corresponding MAP and MMSD estimators.

- In Chapter 4, numerical simulations show that the inclusion of a Bayesian prior on the subspace orthonormal basis can significantly improve the performance of the estimation process. The design of this prior depends, of course, on the considered application and comes from appropriate physical considerations/models. In this scope, we also illustrate the practical use of the proposed methods for STAP detection. The proposed application directly leverages the physical model of [70] in order to improve the performance of low rank detectors [32, 33] at low sample support.

Chapter 2

Bayesian subspace estimators

In this chapter, we focus on the context of low rank CG sources embedded in white Gaussian noise described in Definition 1.4.1 of Part I. This choice is motivated by the fact that the CG distribution has been considered in many modern robust signal processing applications. Hence, the considered model can accurately describe clutter (or power-fluctuating sources) plus thermal noise observations, which are common in plethora of signal processing application [19, 69, 71]. However, these studies were never brought to a Bayesian context, in the sense that they did not incorporate a prior knowledge into the estimation process of signal subspace. We fill this gap by deriving new Bayesian estimators in the context of CG distributed sources embedded in WGN where we assign to the basis of interest $\mathbf{U}$ the CGBL distribution.

2.1 Data model and problem statement

We rely on the data model in (1.4.1) namely, the mixture model. This formulation is useful to model clutter (or power-fluctuating sources) plus thermal noise in several array processing applications, such as radar [19,32,33,69,71]. For this model, the samples $\{\mathbf{z}_k\}_{k=1}^K$ (the columns of $\mathbf{Z}$) are drawn as

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

where

• $\mathbf{s}_k \sim \mathcal{CG}(\mathbf{0}, \mathbf{\Sigma}, \tau_k)$ are the low rank CG distributed sources. Moreover, the source scatter matrix is parameterized by its low-rank SVD as

$$\mathbf{\Sigma} \overset{\text{SVD}}{=} [\mathbf{U}|\mathbf{U}^{\perp}] \begin{bmatrix} \mathbf{\Lambda} & 0 \\ 0 & 0 \end{bmatrix} [\mathbf{U}|\mathbf{U}^{\perp}]^H \tag{2.2}$$

In addition,

i) $\{\tau_k\}$ are the CG textures assumed to be positive unknown deterministic.

ii) $\mathbf{\Lambda} = \text{diag}(\{\lambda_r\})$ is the matrix containing the scatter matrix eigenvalues, which are assumed to be positive unknown deterministic.

iii) $\mathbf{U} \in \mathcal{U}_N^R$ are the eigenvectors of the scatter matrix, whose columns spans the signal subspace basis. The latter follows the distribution $\mathbf{U} \sim \text{CGBL}(\mathbf{C}, \{\mathbf{A}_r\})$.

• $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_N)$ is an additive WGN of known or pre-estimated variance σ^2.

The data matrix can be therefore written as

$$\mathbf{Z} = \mathbf{U}\bar{\mathbf{S}}\mathbf{T} + \mathbf{N} \tag{2.3}$$

with the columns of $\bar{\mathbf{S}} \in \mathbb{C}^{R\times K}$ distributed as $\bar{\mathbf{s}}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Lambda})$ and $\mathbf{T} = \text{diag}(\{\sqrt{\tau_k}\})$. The latter reads as a modified linear model, with unknown power fluctuations for each sample gathered in the matrix $\mathbf{T}$.

Remark 2.1.1 *In this section, we consider the hybrid Bayesian model because our main interest is incorporating a prior knowledge on the signal subspace in the estimation process. Conversely, we choose not to specify the p.d.f. of the texture parameters $\{\tau_k\}$ (and the eigenvalues $\{\lambda_r\}$) which are assumed unknown and deterministic. By doing so, we ensure more robustness to any prior mismatch w.r.t. these parameters. Moreover, this assumption also allows for computational tractability since including a prior distribution on $\{\tau_k\}$ in the considered model leads to integral functions that are complex to handle [72]. In the following, for sake of conciseness and with an abuse of language, the ML-MMSD hybrid Bayesian estimator (respectively ML-MAP) will be simply referred to as MMSD (respectively MAP).*

By denoting

$$\mathbf{\Sigma}_k = \tau_k \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H + \sigma^2 \mathbf{I}_N \quad \forall k \in [\![1, K]\!] \tag{2.4}$$

we have for each sample the conditional representation, $\mathbf{z}_k|\mathbf{U}, \mathbf{\Lambda}, \tau_k \sim \mathcal{CN}(0, \mathbf{\Sigma}_k)$, leading to the conditional p.d.f. of the sample set $\mathbf{Z}$ as

$$p(\mathbf{Z}|\mathbf{U}, \{\lambda_r\}, \{\tau_k\}) = \prod_{k=1}^{K} p(\mathbf{z}_k|\mathbf{U}, \{\lambda_r\}, \tau_k) \propto \prod_{k=1}^{K} \frac{\exp\left\{-\mathbf{z}_k^H \mathbf{\Sigma}_k^{-1} \mathbf{z}_k\right\}}{\det(\mathbf{\Sigma}_k)} \tag{2.5}$$

where $\det(\mathbf{\Sigma}_k) = \sigma^{2(N-R)} \prod_{r=1}^{R} (\tau_k \lambda_r + \sigma^2)$.

Thanks to the Sherman Morrison Woodbury lemma [73,74], the expression of $\mathbf{\Sigma}_k^{-1}$ is simplified as

$$\mathbf{\Sigma}_k^{-1} = \sigma^{-2}\mathbf{I} - \mathbf{U}\mathbf{\Gamma}_k\mathbf{U}^H \tag{2.6}$$

where $\mathbf{\Gamma}_k = \sigma^{-2}\mathbf{I}_R - (\tau_k\mathbf{\Lambda} + \sigma^2\mathbf{I}_R)^{-1}$ is a diagonal matrix of entries

$$[\mathbf{\Gamma}_k]_{r,r} = \gamma_{k,r} = \frac{\tau_k \lambda_r}{\sigma^2(\tau_k\lambda_r + \sigma^2)} \tag{2.7}$$

From (2.6) and (2.5), some manipulations allow to the posterior probability of $\mathbf{U}$ to be rewritten as

$$\begin{aligned}
p(\mathbf{U}|\mathbf{Z}, \{\tau_k\}, \{\lambda_r\}) &\propto p(\mathbf{Z}|\mathbf{U}, \{\tau_k\}, \{\lambda_r\}) p_{\text{CGBL}}(\mathbf{U}) \propto \prod_{k=1}^{K} \frac{\exp\left\{-\mathbf{z}_k^H \left(\mathbf{\Sigma}_k\right)^{-1} \mathbf{z}_k\right\}}{\det(\mathbf{\Sigma}_k)} p_{\text{CGBL}}(\mathbf{U}) \\
&\propto \prod_{k=1}^{K} \left(\prod_{r=1}^{R} \frac{1}{\tau_k\lambda_r + \sigma^2} \right) \exp\left\{-\mathbf{z}_k^H(-\mathbf{U}\mathbf{\Gamma}_k\mathbf{U}^H + \sigma^{-2}\mathbf{I}_N)\mathbf{z}_k\right\} p_{\text{CGBL}}(\mathbf{U}) \\
&\propto \prod_{k=1}^{K} \left(\prod_{r=1}^{R} \frac{1}{\tau_k\lambda_r + \sigma^2} \right) \exp\left\{ \sum_{k=1}^{K} \mathbf{z}_k^H \mathbf{U}\mathbf{\Gamma}_k\mathbf{U}^H\mathbf{z}_k \right\} p_{\text{CGBL}}(\mathbf{U}) \\
&\propto \left(\prod_{k=1}^{K} \prod_{r=1}^{R} \frac{1}{\tau_k\lambda_r + \sigma^2} \right) \exp\left\{ \sum_{r=1}^{R} \mathbf{u}_r^H \mathbf{M}_r \mathbf{u}_r \right\} p_{\text{CGBL}}(\mathbf{U}) \\
&\propto \left(\prod_{k=1}^{K} \prod_{r=1}^{R} \frac{1}{\tau_k\lambda_r + \sigma^2} \right) \exp\left\{ \sum_{r=1}^{R} \text{Re}\{\mathbf{c}_r^H \mathbf{u}_r\} + \mathbf{u}_r^H \left[\mathbf{A}_r + \mathbf{M}_r\right] \mathbf{u}_r \right\}
\end{aligned} \tag{2.8}$$

with

$$\mathbf{M}_r = \sum_{k=1}^{K} \gamma_{k,r} \mathbf{z}_k \mathbf{z}_k^H \tag{2.9}$$

where $\gamma_{k,r}$ is given in (2.7).

In this chapter, we aim to develop Bayesian estimators of the subspace orthonormal basis $\mathbf{U}$ according to the data model (2.1). The first proposal is a MM algorithm to compute the MAP estimator. The second is an algorithm to evaluate the MMSD through MM iterations and a Gibbs sampling scheme. Additionally, we present a special case, referred to as "simplified model" where we consider all the non null eigenvalues of the scatter matrix are all equal to 1 and $\mathbf{U} \sim \text{CIB}(\kappa, \mathbf{A})$. We note that for this special case the MAP and the MMSD estimators coincide, and can be obtained through closed form updates. Considering these approaches, the properties of each method are listed below:

- Theoretically, the MMSD approach offers best performance in terms of expected distance between the estimated and true signal subspace projection matrices. Nevertheless, the computation of the MMSD estimator usually requires a Gibbs sampler scheme which can be computationally expensive.
- The MAP is theoretically sub-optimal (compared to the MMSD), but can generally reach good performance in practice. Moreover, the proposed algorithm to compute this estimator only involves closed form updates, which significantly reduces the computation time.
- The MMSD for the simplified model is interesting because it does not require a Gibbs sampling scheme to be computed. The assumed simplification is not necessarily realistic and introduces a mismatch w.r.t. the true model, however, numerical simulations will illustrate the interest of the approach.

2.2 Maximum a posterior probability estimator

In this section, we derive a subspace MAP estimator for the data model (2.1) that maximizes the posterior probability. The latter reads as the solution of

$$\begin{aligned} \underset{\widehat{\mathbf{U}},\{\tau_k\},\{\lambda_r\}}{\text{maximize}} \quad & p(\widehat{\mathbf{U}}|\mathbf{Z},\{\tau_k\},\{\lambda_r\}) \\ \text{subject to} \quad & \tau_k \geq 0 \ \forall k,\ \lambda_r \geq 0 \ \forall r \\ & \widehat{\mathbf{U}}^H\widehat{\mathbf{U}} = \mathbf{I}_R \end{aligned} \tag{2.10}$$

From (2.8) and (2.10), this problem can be recasted as

$$\begin{aligned} \underset{\{\widehat{\mathbf{u}}_r\},\{\tau_k\},\{\lambda_r\}}{\text{maximize}} \quad & \sum_{r=1}^{R} \text{Re}\{\mathbf{c}_r^H \widehat{\mathbf{u}}_r\} + \widehat{\mathbf{u}}_r^H \left[\mathbf{A}_r + \mathbf{M}_r\right] \widehat{\mathbf{u}}_r - \sum_{k=1}^{K} \ln(\tau_k \lambda_r + \sigma^2) \\ \text{subject to} \quad & \tau_k \geq 0 \ \forall k,\ \lambda_r \geq 0 \ \forall r \\ & \widehat{\mathbf{U}}^H\widehat{\mathbf{U}} = \mathbf{I}_R \ \text{ with } \ \widehat{\mathbf{U}} = [\widehat{\mathbf{u}}_1|\ldots|\widehat{\mathbf{u}}_R] \\ & \mathbf{M}_r = \sum_{k=1}^{K} \frac{\tau_k \lambda_r}{\sigma^2(\tau_k \lambda_r + \sigma^2)} \mathbf{z}_k \mathbf{z}_k^H \end{aligned} \tag{2.11}$$

To solve this problem, we derive an iterative based MM algorithm that sequentially updates the variables $\widehat{\mathbf{U}}^t, \{\tau_k^t\}, \{\lambda_r^t\}$ at the t-th iteration. Specifically, we make use of the block MM

algorithm [75] for this problem. This algorithm performs a block coordinate update of the parameters by minimizing the surrogate (majorizing) function of the objective. The interest of the majorization lies in the possibility to obtain closed form updates and to ensure a monotonic decrement of the objective value at each step. There are general no results on the convergence towards the global minimum, but good performance is observed in practice. Furthermore, the convergence of the objective function is guaranteed under certain mild conditions [76].

The resulting algorithm is summed up in the box Algorithm 1. A brief explanation of the derivations is given below.

2.2.1 Algorithm derivation

First, the variables $\widehat{\mathbf{U}}$, $\{\lambda_r\}$, and $\{\tau_k\}$ are initialized. This initialization can, for example, be taken from the R strongest eigenvectors and eigenvalues of the SCM for $\widehat{\mathbf{U}}$ and $\{\lambda_r\}$ and the norm of the samples for $\boldsymbol{\tau}$.

2.2.1.1 Update of the basis $\widehat{\mathbf{U}}$

By fixing $\{\lambda_r^t\}, \{\tau_k^t\}$, the update of the basis of interest $\widehat{\mathbf{U}}^{t+1}$ is obtained by solving

$$\begin{aligned} \underset{\{\widehat{\mathbf{u}}_r\}}{\text{maximize}} \quad & \sum_{r=1}^{R} \mathrm{Re}\{\mathbf{c}_r^H \widehat{\mathbf{u}}_r\} + \widehat{\mathbf{u}}_r^H \left(\mathbf{A}_r + \mathbf{M}_r^t\right) \widehat{\mathbf{u}}_r \\ \text{subject to} \quad & \widehat{\mathbf{U}}^H \widehat{\mathbf{U}} = \mathbf{I}_R \;\; \text{with} \;\; \widehat{\mathbf{U}} = [\widehat{\mathbf{u}}_1 | \ldots | \widehat{\mathbf{u}}_R] \end{aligned} \tag{2.12}$$

with $\mathbf{M}_r^t = \sum_{k=1}^{K} \gamma_{k,r}^t \mathbf{z}_k \mathbf{z}_k^H$ and $\gamma_{k,r}^t = \frac{\tau_k^t \lambda_r^t}{\sigma^2(\tau_k^t \lambda_r^t + \sigma^2)}$. This problem has not a trivial solution due to the semi-unitary constraint. Therefore, we apply the MM procedure in order to obtain closed form updates that improve the value of the objective at each iteration. An update of the orthonormal basis can be obtained thanks to Proposition B.1 in Appendix B. This update reads as

$$\widehat{\mathbf{U}}^{t+1} = \mathcal{P}_{\mathrm{Proc}}(\mathbf{H}^t) \tag{2.13}$$

where

$$\mathbf{H}^t = \mathbf{S}^t + 1/2\mathbf{C} \;\; \text{and} \;\; \mathbf{S}^t = [\, (\mathbf{A}_1 + \mathbf{M}_1^t)\mathbf{u}_1^t \mid \ldots \mid (\mathbf{A}_R + \mathbf{M}_R^t)\mathbf{u}_R^t \,] \tag{2.14}$$

with $\mathbf{u}_r^t$ is the r-th column of the matrix $\widehat{\mathbf{U}}^t$ and the operator $\mathcal{P}_{\mathrm{Proc}}$ is defined in Definition 3.4.3 in Part I.

2.2.1.2 Update of the texture parameter

The optimization problem in (2.11) w.r.t. $\{\tau_k\}$ for other fixed variables can be expressed as separable sub-problems in τ_k as

$$\begin{aligned} \underset{\tau_k}{\text{minimize}} \quad & \sum_{r=1}^{R} \ln\left(\tau_k \lambda_r^t + \sigma^2\right) - \frac{\tau_k \lambda_r^t}{\tau_k \lambda_r^t + \sigma^2} z_{k,r}^{t+1} \\ \text{subject to} \quad & \tau_k \geq 0 \end{aligned} \tag{2.15}$$

with $z_{k,r}^{t+1} = ||\mathbf{z}_k^H \mathbf{u}_r^{t+1}||^2$. This problem has no direct solution but a closed form update can be obtained thanks to (B.11) from Proposition B.2 in Appendix B for which we identify $\tau_k = a$,

$\lambda_r^t = b_p$, $R = P$, $s_p = z_{k,r}^{t+1}$ and $\alpha_{k,r}^t = \theta_p^t$. Consequently, the update reads

$$\tau_k^{t+1} = \frac{1}{R} \frac{\left(\sum_{r=1}^{R} z_{k,r}^{t+1} \frac{\tau_k^t \lambda_r^t}{\tau_k^t \lambda_r^t + \sigma^2} \right) \left(\sum_{r=1}^{R} \sigma^2 \frac{\alpha_{k,r}^t}{\tau_k^t \lambda_r^t + \sigma^2} \right)}{\sum_{r=1}^{R} \frac{\alpha_{k,r}^t \lambda_r^t}{\tau_k^t \lambda_r^t + \sigma^2}} \tag{2.16}$$

2.2.1.3 Update of the eigenvalues

By fixing the remaining variables, the optimization problem (2.11) w.r.t. $\{\lambda_r\}$ is equivalent to the optimization of the following sub-problems

$$\begin{aligned} \underset{\lambda_r}{\text{minimize}} \quad & \sum_{k=1}^{K} \ln\left(\tau_k^{t+1}\lambda_r + \sigma^2\right) - \frac{\tau_k^{t+1}\lambda_r}{\tau_k^{t+1}\lambda_r + \sigma^2} z_{k,r}^{t+1} \\ \text{subject to} \quad & \lambda_r \geq 0 \end{aligned} \tag{2.17}$$

Similarly to the update of texture and by using (B.11), we can apply Proposition B.2 in Appendix B with $\lambda_r = a$, $\tau_k^{t+1} = b_p$, $K = P$ and $\beta_{k,r}^t = \theta_p$, $z_{k,r}^{t+1} = s_p$. The updates of λ_r are then given by

$$\lambda_r^{t+1} = \frac{1}{K} \frac{\left(\sum_{k=1}^{K} z_{k,r}^{t+1} \frac{\tau_k^{t+1} \lambda_r^t}{\tau_k^{t+1} \lambda_r^t + \sigma^2} \right) \left(\sum_{k=1}^{K} \sigma^2 \frac{\beta_{k,r}^t}{\tau_k^{t+1} \lambda_r^t + \sigma^2} \right)}{\sum_{k=1}^{K} \frac{\beta_{k,r}^t \tau_k^{t+1}}{\tau_k^{t+1} \lambda_r^t + \sigma^2}}. \tag{2.18}$$

Algorithm 1: MAP for the general model (2.1) $\mathbf{U}_{\text{MAP}}$

input : $\mathbf{Z}$, $\mathbf{C}$, $\{\mathbf{A}_r\}$, σ^2, R, K, N
output : MAP estimators of $\mathbf{U}$, $\{\tau_k\}$, $\{\lambda_r\}$
initialize: $\widehat{\mathbf{U}}^0$, $\{\tau_k^0\}$ and $\{\lambda_r^0\}$
for $t = 0 \dots T-1$ **do**
  Update $\widehat{\mathbf{U}}^{t+1} = \mathcal{P}_{\text{Proc}}(\mathbf{H}^t)$, with $\mathbf{H}^t$ in (2.14)
  Update τ_k^{t+1} $\forall k$ with (2.16)
  Update λ_r^{t+1} $\forall r$ with (2.18)
end

2.3 Minimum mean square distance estimator

2.3.1 The subspace MMSD estimator for CG distributed sources

We recall that according to the data model in (2.1), $\mathbf{U} \sim \text{CGBL}(\mathbf{C}, \{\mathbf{A}_r\})$ and $\mathbf{z}_k \sim \mathcal{CN}(0, \tau_k \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H + \sigma^2 \mathbf{I}_N)$. Based on (1.11) and (2.4), the MMSD estimator of $\mathbf{U}$ is expressed as the solution of the following optimization problem

$$\begin{aligned} \underset{\widehat{\mathbf{U}}, \{\tau_k\}, \{\lambda_r\}}{\text{minimize}} \quad & \mathbb{E}_{\mathbf{U},\mathbf{Z}} \left\{ \| \widehat{\mathbf{U}}\widehat{\mathbf{U}}^H - \mathbf{U}\mathbf{U}^H \|_F^2 \right\} \\ \text{subject to} \quad & \tau_k \geq 0 \ \forall k, \ \lambda_r \geq 0 \ \forall r \\ & \widehat{\mathbf{U}}^H \widehat{\mathbf{U}} = \mathbf{I}_R \end{aligned} \tag{2.19}$$

In order to solve this optimization problem, we derive in the following an iterative algorithm that sequentially updates the variables $\widehat{\mathbf{U}}$, $\{\tau_k\}$ and $\{\lambda_r\}$. The update of $\widehat{\mathbf{U}}$ requires a Gibbs sampling scheme, while for updating both the texture $\{\tau_k\}$ and the eigenvalues $\{\lambda_r\}$, we use the MM procedure from Section 2.2. The overall algorithm is summed up in the box Algorithm 2.

2.3.2 Algorithm derivation

The initialization of the variables $\widehat{\mathbf{U}}^0$, $\{\lambda_r^0\}$ and $\{\tau_k^0\}$ is done as for the MAP estimator. The updates of the blocks $\widehat{\mathbf{U}}$, $\{\lambda_r\}$ and $\{\tau_k\}$ are detailed below:

2.3.2.1 Update of the basis $\widehat{\mathbf{U}}$

For fixed blocks $\{\tau_k^t\}$ and $\{\lambda_r^t\}$, the update $\widehat{\mathbf{U}}^{t+1}$ is obtained by solving the following problem

$$\begin{aligned} \underset{\widehat{\mathbf{U}}}{\text{minimize}} \quad & \mathbb{E}_{\mathbf{U},\mathbf{Z}} \left\{ \| \widehat{\mathbf{U}}\widehat{\mathbf{U}}^H - \mathbf{U}\mathbf{U}^H \|_F^2 \right\} \\ \text{subject to} \quad & \widehat{\mathbf{U}}^H\widehat{\mathbf{U}} = \mathbf{I}_R \end{aligned} \tag{2.20}$$

The latter corresponds to the MMSD estimation process. Thanks to the expression (1.12), the update is obtained by

$$\widehat{\mathbf{U}}^{t+1} = \mathcal{P}_R \left\{ \mathbf{M}(p(\mathbf{U}|\mathbf{Z}, \{\tau_k^t\}, \{\lambda_r^t\})) \right\} \tag{2.21}$$

with

$$\mathbf{M}(p(\mathbf{U}|\mathbf{Z}, \{\tau_k^t\}, \{\lambda_r^t\})) = \int \mathbf{U}\mathbf{U}^H p(\mathbf{U}|\mathbf{Z}, \{\tau_k^t\}, \{\lambda_r^t\}) d\mathbf{U} \tag{2.22}$$

The posterior probability in (2.8) is recognized as $(\mathbf{U}|\mathbf{Z}, \{\tau_k^t\}, \{\lambda_r^t\}) \sim \text{CGBL}(\mathbf{C}, \{\mathbf{G}_r^t\})$ with $\mathbf{G}_r^t = \mathbf{A}_r + \mathbf{M}_r^t$, $\forall r \in [\![1, R]\!]$. With this general distribution, there is no closed form for computing the integral in (2.22). Nevertheless, the update can be evaluated by the IAM as given in (1.14) where $\mathbf{U}_{(n)}^t$ are sampled as $\mathbf{U}_{(n)}^t \sim \text{CGBL}(\mathbf{C}, \{\mathbf{G}_r^t\})$. In order to do so, an efficient Gibbs sampling procedure to draw the CGBL distribution is given in Appendix A. Special cases for the sampling scheme required on this update are given in Table 2.1.

2.3.2.2 Update of the eigenvalues and the textures

For fixed $\widehat{\mathbf{U}}^{t+1}$, the update of the eigenvalues and the texture is equivalent to solve respectively (2.17) and (2.15). Consequently, the updates of $\{\tau_k^{t+1}\}$ and $\{\lambda_r^{t+1}\}$ are obtained respectively from (2.16) and (2.18).

2.4 Simplified model

In this Section, we focus on a special case that we refer to as simplified model, where $\mathbf{\Sigma} = \lambda\mathbf{U}\mathbf{U}^H$. The following considered relaxation from the true model, e.g. used in [72], allows for interesting simplifications that significantly reduce the computation time of the estimation procedure. Note that any scaling on the scatter can be absorbed in the textures parameters as $\tilde{\tau} = \lambda\tau$, so we can assume $\mathbf{\Sigma} = \mathbf{U}\mathbf{U}^H$. For this model, $\mathbf{z}_k|\mathbf{U}, \tau_k \sim \mathcal{CN}(0, \mathbf{\Sigma}_k)$, thus, the covariance reads

$$\mathbf{\Sigma}_k = \tau_k\mathbf{U}\mathbf{U}^H + \sigma^2\mathbf{I}_N \tag{2.23}$$

Table 2.1: Posterior distributions for standard priors on $\mathbf{U}$ under simplified (SM) and general (GM) models

The prior complex distribution of $\mathbf{U}$	Posterior distribution $p(\mathbf{U}\|\mathbf{Z})$ for a given model (SM/GM)	to be sampled for MMSD
Generalized Bingham CGB($\{\phi(r)\mathbf{A}\}$) $p(\mathbf{U}) \propto \exp\left\{\sum_{r=1}^{R} \phi(r)\mathbf{u}_r^H \mathbf{A}\mathbf{u}_r\right\}$	SM: $p(\mathbf{U}\|\mathbf{Z})_{\text{SM}} \propto \exp\left\{\sum_{p=1}^{P} \mathbf{u}_r^H(\phi(r)\mathbf{A}+\mathbf{W})\mathbf{u}_r\right\}$	CGB($\{\phi(r)\mathbf{A}+\mathbf{W}\}$)
	GM: $p(\mathbf{U}\|\mathbf{Z})_{\text{GM}} \propto \exp\left\{\sum_{p=1}^{P} \mathbf{u}_r^H(\phi(r)\mathbf{A}+\mathbf{M}_r)\mathbf{u}_r\right\}$	CGB($\{\phi(r)\mathbf{A}+\mathbf{M}_r\}$)
Langevin CL(κ, $\mathbf{C}$) $p(\mathbf{U}) \propto \text{etr}\left\{\kappa\mathbf{C}\mathbf{U}^H\right\}$	SM: $p(\mathbf{U}\|\mathbf{Z})_{\text{SM}} \propto \text{etr}\left\{\text{Re}\{\kappa\mathbf{C}^H\mathbf{U}\}+\mathbf{U}^H\mathbf{W}\mathbf{U}\right\}$	CBL($\kappa\mathbf{C}, \mathbf{I}_R, \mathbf{W}$)
	GM: $p(\mathbf{U}\|\mathbf{Z})_{\text{GM}} \propto \exp\left\{\sum_{p=1}^{P} \text{Re}\{\kappa\mathbf{c}_r^H\mathbf{u}_r\}+\mathbf{u}_r^H\mathbf{M}_r\mathbf{u}_r\right\}$	CGBL($\kappa\mathbf{C}, \{\mathbf{M}_r\}$)
Invariant Bingham CIB(κ, $\mathbf{A}$) $p(\mathbf{U}) \propto \text{etr}\left\{\kappa\mathbf{U}^H\mathbf{A}\mathbf{U}\right\}$	SM: $p(\mathbf{U}\|\mathbf{Z})_{\text{SM}} \propto \text{etr}\left\{\mathbf{U}^H(\kappa\mathbf{A}+\mathbf{W})\mathbf{U}\right\}$	Closed form $\mathcal{P}_R\{\kappa\mathbf{A}+\mathbf{W}\}$
	GM: $p(\mathbf{U}\|\mathbf{Z})_{\text{GM}} \propto \exp\left\{\sum_{p=1}^{P} \mathbf{u}_r^H(\kappa\mathbf{A}+\mathbf{M}_r)\mathbf{u}_r\right\}$	CGB($\{\kappa\mathbf{A}+\mathbf{M}_r\}$)

Algorithm 2: MMSD for the general model (2.1) $\mathbf{U}_{\text{MMSD}}$

input : $\mathbf{Z}$, $\mathbf{C}$, $\{\mathbf{A}_r\}$, σ^2, R, K, N, N_{bi}, N_r
output : $\widehat{\mathbf{U}}_{\text{MMSD}}$, $\{\tau_k\}$, $\{\lambda_r\}$
initialize: $\widehat{\mathbf{U}}^0$, $\{\tau_k^0\}$ and $\{\lambda_r^0\}$
for $t = 0 \dots T-1$ **do**
  for n=1 $\dots N_{bi}+N_r$ **do**
    Sample $\mathbf{U}_{(n)} = \text{CGBL}(\mathbf{C}, \{\mathbf{G}_r^t\})$ ccf **Appendix A**
  end
  Update $\widehat{\mathbf{U}}^{t+1} \approx \mathcal{P}_R\left\{\frac{1}{N_r}\sum_{n=N_{bi}+1}^{N_{bi}+N_r} \mathbf{U}_{(n)}{\mathbf{U}_{(n)}}^H\right\}$
  Update τ_k^{t+1} $\forall k$ with (2.16)
  Update λ_r^{t+1} $\forall r$ with (2.18)
end

Using the Sherman Morrison Woodbury lemma, $\mathbf{\Sigma}_k^{-1}$ reads as

$$\mathbf{\Sigma}_k^{-1} = (\tau_k \mathbf{U}\mathbf{U}^H + \sigma^2\mathbf{I})^{-1} = \sigma^{-2}\mathbf{I} - \frac{\tau_k}{\sigma^2(\tau_k+\sigma^2)}\mathbf{U}\mathbf{U}^H, \ \forall k \tag{2.24}$$

Then, the p.d.f. $p(\mathbf{Z}|\mathbf{U}, \{\tau_k\})$ reduces to

$$\begin{aligned} p(\mathbf{Z}|\mathbf{U}, \{\tau_k\}) &\propto \prod_{k=1}^{K} p(\mathbf{z}_k|\mathbf{U}, \tau_k) \propto \prod_{k=1}^{K} \frac{\exp\left\{-\mathbf{z}_k^H\mathbf{\Sigma}_k^{-1}\mathbf{z}_k\right\}}{\det(\mathbf{\Sigma}_k)} \\ &\propto \exp\left\{\sum_{k=1}^{K} \frac{\tau_k}{\sigma^2(\tau_k+\sigma^2)}\mathbf{U}^H\mathbf{z}_k\mathbf{z}_k\mathbf{U}\right\}\left(\prod_{k=1}^{K}(\tau_k+\sigma^2)^{-R}\right) \\ &\propto \text{etr}\left\{\mathbf{U}^H\mathbf{W}\mathbf{U}\right\}\left(\prod_{k=1}^{K}(\tau_k+\sigma^2)^{-R}\right) \end{aligned} \tag{2.25}$$

where $\mathbf{W} = \mathbf{Z}\mathbf{B}\mathbf{Z}^H$ and $\mathbf{B} = \text{diag}\left(\left\{\frac{\tau_1}{\sigma^2(\tau_1+\sigma^2)} \dots \frac{\tau_K}{\sigma^2(\tau_K+\sigma^2)}\right\}\right)$. In order to obtain closed form expression, we assign to $\mathbf{U}$ a conjugate prior of (2.25), i.e., $\mathbf{U} \sim \text{CIB}(\kappa, \mathbf{A})$, thus its p.d.f. reads

as $p_{\text{CIB}}(\mathbf{U}) \propto \text{etr}\{\kappa \mathbf{U}^H \mathbf{A} \mathbf{U}\}$.

2.4.1 The MMSD estimator for the simplified model

In this case, the MMSD estimator of the basis $\widehat{\mathbf{U}}$ is defined as the minimizer of the following problem

$$\begin{aligned} \underset{\widehat{\mathbf{U}},\{\tau_k\}}{\text{minimize}} \quad & \mathbb{E}_{\mathbf{U},\mathbf{Z}} \left\{ \| \widehat{\mathbf{U}}\widehat{\mathbf{U}}^H - \mathbf{U}\mathbf{U}^H \|_F^2 \right\} \\ \text{subject to} \quad & \tau_k \geq 0, \ \forall k \\ & \widehat{\mathbf{U}}^H \widehat{\mathbf{U}} = \mathbf{I}_R \end{aligned} \tag{2.26}$$

Following the previous lines, we develop an iterative estimation method by solving (2.26) w.r.t. $\widehat{\mathbf{U}}$ and $\{\tau_k\}$ sequentially. The box Algorithm 3 sums-up the main steps of the estimation process which are detailed in the following.

2.4.1.1 Update of the basis $\widehat{\mathbf{U}}$

Now, we assume that the block $\{\tau_k^t\}$ is fixed, consequently, the updated $\widehat{\mathbf{U}}^{t+1}$ is the solution of the following problem

$$\begin{aligned} \underset{\widehat{\mathbf{U}}}{\text{minimize}} \quad & \mathbb{E}_{\mathbf{U},\mathbf{Z}} \left\{ \| \widehat{\mathbf{U}}\widehat{\mathbf{U}}^H - \mathbf{U}\mathbf{U}^H \|_F^2 \right\} \\ \text{subject to} \quad & \widehat{\mathbf{U}}^H \widehat{\mathbf{U}} = \mathbf{I}_R \end{aligned} \tag{2.27}$$

In this case, the updated basis $\widehat{\mathbf{U}}^{t+1}$ is derived as the MMSD estimator in (1.12) leading to

$$\widehat{\mathbf{U}}^{t+1} = \mathcal{P}_R \left\{ \mathbf{M}(p(\mathbf{U}|\mathbf{Z}, \{\tau_k^t\})) \right\} \tag{2.28}$$

where

$$\mathbf{M}(p(\mathbf{U}|\mathbf{Z}, \{\tau_k^t\})) = \int \mathbf{U}\mathbf{U}^H p(\mathbf{U}|\mathbf{Z}, \{\tau_k^t\}) d\mathbf{U} \tag{2.29}$$

From (2.25), the posterior probability $p(\mathbf{U}|\mathbf{Z}, \{\tau_k^t\})$ reads as

$$p(\mathbf{U}|\mathbf{Z}, \{\tau_k^t\}) \propto p(\mathbf{Z}|\mathbf{U}, \{\tau_k^t\}) p_{\text{CIB}}(\mathbf{U}) \propto \text{etr}\left\{\kappa \mathbf{U}^H \mathbf{A} \mathbf{U}\right\} \text{etr}\left\{\mathbf{U}^H \mathbf{W}^t \mathbf{U}\right\} \propto \text{etr}\left\{\mathbf{U}^H (\kappa \mathbf{A} + \mathbf{W}^t) \mathbf{U}\right\} \tag{2.30}$$

with $\mathbf{W}^t = \mathbf{Z}\mathbf{B}^t\mathbf{Z}^H$ and $\mathbf{B}^t = \text{diag}\left(\left\{\frac{\tau_1^t}{\sigma^2(\tau_1^t+\sigma^2)} \cdots \frac{\tau_K^t}{\sigma^2(\tau_K^t+\sigma^2)}\right\}\right)$. Using Proposition 1 from [66], we notice that the updated basis admits the following closed form expression

$$\widehat{\mathbf{U}}^{t+1} = \mathcal{P}_R \left\{ \int \mathbf{U}\mathbf{U}^H \text{etr}\left\{\mathbf{U}^H (\kappa \mathbf{A} + \mathbf{W}^t) \mathbf{U}\right\} d\mathbf{U} \right\} = \mathcal{P}_R \left\{ \kappa \mathbf{A} + \mathbf{W}^t \right\} \tag{2.31}$$

This specific model provides a closed form solution with an interesting interpretations. Indeed, this MMSD appears naturally as the principal subspace of the sum of the SCM using scaled samples and scaled prior subspace projector.

2.4.1.2 Update of the texture parameter

For fixed $\widehat{\mathbf{U}}^{t+1}$, the update of $\{\tau_k\}$ is obtained by maximizing the p.d.f. $p(\mathbf{Z}|\widehat{\mathbf{U}}^{t+1}, \{\tau_k\})$ as

$$\begin{aligned} \underset{\{\tau_k\}}{\text{maximize}} \quad & p(\mathbf{Z}|\widehat{\mathbf{U}}^{t+1}, \{\tau_k\}) \\ \text{subject to} \quad & \tau_k \geq 0, \ \forall k \end{aligned} \tag{2.32}$$

with $\mathbf{\Sigma}_k = \tau_k \widehat{\mathbf{U}}^{t+1}\widehat{\mathbf{U}}^{t+1^H} + \sigma^2 \mathbf{I}_N, \ \forall k$. Minimizing the negative log-likelihood is equivalent to solve

$$\begin{aligned} \underset{\{\tau_k\}}{\text{minimize}} \quad & \sum_{k=1}^{K} \ln\left(\det(\mathbf{\Sigma}_k^{t+1})\right) + \mathbf{z}_k^H \left(\mathbf{\Sigma}_k^{t+1}\right)^{-1} \mathbf{z}_k \\ \text{subject to} \quad & \tau_k \geq 0, \ \forall k \end{aligned} \tag{2.33}$$

which leads to

$$\tau_k^{t+1} = \max\left(\frac{\mathbf{z}_k^H \widehat{\mathbf{U}}^{t+1}\widehat{\mathbf{U}}^{t+1^H}\mathbf{z}_k}{R} - \sigma^2, 0\right), \ \forall k \tag{2.34}$$

2.4.2 The MAP estimator for the simplified model

From (2.30), the update of the basis of interest is the solution of the following problem

$$\begin{aligned} \underset{\widehat{\mathbf{U}}}{\text{maximize}} \quad & \widehat{\mathbf{U}}^H(\kappa\mathbf{A} + \mathbf{W}^t)\widehat{\mathbf{U}} \\ \text{subject to} \quad & \widehat{\mathbf{U}}^H\widehat{\mathbf{U}} = \mathbf{I}_R \end{aligned} \tag{2.35}$$

Given that $\kappa\mathbf{A} + \mathbf{W}^t \in \mathcal{H}_N^+$, the updated basis for the MAP estimator is

$$\widehat{\mathbf{U}}^{t+1} = \mathcal{P}_R\{\kappa\mathbf{A} + \mathbf{W}^t\} \tag{2.36}$$

which corresponds to (2.31). Furthermore, we can notice that the MAP update of $\{\tau_k^{t+1}\}$ is identical to (2.34). Therefore, in this case, the MAP estimator coincides with the MMSD estimator.

Algorithm 3: MMSD estimation of subspace for the simplified model $\widehat{\mathbf{U}}_{\text{sMMSD}}$

input : $\mathbf{Z}$, $\mathbf{A}$, σ^2, R, K, N, κ
output : $\widehat{\mathbf{U}}_{\text{sMMSD}}$, $\{\tau_k\}$
initialize: $\widehat{\mathbf{U}}^0$, $\{\tau_k^0\}$

while *stop criterion unreached* **do**
  Update $\widehat{\mathbf{U}}^t = \mathcal{P}_R\left\{\kappa\mathbf{A} + \mathbf{W}^t\right\}$
  Update $\tau_k^t = \max\left(\frac{\mathbf{z}_k^H \widehat{\mathbf{U}}^t\widehat{\mathbf{U}}^{t^H}\mathbf{z}_k}{R} - \sigma^2, 0\right)$
end

Chapter 3

Bayesian subspace estimators in the presence of outliers

The newly proposed subspace estimators in Chapter 2 have been derived in the context of CG sources embedded in WGN. These estimators may show poor robustness to the introduction of outliers. Notably, samples corrupted by outliers are distorting the dominant eigensubspace of the SCM (phenomenon also referred to as subspace swap [77]). As a remedy, we aim to build an estimate of the basis of interest that is robust to outliers. We derive in the following, the MAP and the MMSD estimators in the context of CG sources embedded in heterogeneous noise (CG outliers plus WGN).

3.1 Adding outliers to the CG plus WGN model

The data is modeled as a sum of CG sources $\mathbf{s}_k$, WGN $\mathbf{n}_k$ and CG outlier $\mathbf{o}_k$. For this model, the samples $\{\mathbf{z}_k\}$ reads as

$$\mathbf{z}_k = \mathbf{s}_k + \mathbf{o}_k + \mathbf{n}_k, \ \forall k \in [\![1, K]\!] \tag{3.1}$$

For each of the parameters, we will assume the following distributions

- The signal $\mathbf{s}_k \sim \mathcal{CG}(0, \mathbf{\Sigma}, \tau_k)$ lies in low rank subspace of dimension R which is assumed pre-established. The SVD of the source scatter matrix is assumed to be $\mathbf{\Sigma} \overset{\text{SVD}}{=} [\mathbf{U}|\mathbf{U}^\perp] \begin{bmatrix} \mathbf{I}_R & 0 \\ 0 & 0 \end{bmatrix} [\mathbf{U}|\mathbf{U}^\perp]^H$. Notice that, as done in [30, 72], the non-null eigenvalues of $\mathbf{\Sigma}$ are assumed to be equal to 1. The hypothesis of eigenvalues equality is a relaxation that is made for tractability purposes but still offers interesting performance in practice [30]. Additionally, any scaling of the signal covariance matrix is absorbed in the textures τ_k of the CG distribution.

- The signal subspace orthonormal basis is distributed as $\mathbf{U} \sim \text{CGBL}(\mathbf{C}, \{\mathbf{A}_r\})$.

- The potential outlier is distributed as $\mathbf{o}_k \sim \mathcal{CG}(0, \mathbf{\Sigma}_o, \beta_k)$, with a covariance matrix defined by $\mathbf{\Sigma}_o = \mathbf{I}_N - \mathbf{U}\mathbf{U}^H = \mathbf{U}^\perp \mathbf{U}^{\perp H}$, where $\mathbf{U}^\perp$ is orthonormal complement of the source subspace basis. This outlier model is interesting in terms of robustness since it corresponds to the less-informative prior, as well as the worst-case outlier in terms of subspace estimation.

- The additive noise $\mathbf{n}_k$ is distributed as $\mathbf{n}_k \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_N)$. For the sake of clarity the variance σ^2 is assumed pre-estimated.

These hypotheses lead to $\mathbf{z}_k|\tau_k,\beta_k,\mathbf{U}\sim\mathcal{CN}(0,\mathbf{\Sigma}_k)$ with

$$\mathbf{\Sigma}_k=\tau_k\mathbf{U}\mathbf{U}^H+\beta_k\mathbf{U}^\perp\mathbf{U}^{\perp H}+\sigma^2\mathbf{I}_N \tag{3.2}$$

Then, the p.d.f. of $\mathbf{Z}$ conditional to $\{\tau_k\},\{\beta_k\},\mathbf{U}$ with $\mathbf{Z}=[\mathbf{z}_1,\ldots,\mathbf{z}_K]$, reads

$$p(\mathbf{Z}|\{\tau_k\},\{\beta_k\},\mathbf{U})=\prod_{k=1}^{K}p(\mathbf{z}_k|\tau_k,\beta_k,\mathbf{U})\propto\prod_{k=1}^{K}\frac{\exp\{-\mathbf{z}_k^H\mathbf{\Sigma}_k^{-1}\mathbf{z}_k\}}{\det(\mathbf{\Sigma}_k)} \tag{3.3}$$

Thanks to the Sherman Morisson Woodbury lemma, the expression of $\mathbf{\Sigma}_k^{-1}$ is given by

$$\begin{aligned}\mathbf{\Sigma}_k^{-1}&=\frac{1}{\sigma^2+\beta_k}\mathbf{I}-\frac{\tau_k-\beta_k}{(\sigma^2+\tau_k)(\sigma^2+\beta_k)}\mathbf{U}\mathbf{U}^H\\&=\frac{1}{\sigma^2+\beta_k}\mathbf{I}-\frac{\tau_k}{\sigma^2(\sigma^2+\tau_k)}\mathbf{U}\mathbf{U}^H+\frac{\beta_k}{\sigma^2(\sigma^2+\beta_k)}\mathbf{U}\mathbf{U}^H\end{aligned} \tag{3.4}$$

From (3.3), the posterior probability of $\mathbf{U}|\mathbf{Z},\{\tau_k\},\{\beta_k\}$ is expressed as

$$\begin{aligned}p\left(\mathbf{U}|\mathbf{Z},\{\tau_k\},\{\beta_k\}\right)&\propto p\left(\mathbf{Z}|\mathbf{U},\{\tau_k\},\{\beta_k\}\right)p_{\mathrm{CGBL}}(\mathbf{U})\\&\propto\prod_{k=1}^{K}\frac{\exp\left\{-\mathbf{z}_k^H\mathbf{\Sigma}_k^{-1}\mathbf{z}_k\right\}}{\det(\mathbf{\Sigma}_k)}p_{\mathrm{CGBL}}(\mathbf{U})\\&\propto\prod_{k=1}^{K}h_k\exp\left\{\frac{\tau_k}{\sigma^2(\sigma^2+\tau_k)}\mathbf{z}_k^H\mathbf{U}\mathbf{U}^H\mathbf{z}_k\right\}\prod_{k=1}^{K}\exp\left\{-\frac{\beta_k}{\sigma^2(\sigma^2+\beta_k)}\mathbf{z}_k^H\mathbf{U}\mathbf{U}^H\mathbf{z}_k\right\}p_{\mathrm{CGBL}}(\mathbf{U})\\&\propto\left(\prod_{k=1}^{K}h_k\right)\mathrm{etr}\{\mathbf{U}^H\mathbf{M}\mathbf{U}\}\exp\left\{\sum_{r=1}^{R}\mathrm{Re}\{\mathbf{c}_r^H\mathbf{u}_r\}+\mathbf{u}_r^H\mathbf{A}_r\mathbf{u}_r\right\}\\&\propto\left(\prod_{k=1}^{K}h_k\right)\exp\left\{\sum_{r=1}^{R}\mathrm{Re}\{\mathbf{c}_r^H\mathbf{u}_r\}+\mathbf{u}_r^H(\mathbf{A}_r+\mathbf{M})\mathbf{u}_r\right\}\end{aligned} \tag{3.5}$$

with

$$\begin{cases}h_k=\dfrac{1}{(\tau_k+\sigma^2)^R(\beta_k+\sigma^2)^{N-R}}\\\mathbf{M}=\mathbf{Z}\mathrm{diag}\left(\dfrac{\tau_1-\beta_1}{(\sigma^2+\tau_1)(\sigma^2+\beta_1)},\ldots,\dfrac{\tau_K-\beta_K}{(\sigma^2+\tau_K)(\sigma^2+\beta_K)}\right)\mathbf{Z}^H\end{cases}$$

3.2 MAP estimator

In this section, we aim to develop the MAP of the subspace orthonormal basis $\mathbf{U}$ according to the data model in (3.1). This subspace estimator $\widehat{\mathbf{U}}$ maximizes the posterior probability $p\left(\mathbf{U}|\mathbf{Z},\{\tau_k\},\{\beta_k\}\right)$. Our proposed MAP estimator reads as the solution of the following problem

$$\begin{aligned}\underset{\widehat{\mathbf{U}},\{\tau_k\},\{\beta_k\}}{\text{maximize}}\quad&p(\widehat{\mathbf{U}}|\mathbf{Z},\{\tau_k\},\{\beta_k\})\\\text{subject to}\quad&\tau_k\geq 0,\ \beta_k\geq 0\ \forall k\\&\widehat{\mathbf{U}}^H\widehat{\mathbf{U}}=\mathbf{I}_R\end{aligned} \tag{3.6}$$

This problem is equivalent to minimizing the negative $\log\Big(p(\widehat{\mathbf{U}}|\mathbf{Z},\{\tau_k\},\{\beta_k\})\Big)$. For (3.6) and (3.5), this optimization problem is recasted as following

$$\begin{array}{ll}\underset{\widehat{\mathbf{U}},\{\tau_k\},\{\beta_k\}}{\text{minimize}} & \displaystyle\sum_{k=1}^{K} R\ln(\tau_k+\sigma^2)+(N-R)\ln(\beta_k+\sigma^2)-\sum_{r=1}^{R}\mathrm{Re}\{\mathbf{c}_r^H\widehat{\mathbf{u}}_r\}+\widehat{\mathbf{u}}_r^H(\mathbf{A}_r+\mathbf{M})\widehat{\mathbf{u}}_r \\ \text{subject to} & \tau_k\geq 0,\ \beta_k\geq 0\ \forall k \\ & \widehat{\mathbf{U}}^H\widehat{\mathbf{U}}=\mathbf{I}_R\end{array} \tag{3.7}$$

This formulation involves non-trivial optimization problem due to the unitary constraint. We propose an iterative algorithm which updates sequentially the parameters $\mathbf{U}$, $\{\tau_k\}$ and $\{\beta_k\}$. For updating $\mathbf{U}$, we make use of the MM algorithm for solving the above problem w.r.t. $\mathbf{U}$ and while fixing the remain variables. Whereas, for updating $\{\tau_k\}$ and $\{\beta_k\}$, we obtain closed form updates for both of them. The main steps of the estimation process are summed up in the box Algorithm 4.

3.2.1 Update of the basis U

Considering $\mathbf{U}$ while fixing the remain variables, the objective of (3.7) w.r.t. the variable $\mathbf{U}$ can be expressed as

$$\begin{array}{ll}\underset{\widehat{\mathbf{U}}}{\text{maximize}} & \displaystyle\sum_{r=1}^{R}\mathrm{Re}\{\mathbf{c}_r^H\widehat{\mathbf{u}}_r\}+\widehat{\mathbf{u}}_r^H(\mathbf{A}_r+\mathbf{M}^t)\widehat{\mathbf{u}}_r \\ \text{subject to} & \widehat{\mathbf{U}}^H\widehat{\mathbf{U}}=\mathbf{I}_R\end{array} \tag{3.8}$$

with

$$\mathbf{M}^t=\mathbf{Z}\mathrm{diag}\left(\left\{\frac{\tau_k^t-\beta_k^t}{(\sigma^2+\tau_k^t)(\sigma^2+\beta_k^t)}\right\}\right)\mathbf{Z}^H \tag{3.9}$$

where the subscript $.^t$ denotes the updated value at iteration t. The above problem does not have a trivial solution. As an alternative, we apply the MM algorithm in order to obtain closed form updates in each step while incrementing the objective value. Using the Proposition B.1, the update of $\mathbf{U}$ is obtained as

$$\widehat{\mathbf{U}}^{t+1}=\mathcal{P}_{\mathrm{Proc}}(\mathbf{F}^t) \tag{3.10}$$

where

$$\mathbf{F}^t=\mathbf{S}^t+1/2\mathbf{C}\ \text{ and }\ \mathbf{S}^t=[\ (\mathbf{A}_1+\mathbf{M}^t)\mathbf{u}_1^t\ |\ \ldots\ |\ (\mathbf{A}_R+\mathbf{M}^t)\mathbf{u}_R^t\ ] \tag{3.11}$$

with $\mathbf{u}_r^t$ is the r-th column of the matrix $\mathbf{U}^t$.

3.2.1.1 Special case for the Bingham distribution

We present a special case ($\mathbf{C}$ is a null matrix and $\mathbf{A}_r=\kappa\mathbf{A}$, $\forall r$), i.e., $\mathbf{U}\sim\mathrm{CIB}(\kappa,\mathbf{A})$ for which the MAP estimator is obtained directly (rather than using a MM update). The above problem (3.8) reduces to

$$\begin{array}{ll}\underset{\widehat{\mathbf{U}}}{\text{maximize}} & \displaystyle\sum_{r=1}^{R}\widehat{\mathbf{u}}_r^H(\kappa\mathbf{A}+\mathbf{M}^t)\widehat{\mathbf{u}}_r \\ \text{subject to} & \widehat{\mathbf{U}}^H\widehat{\mathbf{U}}=\mathbf{I}_R\end{array} \tag{3.12}$$

Equivalently, this problem is expressed as

$$\begin{aligned} \underset{\widehat{\mathbf{U}}}{\text{maximize}} \quad & \text{Tr}\left\{\widehat{\mathbf{U}}^H(\kappa\mathbf{A}+\mathbf{M}^t)\widehat{\mathbf{U}}\right\} \\ \text{subject to} \quad & \widehat{\mathbf{U}}^H\widehat{\mathbf{U}} = \mathbf{I}_R \end{aligned} \tag{3.13}$$

As the matrix $\kappa\mathbf{A}+\mathbf{M}^t \in \mathcal{H}_N^+$, the updated basis of interest has the following expression

$$\widehat{\mathbf{U}}^{t+1} = \mathcal{P}_R\{\kappa\mathbf{A}+\mathbf{M}^t\} \tag{3.14}$$

We present an iterative algorithm to compute this estimator for this case in the box Algorithm 5.

3.2.2 Update of textures

For fixed basis $\widehat{\mathbf{U}}^{t+1}$, the update of texture $\{\tau_k\}$ and $\{\beta_k\}$ are the solutions of the problem

$$\begin{aligned} \underset{\{\tau_k\},\{\beta_k\}}{\text{maximize}} \quad & p(\mathbf{Z}|\widehat{\mathbf{U}}^{t+1},\{\tau_k\},\{\beta_k\}) \\ \text{subject to} \quad & \tau_k \geq 0,\ \beta_k \geq 0\ \forall k \end{aligned} \tag{3.15}$$

The above problem is equivalent to minimizing the negative log-likelihood as

$$\begin{aligned} \underset{\{\tau_k\},\{\beta_k\}}{\text{minimize}} \quad & \sum_{k=1}^{K} \ln\left(\det(\mathbf{\Sigma}_k)\right) + \mathbf{z}_k^H\mathbf{\Sigma}_k^{-1}\mathbf{z}_k \\ \text{subject to} \quad & \tau_k \geq 0,\ \beta_k \geq 0\ \forall k \end{aligned} \tag{3.16}$$

The objective function is separable in the τ_k's and the β_k's. Following the same methodology as in [30], the blocks $\{\tau_k^{t+1}\}$ and $\{\beta_k^{t+1}\}$ are updated as

$$\tau_k^{t+1} = \max\left(\frac{\mathbf{z}_k^H\widehat{\mathbf{U}}^{t+1}\widehat{\mathbf{U}}^{t+1^H}\mathbf{z}_k}{R} - \sigma^2, 0\right),\ \forall k \tag{3.17}$$

and

$$\beta_k^{t+1} = \max\left(\frac{\mathbf{z}_k^H\widehat{\mathbf{U}}^{\perp^{t+1}}\widehat{\mathbf{U}}^{\perp^{t+1^H}}\mathbf{z}_k}{N-R} - \sigma^2, 0\right),\ \forall k \tag{3.18}$$

3.3 Minimum mean square distance estimator

According to the data model (3.1) and the generic expression of the MMSD (1.11), the subspace estimator of $\mathbf{U}$ is expressed as the solution of the following optimization problem

$$\begin{aligned} \underset{\widehat{\mathbf{U}},\{\tau_k\},\{\beta_k\}}{\text{minimize}} \quad & \mathbb{E}_{\mathbf{Z},\mathbf{U}}\left\{\parallel \widehat{\mathbf{U}}\widehat{\mathbf{U}}^H - \mathbf{U}\mathbf{U}^H \parallel_F^2\right\} \\ \text{subject to} \quad & \tau_k \geq 0,\ \beta_k \geq 0\ \ \forall k \\ & \widehat{\mathbf{U}}^H\widehat{\mathbf{U}} = \mathbf{I}_R \end{aligned} \tag{3.19}$$

This problem is hard to solve due to the non-convex constraints and the expectation function. Therefore, we propose a block coordinate descent algorithm that updates sequentially the basis of interest $\mathbf{U}$ and the texture parameters $\{\tau_k\}$ and $\{\beta_k\}$. The algorithm has a guaranteed convergence in terms of objective value and is summed up in the box Algorithm 6.

Algorithm 4: MAP for the outlier model (3.1) $\mathbf{U}_{\text{MAP-OM}}$

input : $\mathbf{Z}$, $\mathbf{C}$, $\{\mathbf{A}_r\}$, σ^2, R, K, N
output : MAP estimators of $\mathbf{U}$, $\{\tau_k\}$, $\{\lambda_r\}$
initialize: $\widehat{\mathbf{U}}^0$, $\{\tau_k^0\}$ and $\{\beta_k^0\}$
for $t = 0 \dots T-1$ **do**
  Update $\widehat{\mathbf{U}}^{t+1} = \mathcal{P}_{\text{Proc}}(\mathbf{F}^t)$, with $\mathbf{F}^t$ in (3.11)
  Update $\tau_k^{t+1} = \max\left(\frac{\mathbf{z}_k^H \widehat{\mathbf{U}}^{t+1}\widehat{\mathbf{U}}^{t+1^H}\mathbf{z}_k}{R} - \sigma^2, 0\right)$
  Update $\beta_k^{t+1} = \max\left(\frac{\mathbf{z}_k^H \widehat{\mathbf{U}}^{\perp^{t+1}}\widehat{\mathbf{U}}^{\perp^{t+1}H}\mathbf{z}_k}{N-R} - \sigma^2, 0\right)$
end

3.3.1 Update of the basis U

For fixed texture $\{\tau_k^t\}$ and $\{\beta_k^t\}$, the update of $\mathbf{U}$ is obtained by solving the following problem

$$\begin{aligned} \underset{\widehat{\mathbf{U}}}{\text{minimize}} \quad & \mathbb{E}_{\mathbf{Z},\mathbf{U}}\left\{ \| \widehat{\mathbf{U}}\widehat{\mathbf{U}}^H - \mathbf{U}\mathbf{U}^H \|_F^2 \right\} \\ \text{subject to} \quad & \widehat{\mathbf{U}}^H\widehat{\mathbf{U}} = \mathbf{I}_R \end{aligned} \tag{3.20}$$

From (3.5) and (1.12), the expression of the updated orthonormal basis $\widehat{\mathbf{U}}^{t+1}$ reads

$$\begin{aligned} \widehat{\mathbf{U}}^{t+1} &= \mathcal{P}_R\left\{ \int \mathbf{U}\mathbf{U}^H p(\mathbf{U}|\mathbf{Z}, \{\tau_k^t\}, \{\beta_k^t\}) d\mathbf{U} \right\} \\ &= \mathcal{P}_R\left\{ \int \mathbf{U}\mathbf{U}^H \text{etr}\left\{ \sum_{r=1}^{R} \text{Re}\{\mathbf{c}_r^H \mathbf{u}_r\} + \mathbf{u}_r^H(\mathbf{A}_r + \mathbf{M}^t)\mathbf{u}_r \right\} d\mathbf{U} \right\} \end{aligned} \tag{3.21}$$

The above posterior probability $p(\mathbf{U}|\mathbf{Z}, \{\tau_k^t\}, \{\beta_k^t\})$ is recognized as the CGBL$(\mathbf{C}, \{\mathbf{L}_r^t\})$, in which $\mathbf{L}_r^t = \mathbf{A}_r + \mathbf{M}^t$, $\forall r$. In this case, there is no closed form of the MMSD estimator. Therefore, a Gibbs sampler scheme can be used to update the basis $\widehat{\mathbf{U}}^{t+1}$ using the so-called IAM as shown below

$$\widehat{\mathbf{U}}^{t+1} = \mathcal{P}_R\left\{ \frac{1}{N_r} \sum_{n=N_{bi}+1}^{N_{bi}+N_r} \mathbf{U}_{(n)}\mathbf{U}_{(n)}{}^H \right\}, \tag{3.22}$$

where $\mathbf{U}_{(n)} \sim$ CGBL$(\mathbf{C}, \{\mathbf{L}_r^t\})$. Special cases of posterior distributions under the data model (3.1) for standard priors of $\mathbf{U}$ are given in Table 3.1.

3.3.1.1 Special case for the Bingham distribution

Consider the special case of the CIB distribution, defined by $\mathbf{C} = \mathbf{0}$ and $\mathbf{A}_r = \kappa\mathbf{A}$, $\forall r \in [\![1, R]\!]$. This case is denoted $\mathbf{U} \sim$ CIB$(\kappa, \mathbf{A})$, i.e., $p_{\text{CIB}}(\mathbf{U}) \propto \text{etr}\{\kappa\mathbf{U}^H\mathbf{A}\mathbf{U}\}$. We have

$$\begin{aligned} \widehat{\mathbf{U}}^{t+1} &= \mathcal{P}_R\left\{ \int \mathbf{U}\mathbf{U}^H p(\mathbf{U}|\mathbf{Z}) d\mathbf{U} \right\} \\ &= \mathcal{P}_R\left\{ \int \mathbf{U}\mathbf{U}^H p(\mathbf{Z}|\mathbf{U}, \boldsymbol{\tau}) p_{\text{CIB}}(\mathbf{U}) d\mathbf{U} \right\} \\ &= \mathcal{P}_R\left\{ \int \mathbf{U}\mathbf{U}^H \text{etr}\left\{ \mathbf{U}^H(\kappa\mathbf{A} + \mathbf{M}^t)\mathbf{U} \right\} d\mathbf{U} \right\}. \end{aligned} \tag{3.23}$$

Algorithm 5: MMSD for the outlier model (3.1) $\mathbf{U}_{\text{MMSD-OM}}$

input : $\mathbf{Z}$, $\mathbf{C}$, $\{\mathbf{A}_r\}$, σ^2, R, K, N, N_{bi}, N_r
output : $\widehat{\mathbf{U}}_{\text{MMSD-OM}}$, $\{\tau_k\}$, $\{\beta_k\}$
initialize: $\widehat{\mathbf{U}}^0$, $\{\tau_k^0\}$ and $\{\lambda_r^0\}$

for $t = 0 \dots T-1$ **do**
  for n=$1 \dots N_{bi}+N_r$ **do**
    Sample $\mathbf{U}_{(n)} = \text{CGBL}(\mathbf{C}, \{\mathbf{L}_r^t\})$ ccf **Appendix A**
  end
  Update $\widehat{\mathbf{U}}^{t+1} \approx \mathcal{P}_R\left\{\frac{1}{N_r}\sum_{n=N_{bi}+1}^{N_{bi}+N_r} \mathbf{U}_{(n)}{\mathbf{U}_{(n)}}^H\right\}$
  Update $\tau_k^{t+1} = \max\left(\frac{\mathbf{z}_k^H \widehat{\mathbf{U}}^{t+1}{\widehat{\mathbf{U}}^{t+1}}^H \mathbf{z}_k}{R} - \sigma^2, 0\right)$
  Update $\beta_k^{t+1} = \max\left(\frac{\mathbf{z}_k^H \widehat{\mathbf{U}}^{\perp^{t+1}}{\widehat{\mathbf{U}}^{\perp^{t+1}}}^H \mathbf{z}_k}{N-R} - \sigma^2, 0\right)$
end

Table 3.1: Posterior distributions for standard priors on $\mathbf{U}$ under outlier model (OM)

The prior complex distribution of $\mathbf{U}$	Posterior distribution $p(\mathbf{U}\|\mathbf{Z})$ for OM	to be sampled for MMSD
Generalized Bingham CGB($\mathbf{A}, \mathbf{\Phi}$)	OM: $p(\mathbf{U}\|\mathbf{Z})_{\text{OM}} \propto \exp\left\{\sum_{p=1}^{P} \mathbf{u}_r^H(\phi(r)\mathbf{A}+\mathbf{M})\mathbf{u}_r\right\}$	CGB($\{\phi(r)\mathbf{A}+\mathbf{M}\}$)
Langevin CL($\mathbf{C}$)	OM: $p(\mathbf{U}\|\mathbf{Z})_{\text{OM}} \propto \text{etr}\left\{\text{Re}\{\mathbf{C}^H\mathbf{U}\} + \mathbf{U}^H\mathbf{M}\mathbf{U}\right\}$	CBL($\mathbf{C}, \mathbf{I}_R, \mathbf{M}$)
Invariant Bingham CIB($\kappa, \mathbf{A}$)	OM: $p(\mathbf{U}\|\mathbf{Z})_{\text{OM}} \propto \text{etr}\left\{\mathbf{U}^H(\kappa\mathbf{A}+\mathbf{M})\mathbf{U}\right\}$	Closed form $\mathcal{P}_R\{\kappa\mathbf{A}+\mathbf{M}\}$

Using Proposition 1 of [66] extended to the complex case, we can obtain a closed form solution of the update as

$$\widehat{\mathbf{U}}^{t+1} = \mathcal{P}_R\left\{\mathbf{M}^t + \kappa\bar{\mathbf{U}}\bar{\mathbf{U}}^H\right\} \tag{3.24}$$

We note that under these circumstances the MMSD (3.24) and the MAP (3.14) coincide. This closed form expression can be a practical tool, as it alleviates the computational cost of both the Gibbs scheme and the MM algorithm.

3.3.2 Update of textures

The updates of these variables $\{\tau_k\}$ and $\{\beta_k\}$ are equivalent to the solutions of the following problem (3.16) given respectively in (3.17) and (3.18).

Algorithm 6: MMSD estimation of subspace for the special case, i.e., $\mathbf{U} \sim \mathrm{CIB}(\kappa, \bar{\mathbf{U}}\bar{\mathbf{U}}^H)$ $\widehat{\mathbf{U}}_{\text{MMSD-OM-B}}$

input : $\mathbf{Z}$, $\mathbf{A}$, σ^2, R, K, N, κ
output : $\widehat{\mathbf{U}}_{\text{MMSD-OM-B}}$, $\{\tau_k\}$, $\{\beta_k\}$
initialize: $\widehat{\mathbf{U}}^0$, $\{\tau_k^0\}$, $\{\beta_k^0\}$

while *stop criterion unreached* **do**
  Update $\widehat{\mathbf{U}}^t = \mathcal{P}_R\left\{\kappa\mathbf{A} + \mathbf{M}^t\right\}$
  Update $\tau_k^{t+1} = \max\left(\frac{\mathbf{z}_k^H \widehat{\mathbf{U}}^{t+1}\widehat{\mathbf{U}}^{t+1^H}\mathbf{z}_k}{R} - \sigma^2, 0\right)$
  Update $\beta_k^{t+1} = \max\left(\frac{\mathbf{z}_k^H \widehat{\mathbf{U}}^{\perp\,t+1}\widehat{\mathbf{U}}^{\perp\,t+1^H}\mathbf{z}_k}{N-R} - \sigma^2, 0\right)$
end

Chapter 4

Simulation and applications

4.1 Simulations

4.1.1 Setup

To illustrate the performance of the proposed estimators, we evaluate their average fraction of energy (AFE) through Monte Carlo simulations. The AFE is considered as an adequate criteria of performance for subspace estimation, since it evaluates the closeness of the true range space $\mathbf{U}\mathbf{U}^H$ towards its estimate $\widehat{\mathbf{U}}\widehat{\mathbf{U}}^H$. The AFE of a given estimator $\widehat{\mathbf{U}}$ is expressed as:

$$\mathrm{AFE}(\widehat{\mathbf{U}}) = \mathbb{E}\{\mathrm{Tr}\{\mathbf{U}^H\widehat{\mathbf{U}}\widehat{\mathbf{U}}^H\mathbf{U}\}\}/R \tag{4.1}$$

The samples are generated from the model in (2.1), i.e., $\mathbf{z}_k \sim \mathcal{CN}(0, \tau_k\mathbf{U}\mathbf{\Lambda}\mathbf{U}^H + \sigma^2\mathbf{I})$. The texture parameters $\{\tau_k\}$ follow a Gamma distribution parameterized by its shape ν which reflects the heterogeneity of the sources, i.e., $\tau_k \sim \Gamma(\nu, \frac{1}{\nu})$, $\forall k$ (thus $\mathbb{E}\{\tau\} = 1$). We set $[\mathbf{\Lambda}]_{r,r} = (R+1-r)/(\sum_{i=1}^{R} i)$ and σ^2 to fix the SNR as SNR$= \mathrm{Tr}\{\mathbf{\Lambda}\}/\sigma^2$. We consider different scenarios for the distribution of $\mathbf{U}$:

- S1: $\mathbf{U}$ follows the complex Langevin distribution $\mathbf{U} \sim \mathrm{CL}(\kappa, \bar{\mathbf{U}})$,
- S2: $\mathbf{U}$ follows the complex generalized Bingham distribution $\mathbf{U} \sim \mathrm{CGB}(\{\kappa\phi_r\bar{\mathbf{U}}\bar{\mathbf{U}}^H\})$ where $\phi_{r,r} = (R+1-r)/(\sum_{i=1}^{R} i)$,
- S3: $\mathbf{U}$ follows the complex invariant Bingham distribution $\mathbf{U} \sim \mathrm{CIB}(\kappa, \bar{\mathbf{U}}\bar{\mathbf{U}}^H)$,

where $\bar{\mathbf{U}} \in \mathcal{U}_N^R$ are the first vectors of the canonical basis and the concentration parameters (κ and $\mathbf{\Phi}$) are set so that $\mathrm{AFE}(\bar{\mathbf{U}})$ has the same value for all the scenarios. We compare the following estimators:

i) $\widehat{\mathbf{U}}_{\mathrm{SCM}} = \mathcal{P}_R\{\mathbf{Z}\mathbf{Z}^H\}$, the estimator built from the SVD of the SCM.

ii) $\widehat{\mathbf{U}}_{\mathrm{MLE}}$ is the subspace estimates based MLE, in Section 1.2.3, computed with eigenvector block majorization minimization (EBMM) algorithm from [69].

iii) $\widehat{\mathbf{U}}_{\mathrm{MAP}}$ denotes the proposed MAP estimator, computed with Algorithm 1.

iv) $\widehat{\mathbf{U}}_{\mathrm{MMSD}}$ represents the proposed MMSD estimator computed with Algorithm 2.

v) $\widehat{\mathbf{U}}_{\mathrm{sMMSD}}$ is the simplified MMSD estimator, that assumes $\mathbf{\Lambda} = \mathbf{I}$ and $\mathbf{U} \sim \mathrm{CIB}(\kappa, \bar{\mathbf{U}}\bar{\mathbf{U}}^H)$, computed with Algorithm 3. This estimator is evaluated for S2 and S3 but the relaxation is not suited for S1 (where the true prior is a complex Langevin).

vi) $\widehat{\mathbf{U}}_{\text{MMSD-OM}}$ denotes the MMSD estimator derived in presence of outliers in Chapter 3, evaluated with Algorithm 5 suited to S1 and S2.

vii) $\widehat{\mathbf{U}}_{\text{MAP-OM}}$ is the MAP estimator for the outlier model (3.1) computed with Algorithm 4, suited to S1 and S2.

viii) $\widehat{\mathbf{U}}_{\text{MMSD-OM-B}}$ stands for the MMSD estimator where we assume presence of outliers and $\mathbf{U}$ follows the CIB distribution. We note that under these specific circumstances, the expression of both the MAP and the MMSD estimators coincide. The estimation process is summed up in Algorithm 6. This estimator is suited to S3.

ix) $\bar{\mathbf{U}}$ is the center of the prior distribution on $\mathbf{U}$.

4.1.2 Simulations results

Figure 4.1 displays the AFE in function of SNR for various sample size K in scenario S1. In this case, the SCM exhibits good performance in the standard regimes (high SNR and/or large K). The textures parameter is $\nu = 0.5$ so the sources are mildly impulsive. Therefore the MLE exhibits performances close to the SCM as it can be expected (differences will be observed in the following). However, both show their limits at low SNR. In this challenging context, Bayesian estimators can leverage the prior information and exhibit better performance in terms of AFE. Interestingly, for the complex Langevin prior, the MMSD outperforms the MAP and the MMSD-OM, which reach both performances close to SCM/MLE as the SNR increases.

Figure 4.2 displays the AFE in function of SNR for various sample size K in scenario S2. The same general observations as in the previous figure can be drawn. For the complex generalized Bingham prior case, the MMSD still outperforms the MAP, but not as significantly as in the scenario S1. We also observe that sMMSD, which assumes equals eigenvalues and a mismatched (averaged) prior, offers an interesting performance versus computational time trade-off when it comes to estimate only the signal subspace. By construction of the true prior, the first column-vectors of $\mathbf{U}$ exhibits less variance than the last ones. By uniformly averaging the prior for each vectors, sMMSD introduces a bias towards the center of distribution, which explains its performance close to $\bar{\mathbf{U}}$ at low SNR.

Figure 4.3 displays the AFE in function of SNR for various sample size K in scenario S3. Here, the sMMSD assumes the true prior and is only mismatched by assuming equals eigenvalues. In this scenario the sMMSD exhibits performance almost identical to the MMSD, which suggests that it is acceptable to relax the eigenvalue estimation when it comes to estimate only the signal subspace. The MMSD-OM-B shows close performance to the remaining Bayesian estimators.

Figure 4.4 displays the AFE in function of SNR for various sample size K in the actual simplified model, i.e., the scenario S3 where $\mathbf{\Lambda} = \mathbf{I}$. In this context, the MMSD and the MAP coincide and we still observe the interest of the Bayesian approach in challenging contexts (low SNR and/or K). We note that the sMMSD outperforms the MMSD-OM-B, since this estimator is derived in the context of corrupted samples by presence of outliers which mismatches the simplified model.

4.1.3 Robustness to the concentration parameter and the signal distribution

First, we study the effect of a miss-selected concentration parameter κ on the AFE of the proposed Bayesian estimator. The setup of Figure 4.5 is the same as for Figure 4.4 and displays the AFE of the MMSD estimator w.r.t. the assumed κ, while the true concentration parameter κ_0 is fixed. This figure illustrates that, for a reasonable range of κ, the AFE of the MMSD estimator remains

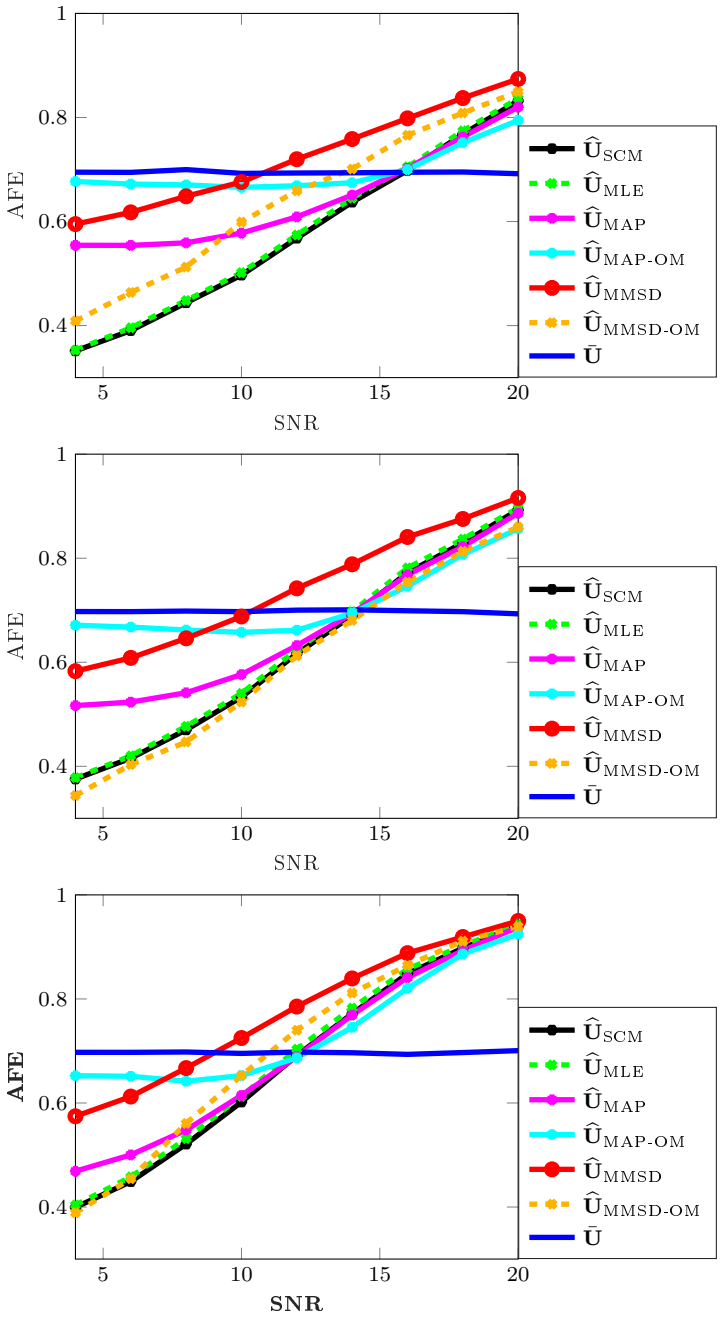


Figure 4.1: AFE w.r.t. SNR for R=5, $N = 20$, $\mathbf{U} \sim \text{CL}(\kappa, \bar{\mathbf{U}})$, $\kappa = 80$, $\nu = 0.5$, from top to bottom: $K = 3R$, $K = 4R$ and $K = 6R$

almost unchanged. Thus, the proposed method appears robust to a reasonable miss-selection of the concentration parameters of the assumed prior distribution.

Second, we study the performance of the proposed method w.r.t. the signal distribution, parameterized by the shape ν. The setup of Figure 4.6 is the same as for Figure 4.4 and displays the AFE of the MMSD estimator w.r.t ν for various SNRs. When $\nu \to 0.5$, the signal tends to be more impulsive (i.e., heavy tailed distributed). In this context, we can notice a

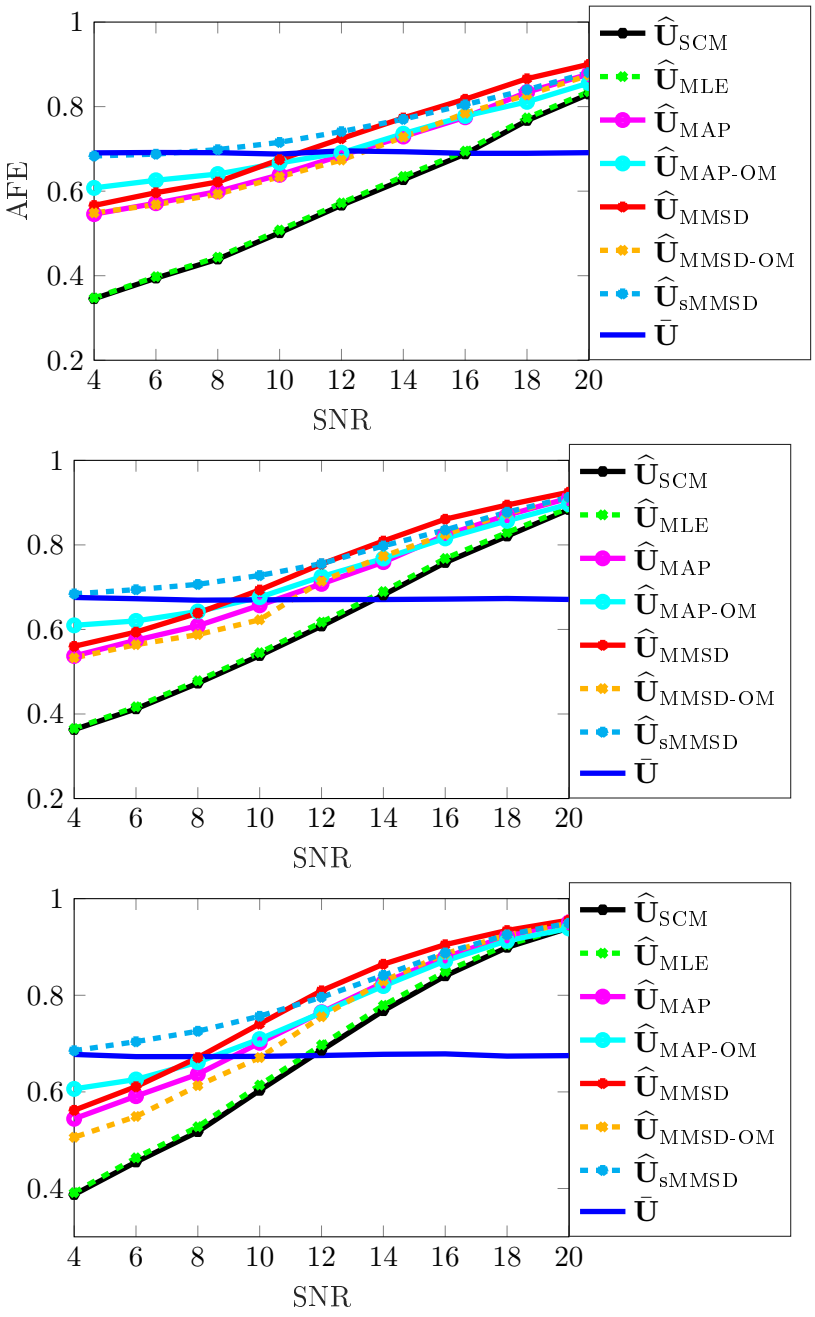


Figure 4.2: AFE w.r.t. SNR for R=5, $N = 20$, $\nu = 0.5$, $\mathbf{U} \sim \text{CGB}(\{\kappa_0 \boldsymbol{\phi}_r \bar{\mathbf{U}}\bar{\mathbf{U}}^H\})$, $\kappa_0 = 300$ from top to bottom: $K = 3R$, $K = 4R$ and $K = 6R$

slight difference between the SCM and the MLE, which illustrates the interest of taking the non-Gaussianity into account. Interestingly, the performance drop happens at lower ν for the Bayesian estimator, which shows the interest of exploiting both the non-Gaussian assumption and the prior knowledge.

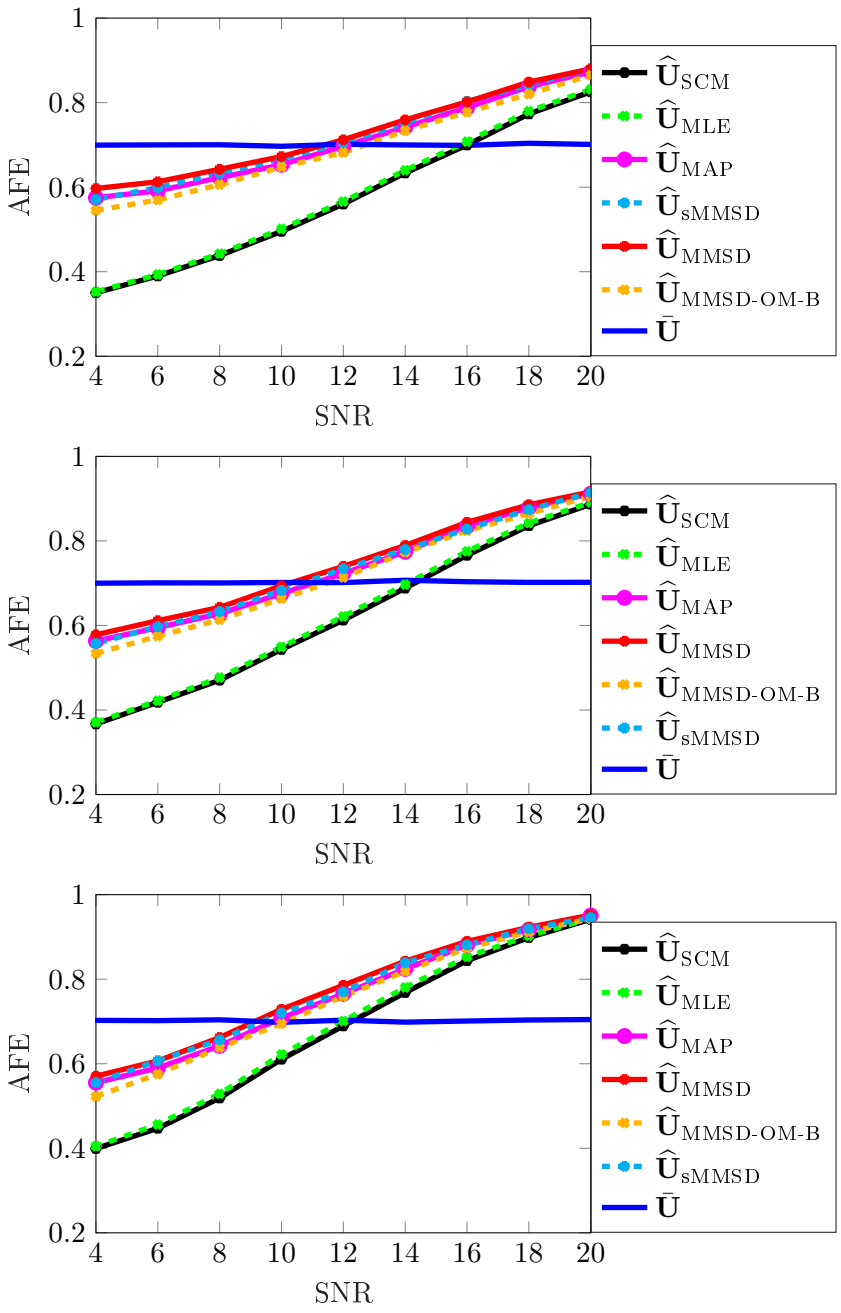


Figure 4.3: AFE w.r.t. SNR for R=5, $N = 20$, $\nu = 0.5$, $\mathbf{U} \sim \text{CIB}(\kappa, \bar{\mathbf{U}}\bar{\mathbf{U}}^H)$, $\kappa = 50$, from top to bottom: $K = 3R$, $K = 4R$ and $K = 6R$

4.1.4 Robustness to outliers

4.1.4.1 Setup

The data matrix $\mathbf{Z}$ is generated according to the model in (3.1). The WGN is generated as $\mathbf{n}_k \sim \mathcal{CN}(0, \sigma^2\mathbf{I})$. The signal follows a CG distribution as in (1.4) with $\mathbf{s}_k \overset{d}{=} \sqrt{\tau_k}\mathbf{g}_k \; \forall k$, $\tau_k \sim \Gamma(\nu, \frac{1}{\nu})$, $\mathbf{g}_k \sim \mathcal{CN}(0, \alpha\mathbf{U}\mathbf{U}^H)$, and SNR=$\log(\frac{\alpha}{\sigma^2})$. When outliers are present in a sample, they

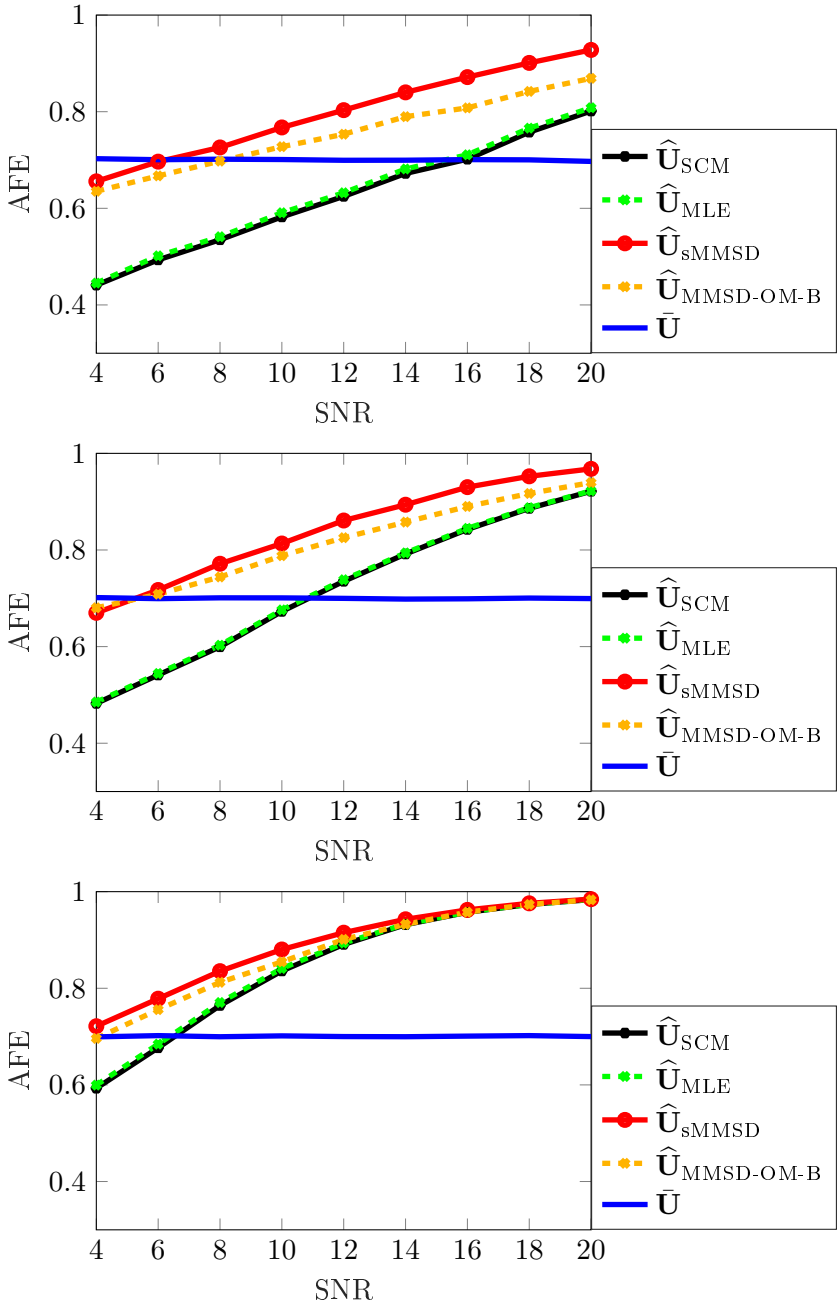


Figure 4.4: AFE w.r.t. SNR for R=5, $N = 20$, $\kappa = 50$, $\nu = 0.5$, for the simplified model from top to bottom: $K = 2R$, $K = 3R$ and $K = 4R$

are generated as $\mathbf{o}_k \overset{d}{=} \sqrt{\beta_k}\mathbf{d}_k \ \forall k$, $\beta_k \sim \Gamma(\nu', \frac{1}{\nu'})$, $\mathbf{d}_k \sim \mathcal{CN}(0, \gamma(\mathbf{I} - \mathbf{U}\mathbf{U}^H))$, and outlier to noise ratio ONR=$\log(\frac{\gamma}{\sigma^2})$.

4.1.4.2 List of estimators

- $\widehat{\mathbf{U}}_{\text{SCM}}$ denotes the estimator built from the SVD of the SCM.

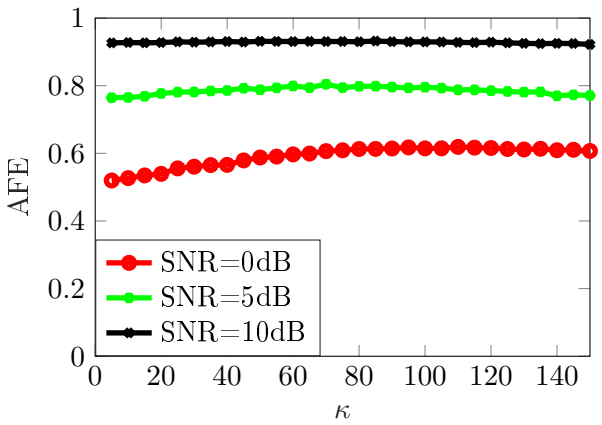


Figure 4.5: AFE of the MMSD w.r.t. assumed κ for various SNR, R=5, N =20, $\nu = 1$, $K = 30$, $\mathbf{U} \sim \text{CIB}(\kappa_0, \bar{\mathbf{U}}\bar{\mathbf{U}}^H)$, with true parameter $\kappa_0 = 60$ for the simplified model, MMSD=sMMSD=MAP

- $\widehat{\mathbf{U}}_{\text{MLE-out}}$ represents the estimator from [30], corresponding to the MLE estimator for the considered context, while assuming no prior distribution on the orthonormal basis $\mathbf{U}$ Section 1.2.4.
- $\widehat{\mathbf{U}}_{\text{MMSD-U}}$ is the MMSD estimator for uniformly distributed sources (1.15).
- $\widehat{\mathbf{U}}_{\text{MMSD-G}}$ stands for the MMSD estimator in presence of Gaussian sources [78].
- $\widehat{\mathbf{U}}_{\text{MMSD-OM-B}}$ denotes the proposed MMSD estimator detailed in Algorithm 6.
- $\bar{\mathbf{U}}$ is the center of distribution on $\mathbf{U}$ (non adaptive estimator).

4.1.4.3 Simulation results

Figure 4.7 displays the AFE of the different estimators w.r.t. the number of corrupted samples where we distinguish two types of samples: the corrupted ones, generated as $\mathbf{z}_k = \mathbf{s}_k + \mathbf{o}_k + \mathbf{n}_k$, and the non corrupted ones, generated as $\mathbf{z}_k = \mathbf{s}_k + \mathbf{n}_k$. In this context, the non-robust estimators $\widehat{\mathbf{U}}_{\text{SCM}}$, $\widehat{\mathbf{U}}_{\text{MMSD-U}}$, and $\widehat{\mathbf{U}}_{\text{MMSD-G}}$ exhibit poor performances due to the presence of outliers. Conversely, the estimators $\widehat{\mathbf{U}}_{\text{MLE-out}}$ and $\widehat{\mathbf{U}}_{\text{MMSD-OM-B}}$, show a better resistance to sample corruptions since they account for outliers in their derivation. The proposed estimator $\widehat{\mathbf{U}}_{\text{MMSD-OM-B}}$ reaches a better AFE than $\widehat{\mathbf{U}}_{\text{MLE-out}}$ thanks to the inclusion of the prior information.

Figure 4.8 displays the AFE in function of ONR where the data are generated as: $\mathbf{z}_k = \mathbf{s}_k + \mathbf{n}_k$, $\forall k \in [\![1, K-1]\!]$ and only the last sample contains an outlier as $\mathbf{z}_K = \mathbf{s}_K + \mathbf{o}_K + \mathbf{n}_K$. In this scenario, the same conclusion as in Figure 4.7 can be drawn. Interestingly, the proposed estimator $\widehat{\mathbf{U}}_{\text{MMSD-OM-B}}$ can exploit both the Bayesian prior information and the outlier modeling to resist high ONR contexts.

4.2 Applications to radar detection

The inclusion of a Bayesian prior can significantly improve the performance of an estimation process. However, the design of this prior depends on the considered application and comes from appropriate physical considerations/models on the system. In the following, we illustrate the

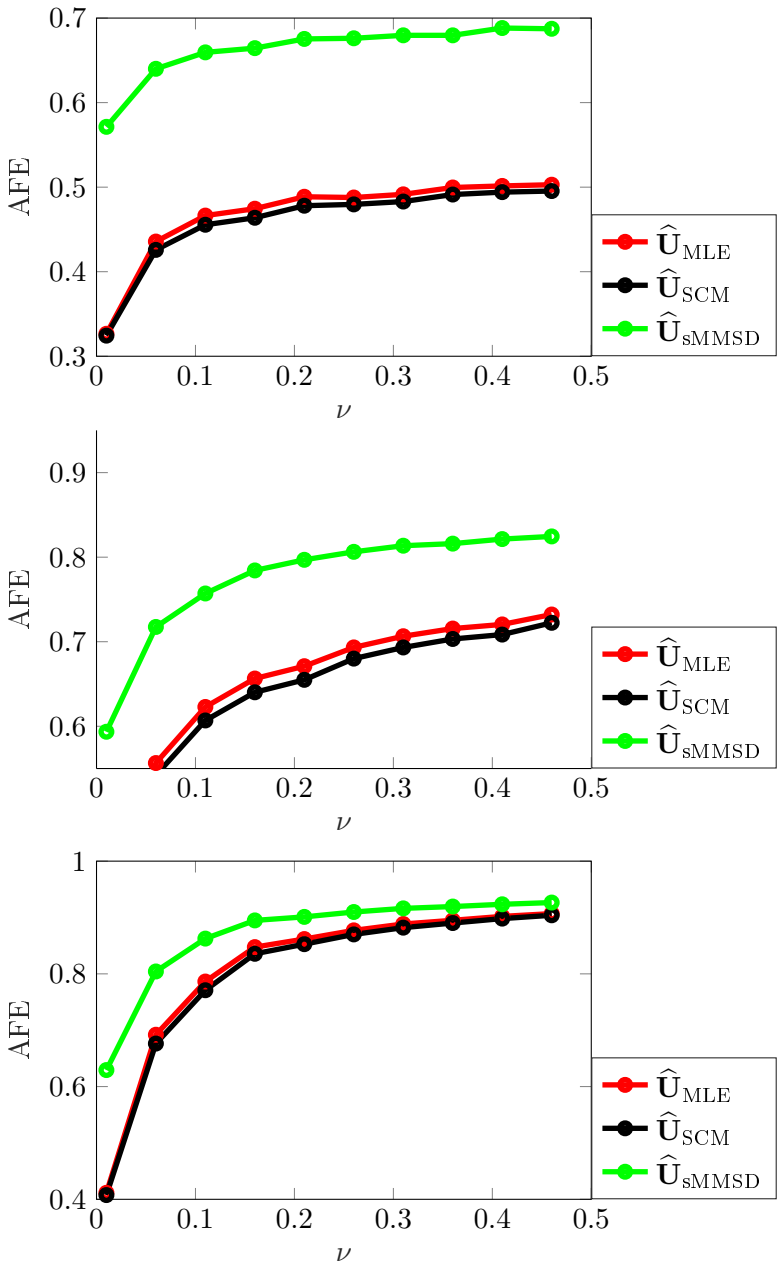


Figure 4.6: AFE w.r.t. ν for R=5, N =20, K = 30, $\mathbf{U} \sim \text{CIB}(\kappa_0, \bar{\mathbf{U}}\bar{\mathbf{U}}^H)$ from top to bottom SNR=0dB, SNR=5dB and SNR=10dB, for the simplified model, MMSD=sMMSD=MAP

practical use of the proposed methods for the context of radar detection. In this application, the clutter (response of the environment) lies in a low dimensional subspace that needs to be estimated to perform adaptive interference cancellation [48, 49]. We consider the approach that directly leverages the physical model of [70] in order to improve the performance of low-rank detectors [32, 33] at low sample support. We illustrate the interest of the proposed approach on a real data set.

In the following, we illustrate the interest of the proposed MMSD estimator for STAP on a real dataset provided by the French agency DGA/MI [79]. STAP is applied to airborne radar in

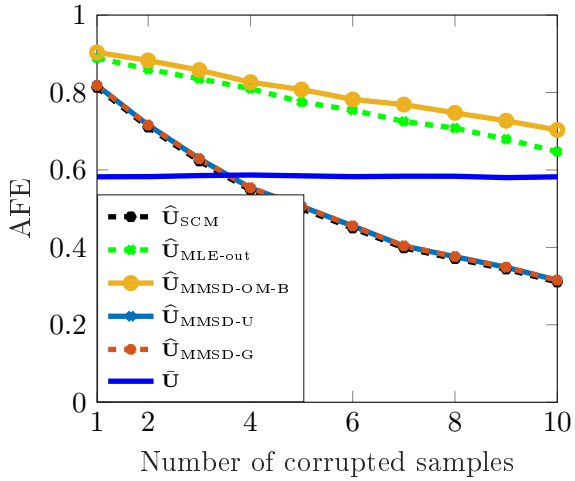


Figure 4.7: AFE w.r.t. number of corrupted samples for N=30, K=20, P=5, $\nu = \nu'$=1, ONR=SNR=15dB, $\mathbf{U} \sim \text{CIB}(\kappa, \bar{\mathbf{U}}\bar{\mathbf{U}}^H)$, $\kappa = 60$.

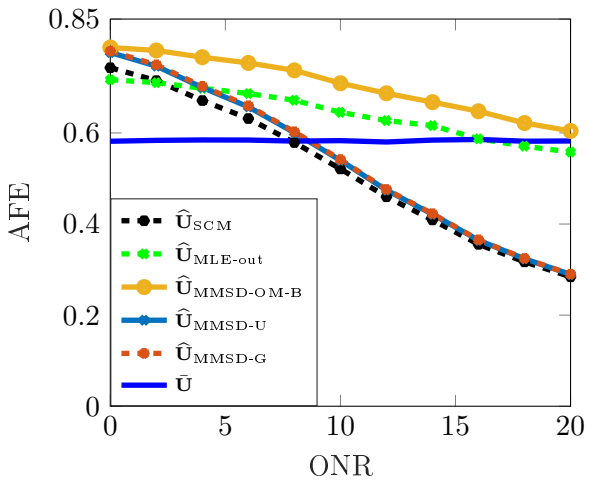


Figure 4.8: AFE w.r.t. ONR (in dB) for N=30, K=20, P=5, ν=ν'=1, SNR=10dB, $\mathbf{U} \sim \text{CIB}(\kappa, \bar{\mathbf{U}}\bar{\mathbf{U}}^H)$, $\kappa = 60$.

order to detect moving target [70]. We consider the hypothesis test

$$\begin{cases} \mathcal{H}_0 : \mathbf{z}_k = \mathbf{c}_k + \mathbf{n}_k, \ \forall k \in [\![0, K]\!] \\ \mathcal{H}_1 : \mathbf{z}_0 = \mathbf{d} + \mathbf{c}_0 + \mathbf{n}_0, \quad \mathbf{z}_k = \mathbf{c}_k + \mathbf{n}_k, \ \forall k \in [\![1, K]\!] \end{cases} \tag{4.2}$$

where the secondary data $\mathbf{z}_k \in \mathbb{C}^N$, $\forall k \in [\![1, K]\!]$ are assumed to be i.i.d.. The additive noise in each sample is the sum of clutter ground response $\mathbf{c}$ plus WGN $\mathbf{n}$. The tested cell $\mathbf{z}_0$ may potentially contain a moving target $\mathbf{d} = \alpha_0 \mathbf{p}$ where α_0 is the amplitude and $\mathbf{p}$ is the steering vector. We also consider the potential presence of an outlier in the secondary data by adding a sample $\mathbf{z}_{K+1}$ which contains the response of moving target and the additive noise, i.e., $\mathbf{z}_{K+1} = \mathbf{d} + \mathbf{c}_{K+1} + \mathbf{n}_{K+1}$. We note that the secondary data $\{\mathbf{z}_k\}$, which can be recasted in matrix form as $\mathbf{Z} = \mathbf{\Pi}_c \mathbf{C} + \mathbf{N}$ with $\mathbf{C} = [\mathbf{c}_1 | \dots | \mathbf{c}_K]$, $\mathbf{N} = [\mathbf{n}_1 | \dots | \mathbf{n}_K]$ and where $\mathbf{\Pi}_c = \mathbf{U}_c \mathbf{U}_c^H$ is the clutter

subspace projector (CSP) of known rank $R \ll N$ where the columns of $\mathbf{U}_c \in \mathcal{U}_N^R$ span the clutter subspace basis. In this context, the ground clutter rank R can be evaluated from the Brennan's rule [60].

We make use of the low rank adaptive matched filter (LR-ANMF) to assess the performance of various CSP estimators:

$$\widehat{w} = \frac{|\mathbf{d}^H \widehat{\mathbf{\Pi}}_c^{\perp} \mathbf{z}_0|^2}{|\mathbf{d}^H \widehat{\mathbf{\Pi}}_c^{\perp} \mathbf{d}||\mathbf{z}_0^H \widehat{\mathbf{\Pi}}_c^{\perp} \mathbf{z}_0|} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \delta \tag{4.3}$$

where $\widehat{\mathbf{\Pi}}_c^{\perp} = \mathbf{I} - \widehat{\mathbf{\Pi}}_c$ is an estimate of the orthogonal complement of the CSP. We compare the following detectors:

i) $\widehat{w}^{\text{SCM}}$ is the LR-ANMF built from the SVD of the SCM of the secondary data.

ii) $\widehat{w}^{\text{SFPE}}$ is the LR-ANMF built from the SVD of the SFPE [65] with regularization parameter γ selected manually to obtain the best results.

iii) $\widehat{w}^{\text{G-MUSIC}}$ is the LR-ANMF using [80,81] to estimate the quadratic forms associated to the CSP.

iv) $\widehat{w}^{\text{sMMSD}}$ denotes the LR-ANMF built from the MMSD estimator for the simplified model computed with Algorithm 3.

v) $\widehat{w}^{\text{MMSD-OM}}$ stands for the LR-ANMF built from the MMSD estimator in presence of outliers where the basis of interest follows CIB distribution.

vi) $\widehat{w}^{\text{MMSD-G}}$ is the LR-ANMF where the CSP is built from the MMSD estimator in presence of Gaussian sources [78].

Remark 4.2.1 *For the MMSD estimators, we leverage the physical model of [70], that allows to build a prior of the CSP basis* $\bar{\mathbf{U}}$ *from the SVD of the STAP covariance matrix model. The physical model of [70] allows to build a theoretical prior of the sources covariance matrix* $\bar{\mathbf{\Sigma}}$*. From this matrix, we construct the center of the prior subspace distribution as* $\mathcal{P}_R\left\{\bar{\mathbf{\Sigma}}\right\}$.

Since the concentration parameter κ is unknown, we display the output of the detectors (sMMSD, MMSD-G and MMSD-OM) for different values of the parameter β denoted by:

$$\beta = \frac{\kappa\sigma^2(1+\text{SNR})}{\text{SNR}} \tag{4.4}$$

Notice that the MMSD detectors present a trade off between the SCM/MLE detector ($\beta = 0$) and the prior only ($\beta \longrightarrow \infty$).

The tested cell contains the response of 10 moving targets to be detected in presence of ground clutter response (and possibly in presence of outliers in the secondary data). The accurate position of these targets is given in [79]. We note that the introduction of outliers is done by inserting the tested cell into the secondary data. We test the aforementioned detectors in the following scenarios

- Figure 4.9 displays the output of the detectors in the asymptotic regime ($K \gg N$) without outliers in the secondary data.
- Figure 4.10 displays the output of the detectors at low sample support ($K = R \ll N$) without outliers in the secondary data.

- Figure 4.11 displays the output of the detectors at low sample support ($K = 2R \ll N$) with outliers in the secondary data.

According to Figure 4.9, all of the detectors allow for target detection for $K > N$. Nevertheless, in non standard regimes, i.e., for small number of samples and/or in presence of outliers (Figure 4.10 and Figure 4.11), several detectors do not allow for target detection and/or exhibit high false alarm rates.

Notably, the SCM detector shows its limited detection performances with a relatively low false alarm rate due to the small number of samples K. In Figure 4.11, the latter does not detect any moving target and does not cancel the interference due to the outliers in the secondary data. These results can be explained by the fact that the SCM is known to be sensitive to missmodeling.

Thanks to its robustness properties, the SFPE detector exhibits better performances compared to the SCM one. However, for undersampled configuration ($K \ll N$), the SFPE subspace estimator suffers from being biased due to a higher regularization parameter γ. The G-MUSIC detector appears robust to outliers in terms of detection but leads to a high false alarm rate.

Conversely, the MMSD detectors (sMMSD, MMSD-G and MMSD-OM) allow for interference rejection and reliable target detection which illustrate the interest of introducing some prior information in an adaptive subspace estimation process. These detectors enjoy best of both SCM detector and prior one for appropriate values of β. Thus, as prospects of this work, it will be interesting to develop adaptive methods for the selection of the parameter β, e.g. from the results of [82]. According to Figure 4.11, the MMSD-OM detector outperforms the remaining MMSD detectors estimators. Indeed, the MMSD-OM detector allows for interference rejection when more than one sample is corrupted, thanks to a robust formulation that also includes prior knowledge.

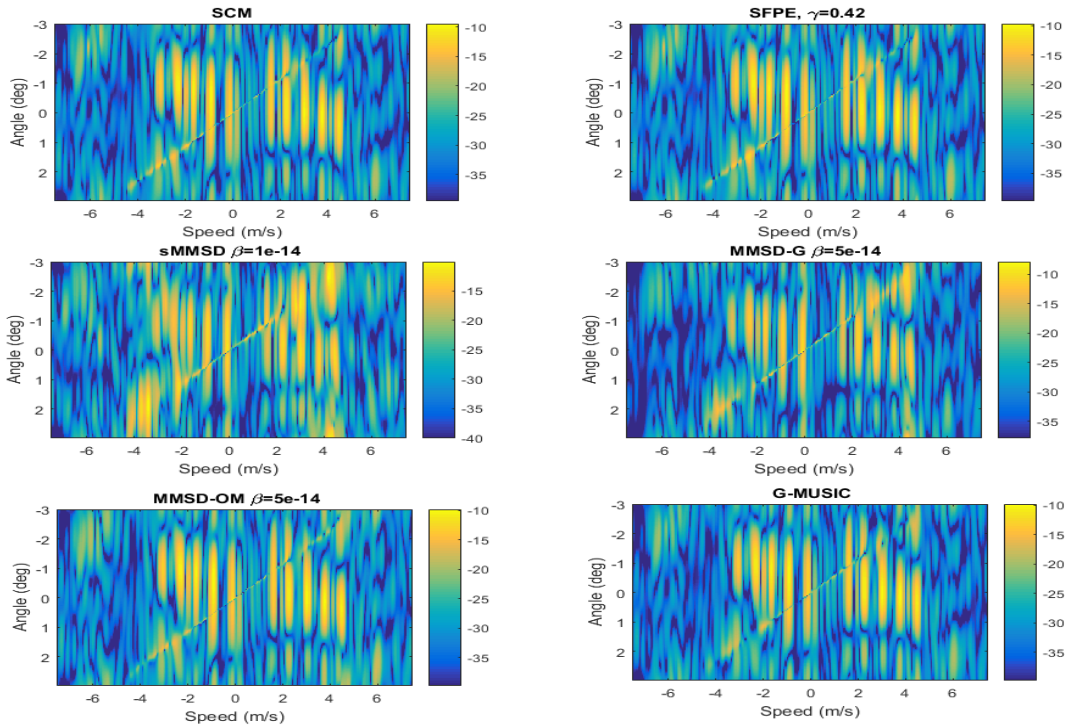


Figure 4.9: The output of detectors, K=397 ($K > N$), $R = 46$, $N = 256$

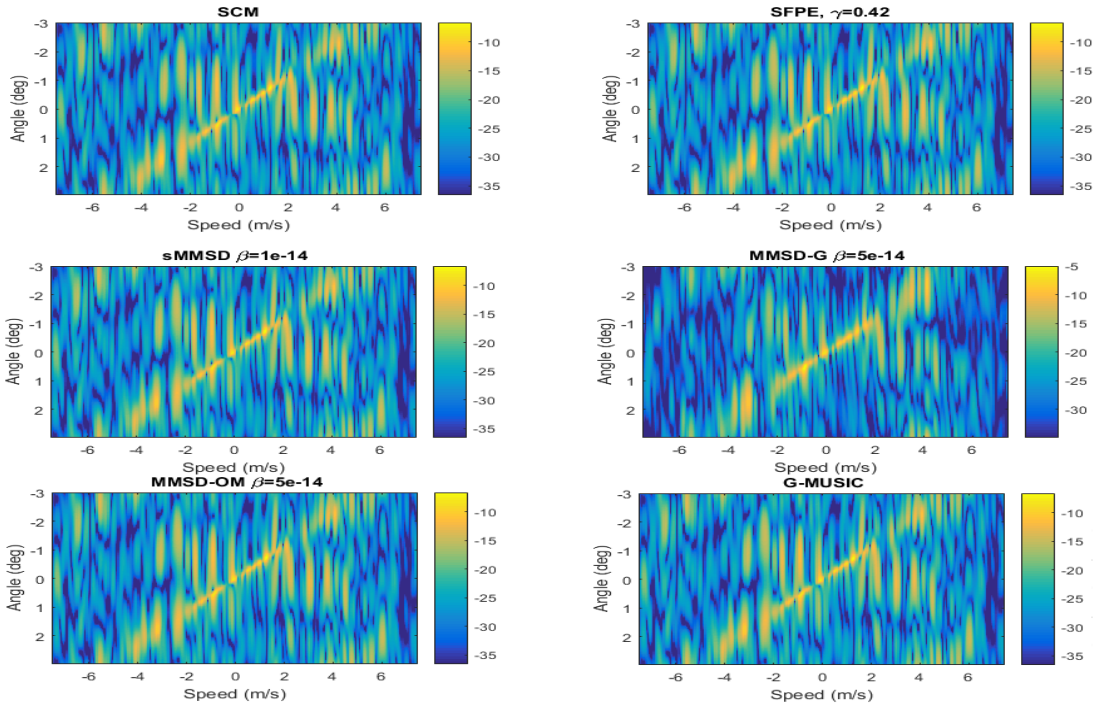


Figure 4.10: The output of detectors, $K = R$ with $K \ll N$, $R = 46$, $N = 256$

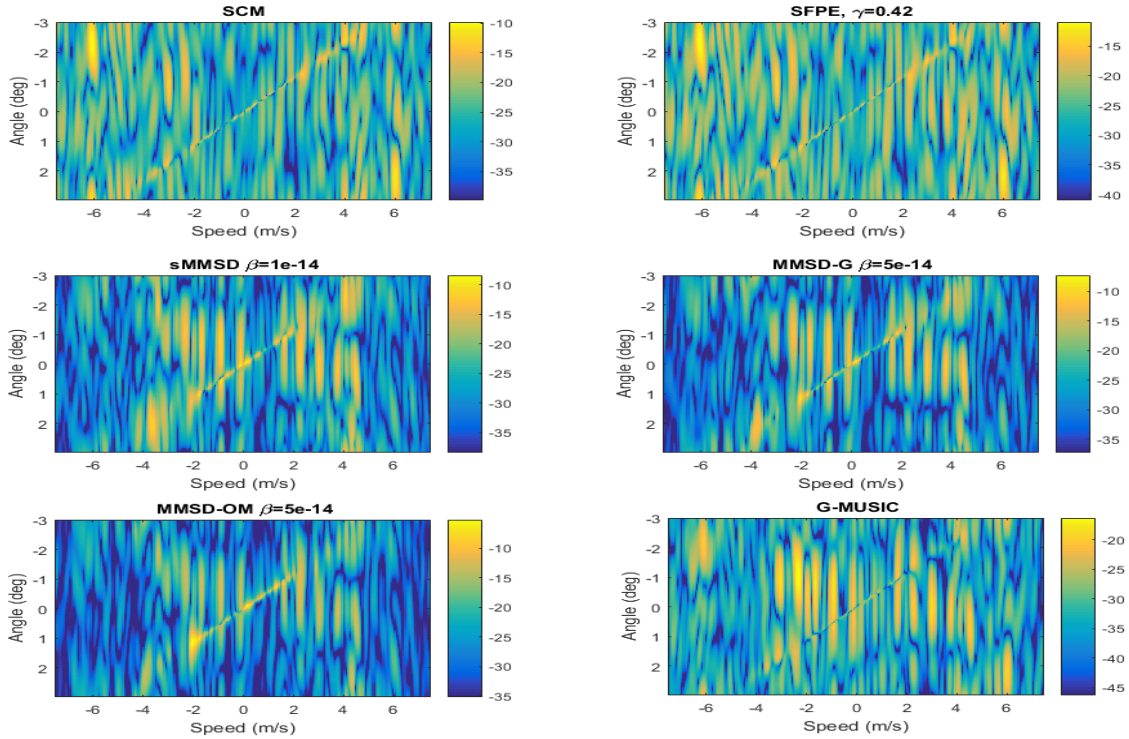


Figure 4.11: The output of detectors in presence of outlier, $K = 2R$ with $K \ll N$, $R = 46$, $N = 256$

Chapter 5

Conclusion and perspectives

In this Part, we considered a Bayesian approach for subspace estimation. First, we cited some standard subspace methods including both deterministic and Bayesian ones. Second, we formulated the MAP and the MMSD estimators of the signal subspace in the context of CG distributed sources and CGBL distributed subspace. Then, in order to develop Bayesian subspace estimators that are robust to presence of impulsive and/or heavy-tailed noise, we derived both the MMSD and the MAP subspace estimators in the context of CG sources embedded in heterogeneous noise (CG outliers plus an additive WGN). Afterwards, simulations illustrated the interest of the proposed approach in critical regimes (low SNR and/or low sample size). Finally, we investigated the practical use of the proposed methods for the context of radar detection where we constructed the prior from a physical model.

Among the possible future prospects which naturally follow this work, we can mention the following points:

- It is interesting to develop adaptive methods for the selection of the parameters matrices of the CGBL distribution. For instance, the estimation of the real Bingham and real Langevin distributions parameters is well investigated theoretically in [82].
- The application of the subspace estimation methods developed in this thesis can be applied to other contexts (e.g. the Low rank filtering for communications can be envisioned).
- In most cases, the rank of principal subspace is unknown and must be estimated in advance [55, 56, 60]. It would therefore be interesting to investigate the problem of rank estimation for the considered context.
- It would seem particularly useful to derive the GLRT of detectors in the context structured covariance matrices where we consider a prior for the principal subspace e.g. CGBL, CL and Bingham distributions.

Part III

Change detection in low rank covariance matrices

Chapter 1

State of the art on the covariance matrices change detection

1.1 Introduction and motivations

Statistical testing of covariance matrix equality (or proportionality) has received increasing interest in the context of signal and image processing [5]. Indeed, this well established hypothesis test has been successfully studied and applied to change/anomaly detection and classification for a plethora of applications as the SAR image processing. Notably, equality testing has been proposed for change detection in SAR image time series [83–90]. A clear overview and statistical analysis of this topic is proposed in [5]. The extension to proportionality testing has been proposed in [91, 92]. In this scope, elliptical noise modeling has also been studied to develop robust covariance matrix-based image processing in [93–96].

In this part of the thesis, we propose new statistical tests for the context of structured covariance matrices. Indeed, the covariance matrix of radar measurements usually exhibit inherent structures. In a very general case, the samples can be modeled as in the Part II of this thesis, i.e., a realizations of a low rank signal component plus WGN. This leads to a covariance matrix structured as

$$\mathbf{\Sigma} = \mathbf{\Sigma}_R + \sigma^2 \mathbf{I} \tag{1.1}$$

where $\mathbf{\Sigma}_R$ is the low rank signal covariance matrix. Taking this prior knowledge in the detection process offers several advantages

i) Introducing relevant prior information (here, the low rank covariance matrix structure) in the model improves detection performances.

ii) Low rank structured models allow to deal with low sample support issues, since less samples are required to estimate the covariance matrix. Especially, it allows to perform tests when the SCM is not invertible, while this condition is restrictive for standard tests.

iii) The considered formulation can go beyond equality and proportionality testing. For example, consider a local power fluctuations of ground response modeled as

$$\mathbf{\Sigma}_i = \tau_i \mathbf{\Sigma}_R + \sigma^2 \mathbf{I} \tag{1.2}$$

with $\tau_i \in \mathbb{R}^+$ and i denotes the index of a homogeneous set. For $i \neq j$, such model leads to

$$\mathbf{\Sigma}_i \neq \mathbf{\Sigma}_j$$
$$\mathbf{\Sigma}_i \not\propto \mathbf{\Sigma}_j$$

If the goal is to detect a signal anomaly in $\mathbf{\Sigma}_R$, detectors based on covariance matrix equality/proportionality testing may lead to excessive false alarms.

Specifically, we propose in the following two novel GLRTs that account for the considered low rank structure, with assumed known rank. Formally, these detectors re-express the equality and proportionality tests, but only on the signal low rank component $\mathbf{\Sigma}_R$ of the total covariance matrix. Then, we are interested in designing a third detection test that is only sensitive to a signal subspace variation. The derivation of both the second and the third tests require to solve some non-trivial optimization problems, for which we tailor appropriate MM algorithms.

Finally, the performance of the proposed detectors are illustrated on simulated data, where they exhibit interesting properties. Furthermore, the benefits of the proposed methods are also illustrated for a change detection application on a UAVSAR dataset (courtesy of NASA/JPL-Caltech, https://uavsar.jpl.nasa.gov).

1.2 State of the art

In this Section, we present two standard GLRTs used for testing the similarity of covariance matrices. The first test is for the equality of covariance matrices, while the second one is derived for the proportionality testing between covariance matrices in the complex-valued Gaussian case. We consider $I+1$ independent sets of samples $\mathbf{z}_k^i \in \mathbb{C}^N$, $i \in [\![0, I]\!]$, $k \in [\![1, K]\!]$, with K_i samples for each set i. The samples $\mathbf{z}_k^i \sim \mathcal{CN}(0, \mathbf{\Sigma}_i)$ are assumed i.i.d. and their corresponding SCM are denoted by $\mathbf{S}_i$

$$\mathbf{S}_i = \frac{1}{K_i} \sum_{k=1}^{K_i} \mathbf{z}_k^i \mathbf{z}_k^{iH} \tag{1.3}$$

1.2.1 Equality testing

The general problem of testing covariance matrix equality [97] in the complex-valued Gaussian case is analyzed in e.g. [5]. The hypothesis test reads as

$$\begin{cases} \mathcal{H}_0 : \ \mathbf{\Sigma}_0 = \mathbf{\Sigma}, \ \mathbf{\Sigma}_i = \mathbf{\Sigma} \ \forall i \in [\![1, I]\!] \\ \mathcal{H}_1 : \ \mathbf{\Sigma}_0 \neq \mathbf{\Sigma}, \ \mathbf{\Sigma}_i = \mathbf{\Sigma} \ \forall i \in [\![1, I]\!] \end{cases} \tag{1.4}$$

The GLRT for this hypothesis test, denoted t_{E}, reads

$$\frac{|\hat{\mathbf{\Sigma}}_{\mathcal{H}_0}|}{\left(|\hat{\mathbf{\Sigma}}^0_{\mathcal{H}_1}|^{\rho_0} |\hat{\mathbf{\Sigma}}^\star_{\mathcal{H}_1}|^{\rho_\star}\right)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \delta^{\mathrm{E}}_{\mathrm{glr}} \tag{1.5}$$

with the quantities

$$K = \sum_{i=0}^{I} K_i$$

$$K_\star = K - K_0 = \sum_{i=1}^{I} K_i$$

$$\rho_0 = K_0/K$$

$$\rho_\star = K_\star/K$$

and the SCMs

$$\begin{aligned}\hat{\mathbf{\Sigma}}_{\mathcal{H}_0} &= \sum_{i=0}^{I} \mathbf{S}_i/K \\ \hat{\mathbf{\Sigma}}^0_{\mathcal{H}_1} &= \mathbf{S}_0/K_0 \\ \hat{\mathbf{\Sigma}}^\star_{\mathcal{H}_1} &= \sum_{i=1}^{I} \mathbf{S}_i/K_\star\end{aligned} \tag{1.6}$$

1.2.2 Proportionality testing

The general problem of testing covariance matrix proportionality [98] in the complex-valued Gaussian case is analyzed in e.g. [91]. The hypothesis test is

$$\begin{cases}\mathcal{H}_0 : \ \mathbf{\Sigma}_0 = \beta_0\mathbf{\Sigma}, \ \mathbf{\Sigma}_i = \beta_i\mathbf{\Sigma} \ \forall i \in [\![1, I]\!] \\ \mathcal{H}_1 : \ \mathbf{\Sigma}_0 \neq \beta_0\mathbf{\Sigma}, \ \mathbf{\Sigma}_i = \beta_i\mathbf{\Sigma} \ \forall i \in [\![1, I]\!]\end{cases} \tag{1.7}$$

The GLRT on proportionality, denoted t_{P}, is given as

$$\left(\frac{|\hat{\beta}^0_{\mathcal{H}_0}\hat{\mathbf{\Sigma}}^{\mathrm{gfp}}_{\mathcal{H}_0}|}{|\hat{\mathbf{\Sigma}}^0_{\mathcal{H}_1}|}\right)^{K_0} \prod_{i=1}^{I}\left(\frac{|\hat{\beta}^i_{\mathcal{H}_0}\hat{\mathbf{\Sigma}}^{\mathrm{gfp}}_{\mathcal{H}_0}|}{|\hat{\beta}^i_{\mathcal{H}_1}\hat{\mathbf{\Sigma}}^{\mathrm{gfp}}_{\mathcal{H}_1}|}\right)^{K_i} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \delta^{\mathrm{P}}_{\mathrm{glr}}, \tag{1.8}$$

where

i) $\{\hat{\beta}^i_{\mathcal{H}_0}\}$ and $\hat{\mathbf{\Sigma}}^{\mathrm{gfp}}_{\mathcal{H}_0}$ are proportionality coefficients and shape matrix estimated with the generalized fixed point estimator (GFPE) [92] applied on the set $\{\mathbf{S}_i\}_{i\in[\![0,I]\!]}$

$$\begin{aligned}\hat{\mathbf{\Sigma}}^{\mathrm{gfp}}_{\mathcal{H}_0} &= \frac{N}{K}\sum_{i=0}^{I}\frac{K_i}{\mathrm{Tr}\{\mathbf{S}_i\hat{\mathbf{\Sigma}}^{\mathrm{gfp}^{-1}}_{\mathcal{H}_0}\}}\mathbf{S}_i \\ \hat{\beta}^i_{\mathcal{H}_0} &= \frac{1}{N}\mathrm{Tr}\{\mathbf{S}_i\hat{\mathbf{\Sigma}}^{\mathrm{gfp}^{-1}}_{\mathcal{H}_0}\}\end{aligned} \tag{1.9}$$

ii) $\{\hat{\beta}^i_{\mathcal{H}_1}\}$ and $\hat{\mathbf{\Sigma}}^{\mathrm{gfp}}_{\mathcal{H}_1}$ are obtained from the GFPE on the set $\{\mathbf{S}_i\}_{i\in[\![1,I]\!]}$

$$\begin{aligned}\hat{\mathbf{\Sigma}}^{\mathrm{gfp}}_{\mathcal{H}_1} &= \frac{N}{K}\sum_{i=1}^{I}\frac{K_i}{\mathrm{Tr}\{\mathbf{S}_i\hat{\mathbf{\Sigma}}^{\mathrm{gfp}^{-1}}_{\mathcal{H}_1}\}}\mathbf{S}_i \\ \hat{\beta}^i_{\mathcal{H}_1} &= \frac{1}{N}\mathrm{Tr}\{\mathbf{S}_i\hat{\mathbf{\Sigma}}^{\mathrm{gfp}^{-1}}_{\mathcal{H}_1}\}\end{aligned} \tag{1.10}$$

iii) $\hat{\mathbf{\Sigma}}^0_{\mathcal{H}_1}$ is the SCM as

$$\hat{\mathbf{\Sigma}}^0_{\mathcal{H}_1} = \mathbf{S}_0/K_0 \tag{1.11}$$

Chapter 2

The proposed detectors for structured models

In this chapter, we propose detectors in the context of low rank structured covariance matrices. The newly proposed detectors are only sensitive to the low rank principal component. To this end, we derive the GLRT to infer on the equality/proportionality of the low rank signal component and on the equality of the principal subspace of signal of interest.

2.1 Structured covariance matrices general model

We consider $I+1$ independent sets of samples $\mathbf{z}_k^i \in \mathbb{C}^N$, $i \in [\![0, I]\!]$, $k \in [\![1, K_i]\!]$, with K_i samples for each set i. The samples $\{\mathbf{z}_k^i\}$ are assumed i.i.d. w.r.t. the following model:

$$\mathbf{z}_k^i = \mathbf{s}_k^i + \mathbf{n}_k^i \tag{2.1}$$

• The signal $\mathbf{s}_k^i \sim \mathcal{CN}(0, \mathbf{\Sigma}_R^i)$, in which the unknown scatter matrix, of rank $R \ll N$, is denoted as $\mathbf{\Sigma}_R^i = \mathbf{U}_i \mathbf{R}_i \mathbf{U}_i^H$ where $\mathbf{U}_i \in \mathcal{U}_N^R$ represents a signal subspace basis and $\mathbf{R}_i \in \mathbb{C}^{R \times R}$ is the signal covariance matrix in this space. The rank R is assumed pre-established. For example, such signal may model the ground response when considering correlation and/or power fluctuations (w.r.t. i).
• The WGN $\mathbf{n}_k^i \sim \mathcal{CN}(0, \sigma^2 \mathbf{I})$ represents the contribution of thermal noise.
Eventually, the samples are drawn as $\mathbf{z}_k^i \sim \mathcal{CN}(0, \mathbf{\Sigma}_i)$, in which, $\mathbf{\Sigma}_i$ has the following expression:

$$\mathbf{\Sigma}_i = \mathbf{U}_i \mathbf{R}_i \mathbf{U}_i^H + \sigma^2 \mathbf{I} \overset{\Delta}{=} \mathbf{\Sigma}_R^i + \sigma^2 \mathbf{I} \tag{2.2}$$

where σ^2 is assumed pre-estimated or known. Consequently, the likelihood of the data set $\{\mathbf{z}_k^i\}$ reads as:

$$\mathcal{L}(\{\mathbf{z}_k^i\}|\boldsymbol{\theta}) = \prod_{i=0}^{I} \frac{\text{etr}\{-\mathbf{S}_i \mathbf{\Sigma}_i^{-1}(\boldsymbol{\theta})\}}{|\ \mathbf{\Sigma}_i(\boldsymbol{\theta})\ |^{K_i}} \tag{2.3}$$

where $\mathbf{S}_i = \sum_{k=1}^{K_i} \mathbf{z}_k^i \mathbf{z}_k^{iH}$ and $\boldsymbol{\theta}$ is an appropriate covariance matrix parameterization of the set $\{\mathbf{\Sigma}_i\}$. For the general model in (2.2)-(2.3), we turn to the problem of testing whether the covariance matrix of the sample set under test $i = 0$ shares some common properties with the secondary sets $i \in [\![1, I]\!]$. These properties are related to the parameters of the decomposition in (2.2) (i.e. $\mathbf{U}_i$ and $\mathbf{R}_i$) and will be specified depending on the proposed test. This problem is relevant to detect, e.g., a local anomaly in the patch,i.e., small rectangular image regions typically squares, w.r.t. adjacent patches, or a temporal change in the last sample of a time-series. We derive

in the following the GLRT of the three detectors, namely the low rank proportionality/equality and the low rank subspace. Some of these GLRTs involves non-trivial optimization problem. In order to evaluate them, we design in the following block-coordinate descent algorithms. As in Part II, we make use of the block MM algorithm for these problems.

2.2 Low rank equality detector

In this section, we develop a GLRT that is sensitive to a variation of any parameter of the low rank signal covariance matrix in the set $i = 0$. Thus, for the covariance matrix model in (2.2), consider the following hypothesis test

$$\begin{cases} \mathcal{H}_0 : & \left| \boldsymbol{\Sigma}_R^i = \boldsymbol{\Sigma}_R \ \forall i \in [\![0, I]\!] \right. \\ \mathcal{H}_1 : & \left| \begin{array}{l} \boldsymbol{\Sigma}_R^i = \boldsymbol{\Sigma}_R \ \forall i \in [\![1, I]\!] \\ \boldsymbol{\Sigma}_R^0 \neq \boldsymbol{\Sigma}_R, \end{array} \right. \end{cases} \tag{2.4}$$

that reads almost identical to a standard equality testing, except that the covariance matrices belong to $\mathcal{H}_R^{\text{LR}}$. Hence, the hypothesis test can be recasted as

$$\begin{cases} \mathcal{H}_0 : & \left| \boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_{R,\mathcal{H}_0} + \sigma^2 \mathbf{I} \ \triangleq \ \boldsymbol{\Sigma}_{\mathcal{H}_0} \in \mathcal{H}_R^{\text{LR}}, \ \forall i \in [\![0, I]\!] \right. \\ \mathcal{H}_1 : & \left| \begin{array}{l} \boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_{R,\mathcal{H}_1}^0 + \sigma^2 \mathbf{I} \ \triangleq \ \boldsymbol{\Sigma}_{\mathcal{H}_1}^0 \in \mathcal{H}_R^{\text{LR}} \\ \boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_{R,\mathcal{H}_1}^\star + \sigma^2 \mathbf{I} \ \triangleq \ \boldsymbol{\Sigma}_{\mathcal{H}_1}^\star \in \mathcal{H}_R^{\text{LR}}, \ \forall i \in [\![1, I]\!] \end{array} \right. \end{cases}$$

The expression of the GLRT is therefore

$$\frac{\max_{\boldsymbol{\theta}_{\mathcal{H}_1}^{\text{lrE}}} \mathcal{L}\left(\{\mathbf{z}_k^i\} | \mathcal{H}_1, \boldsymbol{\theta}_{\mathcal{H}_1}^{\text{lrE}}\right)}{\max_{\boldsymbol{\theta}_{\mathcal{H}_0}^{\text{lrE}}} \mathcal{L}(\{\mathbf{z}_k^i\} | \mathcal{H}_0, \boldsymbol{\theta}_{\mathcal{H}_0}^{\text{lrE}})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \delta_{\text{glr}}^{\text{lrE}}, \tag{2.5}$$

with sets

$$\begin{aligned} \boldsymbol{\theta}_{\mathcal{H}_1}^{\text{lrE}} &= \left\{ \boldsymbol{\Sigma}_{\mathcal{H}_1}^0, \boldsymbol{\Sigma}_{\mathcal{H}_1}^\star \right\} \\ \boldsymbol{\theta}_{\mathcal{H}_0}^{\text{lrE}} &= \{ \boldsymbol{\Sigma}_{\mathcal{H}_0} \} \end{aligned}$$

In the context of Gaussian data, the MLE of low rank structured covariance matrix is obtained by thresholding the eigenvalues of the SCM [99] with the operator $\mathcal{T}_R$ defined in Definition 3.4.4 in Part I where $\psi = \sigma^2$. We recall what $\mathcal{T}_R$ stands for:

$$\begin{array}{rccc} \mathcal{P}_{\text{Proc}} : & \mathbb{C}^{N \times R} & \longrightarrow & \mathcal{U}_N^R \\ & \mathbf{Z} \overset{\text{TSVD}}{=} \mathbf{U}\mathbf{D}\mathbf{V}^H & \longmapsto & \mathcal{P}_{\text{Proc}}\{\mathbf{Z}\} = \mathbf{U}\mathbf{V}^H \end{array} \tag{2.6}$$

$$\begin{array}{rccc} \mathcal{T}_R : & \mathcal{H}_N^+ & \longrightarrow & \mathcal{H}_N^+ \\ & \boldsymbol{\Sigma} \overset{\text{SVD}}{=} [\mathbf{U}|\mathbf{U}^\perp] \begin{bmatrix} \boldsymbol{\Lambda}_R & 0 \\ 0 & \boldsymbol{\Lambda}_{N-R} \end{bmatrix} [\mathbf{U}|\mathbf{U}^\perp]^H & \longmapsto & \mathcal{T}_R\{\boldsymbol{\Sigma}\} \overset{\text{SVD}}{=} [\mathbf{U}|\mathbf{U}^\perp] \begin{bmatrix} \tilde{\boldsymbol{\Lambda}}_R & 0 \\ 0 & \sigma^2 \mathbf{I}_{N-R} \end{bmatrix} [\mathbf{U}|\mathbf{U}^\perp]^H \end{array} \tag{2.7}$$

with

$$[\tilde{\boldsymbol{\Lambda}}]_{i,i} = \begin{cases} \max([\boldsymbol{\Lambda}]_{i,i}, \sigma^2), & i \leq R \\ \sigma^2, & i > R \end{cases} \tag{2.8}$$

Therefore, it can be easily shown that

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{\mathcal{H}_0}^{\text{lrE}} &= \left\{ \mathcal{T}_R\{\hat{\boldsymbol{\Sigma}}_{\mathcal{H}_0}\} \right\} \\ \hat{\boldsymbol{\theta}}_{\mathcal{H}_1}^{\text{lrE}} &= \left\{ \mathcal{T}_R\{\hat{\boldsymbol{\Sigma}}_{\mathcal{H}_1}^0\}, \mathcal{T}_R\{\hat{\boldsymbol{\Sigma}}_{\mathcal{H}_1}^\star\} \right\} \end{aligned}$$

with $\hat{\mathbf{\Sigma}}_{\mathcal{H}_0}$, $\hat{\mathbf{\Sigma}}^0_{\mathcal{H}_1}$ and $\hat{\mathbf{\Sigma}}^\star_{\mathcal{H}_1}$ defined in (1.6). Finally, the GLRT for testing the equality of low rank structured matrices, denoted $t^{\mathrm{LR}}_{\mathrm{E}}$, reads as

$$\frac{\mathcal{L}(\{\mathbf{z}^i_k\}|\mathcal{H}_1,\hat{\boldsymbol{\theta}}^{\mathrm{lrE}}_{\mathcal{H}_1})}{\mathcal{L}(\{\mathbf{z}^i_k\}|\mathcal{H}_0,\hat{\boldsymbol{\theta}}^{\mathrm{lrE}}_{\mathcal{H}_0})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \delta^{\mathrm{lrE}}_{\mathrm{glr}} \tag{2.9}$$

2.3 Low rank proportionality detector

In this section, we derive a GLRT to infer on the proportionality of the low rank signal component of the covariance matrix. Notice that this test differs from strict proportionality testing. Indeed, scaling fluctuations should only apply on the low rank part of the signal covariance matrix, and not to the identity, related to the thermal noise. For the covariance matrix model in (2.2), this leads to the following hypothesis test

$$\begin{cases} \mathcal{H}_0 : & \left|\mathbf{\Sigma}^i_R = \tau_i \mathbf{\Sigma}_R \ \forall i \in [\![0,I]\!]\right. \\ \mathcal{H}_1 : & \left|\begin{array}{l}\mathbf{\Sigma}^i_R = \tau_i \mathbf{\Sigma}_R \ \forall i \in [\![1,I]\!] \\ \mathbf{\Sigma}^0_R \neq \tau_0 \mathbf{\Sigma}_R\end{array}\right. \end{cases} \tag{2.10}$$

which reads as signals sharing the same covariance matrix structure, but with fluctuating power τ_i w.r.t. sample set i. The above test can be recasted as

$$\begin{cases} \mathcal{H}_0 : \left|\mathbf{\Sigma}_i = \tau_i \mathbf{U}_{\mathcal{H}_0}\mathbf{\Lambda}_{\mathcal{H}_0}(\mathbf{U}_{\mathcal{H}_0})^H + \sigma^2\mathbf{I},\ \forall i \in [\![0,I]\!]\right. \\ \mathcal{H}_1 : \left|\begin{array}{l}\mathbf{\Sigma}_0 = \ \mathbf{U}^0_{\mathcal{H}_1}\mathbf{R}^0_{\mathcal{H}_1}(\mathbf{U}^0_{\mathcal{H}_1})^H + \sigma^2\mathbf{I} \\ \mathbf{\Sigma}_i = \tau_i \mathbf{U}^\star_{\mathcal{H}_1}\mathbf{\Lambda}^\star_{\mathcal{H}_1}(\mathbf{U}^\star_{\mathcal{H}_1})^H + \sigma^2\mathbf{I},\ \forall i \in [\![1,I]\!]\end{array}\right. \end{cases} \tag{2.11}$$

where $\mathbf{R}^i_{\mathcal{H}_0} = \tau_i \mathbf{\Lambda}_{\mathcal{H}_0},\ \forall i \in [\![0,I]\!]$ and $\mathbf{R}^i_{\mathcal{H}_1} = \tau_i \mathbf{\Lambda}^\star_{\mathcal{H}_1},\ \forall i \in [\![1,I]\!]$ (that are diagonal matrices). The expression of the corresponding GLRT, denoted $t^{\mathrm{LR}}_{\mathrm{P}}$ is given by

$$\frac{\max_{\boldsymbol{\theta}^{\mathrm{lrP}}_{\mathcal{H}_1}} \mathcal{L}\left(\{\mathbf{z}^i_k\}|\mathcal{H}_1,\boldsymbol{\theta}^{\mathrm{lrP}}_{\mathcal{H}_1}\right)}{\max_{\boldsymbol{\theta}^{\mathrm{lrP}}_{\mathcal{H}_0}} \mathcal{L}(\{\mathbf{z}^i_k\}|\mathcal{H}_0,\boldsymbol{\theta}^{\mathrm{lrP}}_{\mathcal{H}_0})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \delta^{\mathrm{lrP}}_{\mathrm{glr}}, \tag{2.12}$$

with sets of parameters

$$\begin{aligned} \boldsymbol{\theta}^{\mathrm{lrP}}_{\mathcal{H}_0} &= \left\{\{\tau_i\}_{i\in[\![0,I]\!]}, \mathbf{U}_{\mathcal{H}_0}, \mathbf{\Lambda}_{\mathcal{H}_0}\right\}, \\ \boldsymbol{\theta}^{\mathrm{lrP}}_{\mathcal{H}_1} &= \left\{\{\tau_i\}_{i\in[\![1,I]\!]}, \mathbf{U}^\star_{\mathcal{H}_1}, \mathbf{\Lambda}^\star_{\mathcal{H}_1}, \mathbf{U}^0_{\mathcal{H}_1}, \mathbf{R}^0_{\mathcal{H}_1}\right\}. \end{aligned} \tag{2.13}$$

It is clear that the maximization of the likelihood function is not trivial, due notably to the unitary constraints on the eigenvectors. To overcome this issue, we propose the use of the block-MM algorithm and from the Part II of this thesis. The MM algorithm is described in Section B.1 in Appendix B and useful surrogates for our considered problem are derived in Section B.3 in Appendix B.

Eventually, these algorithms allow to evaluate the MLEs $\hat{\boldsymbol{\theta}}^{\mathrm{lrP}}_{\mathcal{H}_0}$ and $\hat{\boldsymbol{\theta}}^{\mathrm{lrP}}_{\mathcal{H}_1}$ and the GLRT as

$$\frac{\mathcal{L}(\{\mathbf{z}^i_k\}|\mathcal{H}_1,\hat{\boldsymbol{\theta}}^{\mathrm{lrP}}_{\mathcal{H}_1})}{\mathcal{L}(\{\mathbf{z}^i_k\}|\mathcal{H}_0,\hat{\boldsymbol{\theta}}^{\mathrm{lrP}}_{\mathcal{H}_0})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \delta^{\mathrm{lrP}}_{\mathrm{glr}} \tag{2.14}$$

2.3.1 MLE of $\boldsymbol{\theta}_{\mathcal{H}_0}^{\text{lrP}}$ under $\mathcal{H}_0$

Under $\mathcal{H}_0$, we have the following optimization problem

$$\begin{array}{ll} \underset{\boldsymbol{\theta}_{\mathcal{H}_0}^{\text{lrP}}}{\text{maximize}} & \mathcal{L}(\{\mathbf{z}_k^i\}|\mathcal{H}_0, \boldsymbol{\theta}_{\mathcal{H}_0}^{\text{lrP}}) \\ \text{subject to} & \tau_i \geq 0,\ \forall i \in [\![0, I]\!] \\ & [\boldsymbol{\Lambda}_{\mathcal{H}_0}]_{r,r} \geq 0,\ \forall r \in [\![1, R]\!] \\ & (\mathbf{U}_{\mathcal{H}_0})^H \mathbf{U}_{\mathcal{H}_0} = \mathbf{I}. \end{array} \tag{2.15}$$

This problem is equivalent to maximizing the negative log-likelihood as

$$\begin{array}{ll} \underset{\boldsymbol{\theta}_{\mathcal{H}_0}^{\text{lrP}}}{\text{minimize}} & \sum_{i=0}^{I} \left[K_i \ln(|\boldsymbol{\Sigma}_i|) + \text{Tr}\left\{\mathbf{S}_i \boldsymbol{\Sigma}_{\boldsymbol{i}}^{-1}\right\}\right] \\ \text{subject to} & \boldsymbol{\Sigma}_i = \tau_i \mathbf{U}_{\mathcal{H}_0} \boldsymbol{\Lambda}_{\mathcal{H}_0} (\mathbf{U}_{\mathcal{H}_0})^H + \sigma^2 \mathbf{I},\ \forall i \in [\![0, I]\!] \\ & \tau_i \geq 0,\ \forall i \in [\![0, I]\!] \\ & [\boldsymbol{\Lambda}_{\mathcal{H}_0}]_{r,r} \geq 0,\ \forall r \in [\![1, R]\!] \\ & (\mathbf{U}_{\mathcal{H}_0})^H \mathbf{U}_{\mathcal{H}_0} = \mathbf{I}, \end{array} \tag{2.16}$$

for which we derive MM updates in the following. The corresponding algorithm is summed up in the box Algorithm 7.

Algorithm 7: MM algorithm to compute $\hat{\boldsymbol{\theta}}_{\mathcal{H}_0}^{\text{lrP}}$ for t_{P}^{LR}

Input: $\{\mathbf{S}_i, K_i\}$ for $i \in [\![0, I]\!]$, R, and σ^2.
Repeat
$t \leftarrow t + 1$
Update $\{\tau_i^{(t)}\}$, $\forall i \in [\![0, I]\!]$ with (B.15)
Update $\{\lambda_r^{(t)}\}$, $\forall r \in [\![1, R]\!]$ with (2.21)
Update $\hat{\mathbf{U}}_{\mathcal{H}_0}^{(t)}$ with (2.24)
Until Some convergence criterion is met.
Output: $\hat{\boldsymbol{\theta}}_{\mathcal{H}_0}^{\text{lrP}} = \left\{\{\hat{\tau}_i\}_{i \in [\![0,I]\!]},\ \hat{\mathbf{U}}_{\mathcal{H}_0},\ \hat{\boldsymbol{\Lambda}}_{\mathcal{H}_0}\right\}$

2.3.1.1 Update $\{\tau_i\}$

First, remark that $\boldsymbol{\Sigma}_i^{-1}$ can be expressed thanks to the matrix inversion lemma as

$$\boldsymbol{\Sigma}_i^{-1} = (\tau_i \mathbf{V}_{\mathcal{H}_0} \boldsymbol{\Lambda}_{\mathcal{H}_0} (\mathbf{V}_{\mathcal{H}_0})^H + \sigma^2 \mathbf{I})^{-1} = \sigma^{-2} \mathbf{I} - \mathbf{V}_{\mathcal{H}_0} \boldsymbol{\Gamma}_i (\mathbf{V}_{\mathcal{H}_0})^H,\ \forall i \in [\![0, I]\!] \tag{2.17}$$

with

$$\begin{aligned} \boldsymbol{\Gamma}_i &= \text{diag}\left(\alpha_{i,1},\ \ldots,\ \alpha_{i,R}\right) \\ \alpha_{i,r} &= \frac{\tau_i \lambda_r}{\sigma^2(\tau_i \lambda_r + \sigma^2)},\ \forall r \in [\![1, R]\!],\ \forall i \in [\![0, I]\!] \\ \lambda_r &= [\boldsymbol{\Lambda}_{\mathcal{H}_0}]_{r,r},\ \forall r \in [\![1, R]\!] \end{aligned} \tag{2.18}$$

For other variables fixed, the problem in (2.16) is separable for each τ_i. After some direct calculus, the objective of (2.16) w.r.t. the variable τ_i only can be obtained as

$$\mathcal{L}_i(\tau_i) = \sum_{r=1}^{R} \left[K_i \ln\left(\tau_i \lambda_r + \sigma^2\right) - \sum_{k=1}^{K_i} \frac{\tau_i\ \lambda_r\ s_{k,i}^r}{\tau_i \lambda_r + \sigma^2}\right] + \text{const.} \tag{2.19}$$

with $s_{k,i}^r = \mathbf{v}_r^H \mathbf{z}_k^i \mathbf{z}_k^{iH} \mathbf{v}_r$ and where $\mathbf{V}_{\mathcal{H}_0} = [\mathbf{v}_1, \ \ldots, \ \mathbf{v}_R]$. With this formulation of the objective over τ_i, we can directly apply Proposition B.3 in Appendix B to obtain the update of all τ_i's as in (B.15).

2.3.1.2 Update $\mathbf{\Lambda}_{\mathcal{H}_0}$

First recall that $\mathbf{\Lambda}_{\mathcal{H}_0}$ is diagonal and that we have the notation $\lambda_r = [\mathbf{\Lambda}_{\mathcal{H}_0}]_{r,r}$, $\forall r \in [\![1, R]\!]$. Following the derivations of the previous section, the objective of (2.16) is separable in λ_r and reads for each λ_r as

$$\mathcal{L}_r(\lambda_r) = \sum_{i=1}^{I} \left[K_i \ln(\tau_i \lambda_r + \sigma^2) - \sum_{k=1}^{K_i} \frac{\tau_i \lambda_r s_{k,i}^r}{\tau_i \lambda_r + \sigma^2} \right] + \text{const.} \tag{2.20}$$

Thus, one can note that $\{\tau_i\}$ and $\{\lambda_r\}$ play similar role in the objective. We can therefore obtain a surrogate by adapting the formulation in Proposition B.4 in Appendix B. The resulting update for these variables is given as

$$\lambda_r^{(t+1)} = \frac{D_r C_r}{(A_r - D_r) B_r} \tag{2.21}$$

with

$$\begin{cases} \beta_i^r = \displaystyle\sum_{k=1}^{K_i} s_{k,i}^r \frac{\tau_i \lambda_r^{(t)}}{\tau_i \lambda_r^{(t)} + \sigma^2} & \gamma_i^r = K_i + \displaystyle\sum_{k=1}^{K_i} s_{k,i}^r \frac{\tau_i \lambda_r^{(t)}}{\tau_i \lambda_r^{(t)} + \sigma^2} & A_r = \displaystyle\sum_{i=1}^{I} \gamma_i^r \\ B_r = \dfrac{\sum_{i=1}^{I} \frac{\gamma_i^r \tau_i}{\tau_i \lambda_r^{(t)} + \sigma^2}}{\sum_{i=1}^{I} \gamma_i^r} & C_r = \dfrac{\sum_{i=1}^{I} \frac{\gamma_i^r \sigma_n^2}{\tau_i \lambda_r^{(t)} + \sigma^2}}{\sum_{i=1}^{I} \gamma_i^r} & D_r = \displaystyle\sum_{i=1}^{I} \beta_i^r \end{cases} \tag{2.22}$$

2.3.1.3 Update $\mathbf{U}_{\mathcal{H}_0}$

Using (2.17), the objective in (2.16) w.r.t. $\mathbf{U}_{\mathcal{H}_0}$ fixing remaining variables is expressed as

$$\mathcal{L}_v(\mathbf{U}_{\mathcal{H}_0}) = -\sum_{i=0}^{I} \text{Tr}\{(\mathbf{U}_{\mathcal{H}_0})^H \mathbf{S}_i \mathbf{U}_{\mathcal{H}_0} \mathbf{\Gamma}_i\} + \text{const.} \tag{2.23}$$

with $\mathbf{\Gamma}_i$ in (2.18). Thus, we can directly apply Proposition B.3 in Appendix B to obtain the update

$$\mathbf{U}_{\mathcal{H}_0}^{(t+1)} = \mathbf{U}_{\text{left}} \mathbf{U}_{\text{right}}^H, \quad \text{with} \quad \sum_{i=0}^{I} \left(\mathbf{S}_i \mathbf{U}_{\mathcal{H}_0}^{(t)} \mathbf{\Gamma}_i \right) \overset{\text{TSVD}}{=} \mathbf{U}_{\text{left}} \mathbf{D} \mathbf{U}_{\text{right}}^H \tag{2.24}$$

2.3.2 MLE of $\boldsymbol{\theta}_{\mathcal{H}_1}^{\text{lrP}}$ under $\mathcal{H}_1$

Under $\mathcal{H}_1$, we have the following optimization problem

$$\begin{array}{ll} \underset{\boldsymbol{\theta}_{\mathcal{H}_1}^{\text{lrP}}}{\text{maximize}} & \mathcal{L}(\{\mathbf{z}_k^i\} | \mathcal{H}_1, \boldsymbol{\theta}_{\mathcal{H}_1}^{\text{lrP}}) \\ \text{subject to} & \tau_i \geq 0, \ \forall i \in [\![1, I]\!] \\ & \left[\mathbf{\Lambda}_{\mathcal{H}_1}^\star\right]_{r,r} \geq 0, \ \forall r \in [\![1, R]\!] \\ & \left[\mathbf{R}_{\mathcal{H}_1}^0\right]_{r,r} \geq 0, \ \forall r \in [\![1, R]\!] \\ & (\mathbf{U}_{\mathcal{H}_1}^\star)^H \mathbf{U}_{\mathcal{H}_1}^\star = \mathbf{I} \\ & (\mathbf{U}_{\mathcal{H}_1}^0)^H \mathbf{U}_{\mathcal{H}_1}^0 = \mathbf{I} \end{array} \tag{2.25}$$

This problem is equivalent to two separate ones in respectively $\{\{\tau_i\}_{i\in[\![1,I]\!]},\ \mathbf{\Lambda}^\star_{\mathcal{H}_1},\ \mathbf{U}^\star_{\mathcal{H}_1}\}$ and $\{\hat{\mathbf{R}}^0_{\mathcal{H}_1},\ \hat{\mathbf{U}}^0_{\mathcal{H}_1}\}$, for which we derive appropriate solutions in the following. The corresponding global algorithm is summed up in the box Algorithm 8.

Algorithm 8: MM algorithm to compute $\hat{\boldsymbol{\theta}}^{\text{lrP}}_{\mathcal{H}_1}$ for t^{LR}_{P}

1 **Input**: $\{\mathbf{S}_i, K_i\}$ for $i \in [\![0, I]\!]$, R, and σ^2.
2 Call Algorithm 7 on the restricted set $\{\mathbf{S}_i\}$ for $i \in [\![1, I]\!]$, the output is $\{\hat{\tau}_i\}, \hat{\mathbf{\Lambda}}^\star_{\mathcal{H}_1}, \hat{\mathbf{U}}^\star_{\mathcal{H}_1}$
3 Compute $\hat{\mathbf{R}}^0_{\mathcal{H}_1}$ and $\hat{\mathbf{U}}^0_{\mathcal{H}_1}$ with (2.29)
4 **Output**: $\hat{\boldsymbol{\theta}}^{\text{lrP}}_{\mathcal{H}_1} = \left\{\{\hat{\tau}_i\}_{i\in[\![1,I]\!]}, \hat{\mathbf{U}}^\star_{\mathcal{H}_1}, \hat{\mathbf{\Lambda}}^\star_{\mathcal{H}_1}, \hat{\mathbf{U}}^0_{\mathcal{H}_1}, \hat{\mathbf{R}}^0_{\mathcal{H}_1}\right\}$

2.3.2.1 Update $\{\tau_i\}_{i\in[\![1,I]\!]}$, $\mathbf{\Lambda}^\star_{\mathcal{H}_1}$, and $\mathbf{U}^\star_{\mathcal{H}_1}$ under $\mathcal{H}_1$

This problem requires solving

$$\begin{array}{ll} \underset{\{\tau_i\},\mathbf{\Lambda}^\star_{\mathcal{H}_1},\mathbf{U}^\star_{\mathcal{H}_1}}{\text{minimize}} & \sum_{i=1}^{I}\left[K_i \ln(|\mathbf{\Sigma}_i|) + \text{Tr}\left\{\mathbf{S}_i\mathbf{\Sigma}_i^{-1}\right\}\right] \\ \text{subject to} & \mathbf{\Sigma}_i = \tau_i\mathbf{U}^\star_{\mathcal{H}_1}\mathbf{\Lambda}^\star_{\mathcal{H}_1}(\mathbf{U}^\star_{\mathcal{H}_1})^H + \sigma^2\mathbf{I},\ \forall i \in [\![1, I]\!] \\ & \tau_i \geq 0,\ \forall i \in [\![1, I]\!] \\ & \left[\mathbf{\Lambda}^\star_{\mathcal{H}_1}\right]_{r,r} \geq 0,\ \forall r \in [\![1, R]\!] \\ & (\mathbf{U}^\star_{\mathcal{H}_1})^H\mathbf{U}^\star_{\mathcal{H}_1} = \mathbf{I}, \end{array} \tag{2.26}$$

which is identical to (2.16) except that the set $i = 0$ is excluded. Hence, we can directly apply Algorithm 7 to obtain the solutions $\{\hat{\tau}_i\}, \hat{\mathbf{\Lambda}}^\star_{\mathcal{H}_1}, \hat{\mathbf{U}}^\star_{\mathcal{H}_1}$.

2.3.2.2 Update $\hat{\mathbf{R}}^0_{\mathcal{H}_1}$ and $\hat{\mathbf{U}}^0_{\mathcal{H}_1}$ under $\mathcal{H}_1$

This problem requires solving

$$\begin{array}{ll} \underset{\mathbf{R}^0_{\mathcal{H}_1},\mathbf{U}^0_{\mathcal{H}_1}}{\text{minimize}} & K_0 \ln(|\mathbf{\Sigma}_0|) + \text{Tr}\left\{\mathbf{S}_0\mathbf{\Sigma}_\mathbf{0}^{-1}\right\} \\ \text{subject to} & \mathbf{\Sigma}_0 = \mathbf{U}^0_{\mathcal{H}_1}\mathbf{R}^0_{\mathcal{H}_1}(\mathbf{U}^0_{\mathcal{H}_1})^H + \sigma^2\mathbf{I} \\ & \left[\mathbf{R}^0_{\mathcal{H}_1}\right]_{r,r} \geq 0,\ \forall r \in [\![1, R]\!] \\ & (\mathbf{U}^0_{\mathcal{H}_1})^H\mathbf{U}^0_{\mathcal{H}_1} = \mathbf{I} \end{array} \tag{2.27}$$

The solution corresponds to the MLE of low rank structured covariance matrix in the context of Gaussian data. Let us denote the SVD of the SCM as follows

$$\mathbf{S}_0/K_0 \overset{\text{SVD}}{=} \left[\begin{array}{cc} \mathbf{U}_R & \mathbf{U}_R^\perp \end{array}\right] \begin{bmatrix} \mathbf{D}_R & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{N-R} \end{bmatrix} \left[\begin{array}{cc} \mathbf{U}_R & \mathbf{U}_R^\perp \end{array}\right]^H \tag{2.28}$$

The solution is given as (cf. section IV.A. in [100] and [99]).

$$\begin{cases} \left[\hat{\mathbf{R}}^0_{\mathcal{H}_1}\right]_{r,r} = \max([\mathbf{D}_R]_{r,r} - \sigma^2,\ 0),\ \forall r \in [\![1, R]\!] \\ \hat{\mathbf{U}}^0_{\mathcal{H}_1} = \mathbf{U}_R \end{cases} \tag{2.29}$$

2.4 Low rank subspace detector

The aforementioned detectors are based on covariance matrix equality or proportionality testing. Conversely, we focus here on the signal subspace equality testing, i.e., we aim to build a test that accounts only for a change in the signal subspace for covariance matrices as in (2.2). Specifically, we test the following hypothesis

$$\begin{cases} \mathcal{H}_0 : \Big| \mathcal{R}_R\{\boldsymbol{\Sigma}_i\} = \mathcal{R}_R\{\boldsymbol{\Sigma}\},\ \forall i \in [\![0, I]\!] \\ \mathcal{H}_1 : \Big| \begin{matrix} \mathcal{R}_R\{\boldsymbol{\Sigma}_i\} = \mathcal{R}_R\{\boldsymbol{\Sigma}\},\ \forall i \in [\![1, I]\!] \\ \mathcal{R}_R\{\boldsymbol{\Sigma}_0\} \neq \mathcal{R}_R\{\boldsymbol{\Sigma}\} \end{matrix} \end{cases} \tag{2.30}$$

where $\mathcal{R}_R\{.\}$ is the operator that extracts the dominant range space of a matrix (defined in Definition 3.4.2 of part I).

Remark 2.4.1 *This proposed test can assess for a specific underlying physical mechanism. For example, in a radar context the source power and correlations can fluctuate (change in the signal covariance matrices* $\mathbf{R}_i$ *in* (2.2)*) while spanning the same signal subspace. This leads to both*

$$\boldsymbol{\Sigma}_0 \neq \boldsymbol{\Sigma}_i \text{ and } \boldsymbol{\Sigma}_0 \not\propto \boldsymbol{\Sigma}_i,\ \forall i \tag{2.31}$$

and

$$\mathcal{R}_R\{\boldsymbol{\Sigma}_0\} = \mathcal{R}_R\{\boldsymbol{\Sigma}_i\},\ \forall i \tag{2.32}$$

Notice that (2.31) *is considered as* $\mathcal{H}_1$ *for the standard tests* (1.5) *and* (1.8) *while the relation* (2.32) *gives* $\mathcal{H}_0$ *for the test* (2.30)*. Thus, the proposed detector is insensitive to the sources correlations/power fluctuations, which can lead to lower false alarm rates in specific detection applications [91].*

We test whether the sample sets share the same principal subspace. From (2.2) and (2.30), the hypothesis test can be reformulated as

$$\begin{cases} \mathcal{H}_0 : \Big| \boldsymbol{\Sigma}_i = \mathbf{U}_{\mathcal{H}_0} \mathbf{R}^i_{\mathcal{H}_0} (\mathbf{U}_{\mathcal{H}_0})^H + \sigma^2 \mathbf{I},\ \forall i \in [\![0, I]\!] \\ \mathcal{H}_1 : \Big| \begin{matrix} \boldsymbol{\Sigma}_i = \mathbf{U}^\star_{\mathcal{H}_1} \mathbf{R}^i_{\mathcal{H}_1} (\mathbf{U}^\star_{\mathcal{H}_1})^H + \sigma^2 \mathbf{I},\ \forall i \in [\![1, I]\!] \\ \boldsymbol{\Sigma}_0 = \mathbf{U}^0_{\mathcal{H}_1} \mathbf{R}^0_{\mathcal{H}_1} (\mathbf{U}^0_{\mathcal{H}_1})^H + \sigma^2 \mathbf{I} \end{matrix} \end{cases} \tag{2.33}$$

where $\mathbf{U}_{\mathcal{H}_0}(\mathbf{U}_{\mathcal{H}_0})^H$, $\mathbf{U}^\star_{\mathcal{H}_1}(\mathbf{U}^\star_{\mathcal{H}_1})^H$ and $\mathbf{U}^0_{\mathcal{H}_1}(\mathbf{U}^0_{\mathcal{H}_1})^H$ correspond to the range spaces of respectively the secondary sets under $\mathcal{H}_0$, under $\mathcal{H}_1$ and of the tested set under $\mathcal{H}_1$. The quantities $\mathbf{R}^i_{\mathcal{H}_0}$, $\mathbf{R}^i_{\mathcal{H}_1}$ and $\mathbf{R}^0_{\mathcal{H}_1}$ denote the signal covariance matrix in low rank subspace of respectively the secondary sets under $\mathcal{H}_0$, under $\mathcal{H}_1$ and the tested set under $\mathcal{H}_1$. The GLRT for principal subspace equality is given as

$$\frac{\max_{\boldsymbol{\theta}^{\text{sub}}_{\mathcal{H}_1}} \mathcal{L}\left(\{\mathbf{z}^i_k\}|\mathcal{H}_1, \boldsymbol{\theta}^{\text{sub}}_{\mathcal{H}_1}\right)}{\max_{\boldsymbol{\theta}^{\text{sub}}_{\mathcal{H}_0}} \mathcal{L}(\{\mathbf{z}^i_k\}|\mathcal{H}_0, \boldsymbol{\theta}^{\text{sub}}_{\mathcal{H}_0})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \delta^{\text{sub}}_{\text{glr}} \tag{2.34}$$

with sets of parameters

$$\begin{aligned} \boldsymbol{\theta}^{\text{sub}}_{\mathcal{H}_0} &= \left\{ \left\{ \mathbf{R}^i_{\mathcal{H}_0} \right\}_{i \in [\![0,I]\!]}, \mathbf{U}_{\mathcal{H}_0} \right\} \\ \boldsymbol{\theta}^{\text{sub}}_{\mathcal{H}_1} &= \left\{ \left\{ \mathbf{R}^i_{\mathcal{H}_1} \right\}_{i \in [\![0,I]\!]}, \mathbf{U}^\star_{\mathcal{H}_1}, \mathbf{U}^0_{\mathcal{H}_1} \right\} \end{aligned} \tag{2.35}$$

and where functions $\mathcal{L}\left(\{\mathbf{z}_k^i\}|\mathcal{H}_1, \boldsymbol{\theta}_{\mathcal{H}_1}^{\text{sub}}\right)$, $\mathcal{L}\left(\{\mathbf{z}_k^i\}|\mathcal{H}_0, \boldsymbol{\theta}_{\mathcal{H}_0}^{\text{sub}}\right)$ denote the likelihood of the dataset $\{\mathbf{z}_k^i\}$ under respectively $\mathcal{H}_1$ and $\mathcal{H}_0$. Then, the GLRT for the proposed test reads as

$$\frac{\mathcal{L}\left(\{\mathbf{z}_k^i\}|\mathcal{H}_1, \widehat{\boldsymbol{\theta}}_{\mathcal{H}_1}^{\text{sub}}\right)}{\mathcal{L}(\{\mathbf{z}_k^i\}|\mathcal{H}_0, \widehat{\boldsymbol{\theta}}_{\mathcal{H}_0}^{\text{sub}})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \delta_{\text{glr}}^{\text{sub}} \tag{2.36}$$

where $\widehat{\boldsymbol{\theta}}_{\mathcal{H}_1}^{\text{sub}}$ and $\widehat{\boldsymbol{\theta}}_{\mathcal{H}_0}^{\text{sub}}$ are the MLE of respectively $\boldsymbol{\theta}_{\mathcal{H}_1}^{\text{sub}}$ and $\boldsymbol{\theta}_{\mathcal{H}_0}^{\text{sub}}$. In order to evaluate this GLRT, we design in the following block-coordinate descent algorithms to compute the MLE of $\boldsymbol{\theta}_{\mathcal{H}_0}^{\text{sub}}$ and $\boldsymbol{\theta}_{\mathcal{H}_1}^{\text{sub}}$. Specifically, we make use of the block MM algorithm [101] for this problem.

2.4.1 MLE of $\boldsymbol{\theta}_{\mathcal{H}_0}^{\text{sub}}$ under $\mathcal{H}_0$

Under $\mathcal{H}_0$, the likelihood optimization reduces to

$$\begin{array}{ll} \underset{\boldsymbol{\theta}_{\mathcal{H}_0}^{\text{sub}}}{\max} & \mathcal{L}(\{\mathbf{z}_k^i\}|\mathcal{H}_0, \boldsymbol{\theta}_{\mathcal{H}_0}^{\text{sub}}) \\ \text{subject to} & \mathbf{R}_{\mathcal{H}_0}^i \succcurlyeq \mathbf{0},\ \forall i \in [\![0, I]\!] \\ & \mathbf{U}_{\mathcal{H}_0}^H \mathbf{U}_{\mathcal{H}_0} = \mathbf{I} \end{array} \tag{2.37}$$

This problem is equivalent to minimizing the negative log-likelihood as

$$\begin{array}{ll} \underset{\boldsymbol{\theta}_{\mathcal{H}_0}^{\text{sub}}}{\min} & \sum_{i=0}^{I} \left[K_i \ln(|\boldsymbol{\Sigma}_i|) + \text{Tr}\left\{\mathbf{S}_i \boldsymbol{\Sigma}_i^{-1}\right\}\right] \\ \text{subject to} & \boldsymbol{\Sigma}_i = \mathbf{U}_{\mathcal{H}_0} \mathbf{R}_{\mathcal{H}_0}^i (\mathbf{U}_{\mathcal{H}_0})^H + \sigma^2 \mathbf{I},\ \forall i \in [\![0, I]\!] \\ & \mathbf{R}_{\mathcal{H}_0}^i \succcurlyeq \mathbf{0} \\ & \mathbf{U}_{\mathcal{H}_0}^H \mathbf{U}_{\mathcal{H}_0} = \mathbf{I} \end{array} \tag{2.38}$$

To solve this problem, we derive an iterative alternating algorithm that sequentially updates the variables $\{\mathbf{R}_{\mathcal{H}_0}^i\}$ and $\mathbf{U}_{\mathcal{H}_0}$. The main steps of the estimation process are summed up in the box Algorithm 9, and are briefly explained in the following.

Algorithm 9: MM algorithm to compute $\hat{\boldsymbol{\theta}}_{\mathcal{H}_0}^{\text{sub}}$

Input : $\{\mathbf{S}_i\}$ for $i \in [\![0, I]\!]$, K, R and σ^2.

1 **Repeat**

2 $t \leftarrow t + 1$

3 Update $\mathbf{R}_{\mathcal{H}_0}^{i(t)},\ \forall i \in [\![0, I]\!]$ with (2.41)

4 Update $\mathbf{U}_{\mathcal{H}_0}^{(t)}$ with (2.46)

5 **Until** a convergence criterion is met

6 **Output**: $\hat{\boldsymbol{\theta}}_{\mathcal{H}_0}^{\text{sub}} = \left\{\left\{\hat{\mathbf{R}}_{\mathcal{H}_0}^i\right\}_{i \in [\![0,I]\!]}, \hat{\mathbf{U}}_{\mathcal{H}_0}\right\}$

2.4.1.1 Update $\{\mathbf{R}_{\mathcal{H}_0}^i\}$

Considering $\{\mathbf{R}_{\mathcal{H}_0}^i\}$ while fixing the remain variables, the problem in (2.38) is separable for each $\mathbf{R}_{\mathcal{H}_0}^i$. The objective of (2.38) w.r.t. the variable $\mathbf{R}_{\mathcal{H}_0}^i$ can be expressed as

$$\mathcal{L}_r(\mathbf{R}_{\mathcal{H}_0}^i) = K_i \ln\left|(\mathbf{R}_{\mathcal{H}_0}^i + \sigma^2 \mathbf{I})\right| \text{Tr}\left\{\tilde{\mathbf{S}}_i (\mathbf{R}_{\mathcal{H}_0}^i + \sigma^2 \mathbf{I})^{-1}\right\} \tag{2.39}$$

with $\tilde{\mathbf{S}}_i = \mathbf{U}_{\mathcal{H}_0}^H \mathbf{S}_i \mathbf{U}_{\mathcal{H}_0}$. The minimizer of this objective w.r.t. $\mathbf{R}_{\mathcal{H}_0}^i$ corresponds therefore to the MLE of a structured covariance matrix for dimension reduced Gaussian variables. Denoting the SVD of the dimension reduced SCM as

$$\tilde{\mathbf{S}}_i / K_i \overset{\text{SVD}}{=} [\mathbf{Q}_R | \mathbf{Q}_R^\perp] \begin{bmatrix} \mathbf{D}_R & 0 \\ 0 & \mathbf{D}_{N-R} \end{bmatrix} [\mathbf{Q}_R | \mathbf{Q}_R^H]^H \tag{2.40}$$

Thus, the update is given as

$$\begin{cases} \mathbf{R}_{\mathcal{H}_0}^{i(t+1)} = \mathbf{Q}_R \tilde{\mathbf{D}}_R \mathbf{Q}_R^H \\ [\tilde{\mathbf{D}}_R]_{r,r} = \max([\mathbf{D}_R]_{r,r} - \sigma^2,\ 0),\ \forall r \in [\![1, R]\!] \end{cases} \tag{2.41}$$

2.4.1.2 Update $\mathbf{U}_{\mathcal{H}_0}$

First, remark that $\mathbf{\Sigma}_i^{-1}$ can be expressed thanks to the matrix inversion lemma as

$$\begin{aligned} \mathbf{\Sigma}_i^{-1} &= (\mathbf{U}_{\mathcal{H}_0} \mathbf{R}_{\mathcal{H}_0}^i (\mathbf{U}_{\mathcal{H}_0})^H + \sigma^2 \mathbf{I})^{-1} \\ &= \sigma^{-2}\mathbf{I} - \mathbf{U}_{\mathcal{H}_0} \underbrace{\sigma^{-4}((\mathbf{R}_{\mathcal{H}_0}^i)^{-1} + \sigma^{-2}\mathbf{I})^{-1}}_{\mathbf{W}_i} \mathbf{U}_{\mathcal{H}_0}^H \end{aligned} \tag{2.42}$$

After some calculus, the objective in (2.38) w.r.t. $\mathbf{U}_{\mathcal{H}_0}$ for other fixed variables can be expressed

$$f(\mathbf{U}_{\mathcal{H}_0}) = \sum_{i=0}^{I} \text{Tr}\{(\mathbf{U}_{\mathcal{H}_0})^H \mathbf{S}_i \mathbf{U}_{\mathcal{H}_0} \mathbf{W}_i\} \tag{2.43}$$

An update of $\mathbf{U}_{\mathcal{H}_0}$ can be obtained in closed form by following the MM approach [101]. Applying the proposition B.1 in Appendix B, the objective can be majorized with equality at point $\mathbf{U}_{\mathcal{H}_0}^{(t)}$ as

$$f(\mathbf{U}_{\mathcal{H}_0}) \leq g(\mathbf{U}_{\mathcal{H}_0} | \mathbf{U}_{\mathcal{H}_0}^{(t)}) \tag{2.44}$$

with

$$g(\mathbf{U}_{\mathcal{H}_0} | \mathbf{U}_{\mathcal{H}_0}^{(t)}) = \sum_{i=1}^{I} 2\text{Re}\left[\text{Tr}\{(\mathbf{U}_{\mathcal{H}_0})^H \mathbf{S}_i (\mathbf{U}_{\mathcal{H}_0})^{(t)} \mathbf{W}_i\}\right]$$

Then, $\mathbf{U}_{\mathcal{H}_0}^{(t+1)}$ is obtained as the minimizer of the following problem

$$\begin{aligned} \min_{\mathbf{U}_{\mathcal{H}_0}} \quad & g(\mathbf{U}_{\mathcal{H}_0} | \mathbf{U}_{\mathcal{H}_0}^{(t)}) \\ \text{subject to} \quad & \mathbf{U}_{\mathcal{H}_0}^H \mathbf{U}_{\mathcal{H}_0} = \mathbf{I} \end{aligned} \tag{2.45}$$

Solving this optimization problem under the orthonormality constraint leads to an update of the form

$$\mathbf{U}_{\mathcal{H}_0}^{(t+1)} = \mathbf{U}_{\text{left}} \mathbf{U}_{\text{right}}^H, \tag{2.46}$$

with $\mathbf{U}_{\text{left}}$ and $\mathbf{U}_{\text{right}}^H$ are the left and right eigenvectors of the TSVD, that is

$$\sum_{i=0}^{I} \left(\mathbf{S}_i \mathbf{U}_{\mathcal{H}_0}^{(t)} \mathbf{W}_i\right) \overset{\text{TSVD}}{=} \mathbf{U}_{\text{left}} \mathbf{D} \mathbf{U}_{\text{right}}^H \tag{2.47}$$

2.4.2 MLE of $\boldsymbol{\theta}_{\mathcal{H}_1}^{\text{sub}}$ under $\mathcal{H}_1$

Under $\mathcal{H}_1$, the optimization problem reads

$$\begin{array}{ll} \underset{\boldsymbol{\theta}_{\mathcal{H}_1}^{\text{sub}}}{\max} & \mathcal{L}(\{\mathbf{z}_k^i\}|\mathcal{H}_1, \boldsymbol{\theta}_{\mathcal{H}_1}^{\text{sub}}) \\ \text{subject to} & \mathbf{R}_{\mathcal{H}_0}^i \succcurlyeq \mathbf{0} \\ & (\mathbf{U}_{\mathcal{H}_1}^0)^H \mathbf{U}_{\mathcal{H}_1}^0 = \mathbf{I} \\ & (\mathbf{U}_{\mathcal{H}_1}^\star)^H \mathbf{U}_{\mathcal{H}_1}^\star = \mathbf{I} \end{array} \tag{2.48}$$

which is separable in $\{\{\mathbf{R}_{\mathcal{H}_1}^i\}_{i\in[\![1,I]\!]},\ \mathbf{U}_{\mathcal{H}_1}^\star\}$ and $\{\mathbf{R}_{\mathcal{H}_1}^0,\ \mathbf{U}_{\mathcal{H}_1}^0\}$, for which we derive appropriate solutions in the following. The corresponding algorithm is summed up in the box Algorithm 10 and a brief explanation of the procedure updates are given below.

Algorithm 10: MM algorithm to compute $\hat{\boldsymbol{\theta}}_{\mathcal{H}_1}^{\text{sub}}$

input : $\{\mathbf{S}_i\}$ for $i \in [\![0, I]\!]$, K, R and σ^2.

1 Call Algorithm 9 on the restricted set $\{\mathbf{S}_i\}$ for $i \in [\![1, I]\!]$, the output is $\{\hat{\mathbf{R}}_{\mathcal{H}_1}^i\}_{i\in[\![1,I]\!]}$ and $\hat{\mathbf{U}}_{\mathcal{H}_1}^\star$

2 Compute $\hat{\mathbf{R}}_{\mathcal{H}_1}^0$ and $\hat{\mathbf{U}}_{\mathcal{H}_1}^0$ with (2.51)

3 **Output**: $\hat{\boldsymbol{\theta}}_{\mathcal{H}_1}^{\text{sub}} = \left\{\{\hat{\mathbf{R}}_{\mathcal{H}_1}^i\}_{i\in[\![0,I]\!]}, \hat{\mathbf{U}}_{\mathcal{H}_1}^\star, \hat{\mathbf{U}}_{\mathcal{H}_1}^0\right\}$

2.4.2.1 Update $\{\{\mathbf{R}_{\mathcal{H}_1}^i\}_{i\in[\![1,I]\!]},\ \mathbf{U}_{\mathcal{H}_1}^\star\}$ under $\mathcal{H}_1$

This problem is identical to (2.38) except that the set $i = 0$ is excluded. Hence, we can directly use Algorithm 9 to obtain the solutions $\{\hat{\mathbf{R}}_{\mathcal{H}_1}^i\}_{i\in[\![1,I]\!]}$ and $\hat{\mathbf{U}}_{\mathcal{H}_1}^\star$.

2.4.2.2 Update $\{\mathbf{R}_{\mathcal{H}_1}^0,\ \mathbf{U}_{\mathcal{H}_1}^0\}$ under $\mathcal{H}_1$

This problem reduces to

$$\begin{array}{ll} \underset{\mathbf{R}_{\mathcal{H}_1}^0,\ \mathbf{U}_{\mathcal{H}_1}^0}{\min} & K_i \ln(|\boldsymbol{\Sigma}_i|) + \text{Tr}\left\{\mathbf{S}_i \boldsymbol{\Sigma}_i^{-1}\right\} \\ \text{subject to} & \boldsymbol{\Sigma}_0 = \mathbf{U}_{\mathcal{H}_1}^0 \mathbf{R}_{\mathcal{H}_1}^0 (\mathbf{U}_{\mathcal{H}_1}^0)^H + \sigma^2 \mathbf{I} \\ & \mathbf{R}_{\mathcal{H}_1}^0 \succcurlyeq \mathbf{0} \\ & (\mathbf{U}_{\mathcal{H}_1}^0)^H \mathbf{U}_{\mathcal{H}_1}^0 = \mathbf{I} \end{array} \tag{2.49}$$

Then, the solution corresponds to the MLE of the low rank structured covariance matrix in the context of Gaussian data [99]. Let us denote the SVD of the SCM as follows

$$\mathbf{S}_0/K_0 \overset{\text{SVD}}{=} \left[\begin{array}{cc} \mathbf{U}_R & \mathbf{U}_R^\perp \end{array}\right] \left[\begin{array}{cc} \mathbf{D}_R & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{N-R} \end{array}\right] \left[\begin{array}{cc} \mathbf{U}_R & \mathbf{U}_R^\perp \end{array}\right]^H \tag{2.50}$$

The solution reads

$$\begin{cases} \left[\hat{\mathbf{R}}_{\mathcal{H}_1}^0\right]_{r,r} = \max([\mathbf{D}_R]_{r,r} - \sigma^2,\ 0),\ \forall r \in [\![1, R]\!] \\ \hat{\mathbf{U}}_{\mathcal{H}_1}^0 = \mathbf{U}_R \end{cases} \tag{2.51}$$

Chapter 3

Simulations and applications

In the following, we illustrate the performance detection of the proposed methods through simulated data. Furthermore, we assess the interest of these detectors for a change detection application on a SAR dataset.

3.1 Simulations

3.1.1 Simulation setup

This section presents numerical simulations to assess the performance of the following detectors:

- t_{E} stands for the GLRT for the equality testing [5].
- $t_{\mathrm{E}}^{\mathrm{LR}}$ denotes the GLRT for the low rank structured covariance matrix equality testing.
- t_{P} denotes the GLRT for the proportionality testing.
- $t_{\mathrm{P}}^{\mathrm{LR}}$ is the GLRT for testing the proportionality of low rank signal covariance matrix component.
- t_{sub} stands for the GLRT detector for the subspace equality.

To this end, we consider, $N = 20$, $R = 5$, $K_i = K = 25$, $\forall i \in [\![0, I]\!]$, $I = 3$, the samples $\mathbf{z}_k^i$ are drawn from i.i.d. complex normal distribution, i.e., $\mathbf{z}_k^i \sim \mathcal{CN}(0, \mathbf{\Sigma}_i)$ and $\mathbf{\Sigma}_i = \tau_i \mathbf{U}_i \mathbf{\Lambda} \mathbf{U}_i^H + \sigma^2 \mathbf{I}$ where $\mathbf{U}_i \in \mathcal{U}_N^R$, $\tau_i \in \mathbb{R}^+$ and $\mathbf{\Lambda}$ is a diagonal matrix. The eigenvalues $[\mathbf{\Lambda}]_{r,r} = \alpha(R+1-r)$, the signal to noise ratio $\mathrm{SNR} = \mathrm{Tr}\{\mathbf{\Lambda}\}/R\sigma^2$ with $\sigma^2 = 1$. As comparative criteria, we consider the receiver operating characteristic curve (ROC) for the following scenarios:

- Scenario 1: Under $\mathcal{H}_0$ and $\mathcal{H}_1$, $\tau_i = 1$ and $\mathbf{U}_i = \mathbf{U}$, $\forall i \in [\![1, I]\!]$ where $\mathbf{U} \in \mathcal{U}_N^R$ is built from the R first elements of the canonical basis. The covariance matrix of the tested set, under $\mathcal{H}_1$, reads as $\mathbf{\Sigma}_0 = \mathbf{U}_0 \mathbf{\Lambda} \mathbf{U}_0^H + \sigma^2 \mathbf{I}$ where $\mathbf{U}_0 \in \mathcal{U}_N^R$ is generated by changing the first 2 eigenvectors of $\mathbf{U}$, i.e., $\mathbf{U}_0 \neq \mathbf{U}$ so that $\mathcal{R}_R\{\mathbf{\Sigma}_0\} \neq \mathcal{R}_R\{\mathbf{\Sigma}_i\}$, $\forall i$. **This change is occurred by inverting the order of the eigenvalues of $\mathbf{\Sigma}_i$.** This scenario corresponds to a strict change in the covariance matrix where all the secondary sets are homogeneous.

- Scenario 2: Under $\mathcal{H}_0$ and $\mathcal{H}_1$, $\tau_i \sim \Gamma(\nu, 1/\nu)$ (with $\nu = 1$) and $\mathbf{U}_i = \mathbf{U}$, $\forall i \in [\![1, I]\!]$ where $\mathbf{U} \in \mathcal{U}_N^R$ is built from the R first elements of the canonical basis. Under $\mathcal{H}_1$, the anomaly in the covariance matrix of the tested set $\mathbf{\Sigma}_0 = \mathbf{U}_0 \mathbf{\Lambda} \mathbf{U}_0^H + \sigma^2 \mathbf{I}$ is generated again by changing its principal subspace, i.e., $\mathcal{R}_R\{\mathbf{\Sigma}_0\} \neq \mathcal{R}_R\{\mathbf{\Sigma}_i\}$, $\forall i$ as in Scenario 1. This

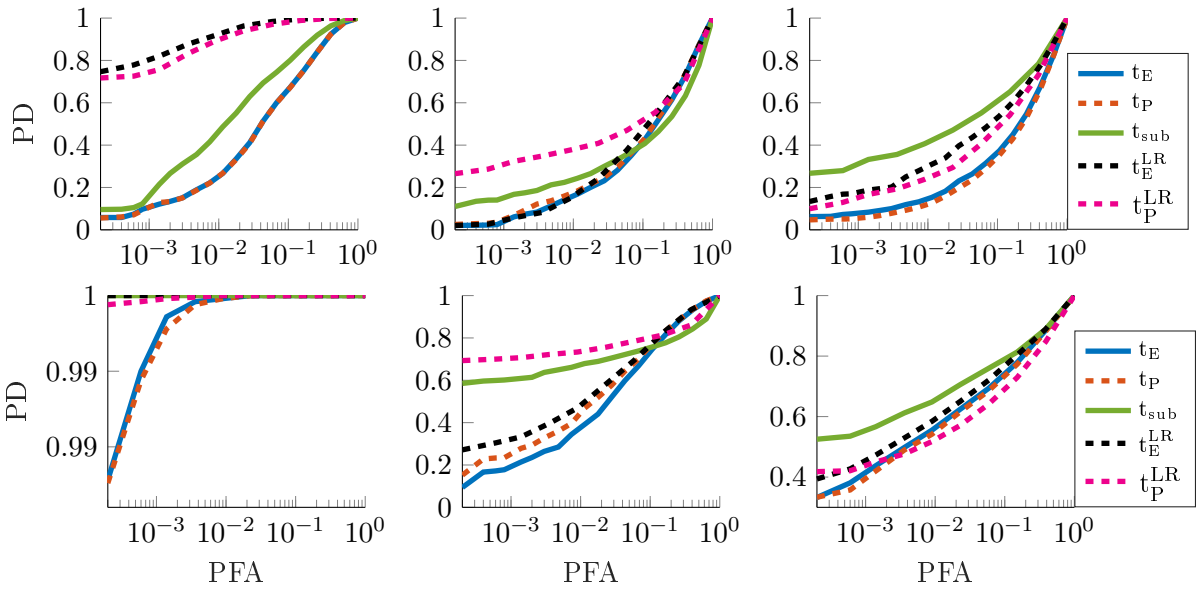


Figure 3.1: ROC curves for Scenario 1 to 3 (from left to right column) where SNR = 0dB (top) and SNR = 5dB (bottom).

scenario accounts for power fluctuations of the signal w.r.t. the batches i. Hence, the signal covariance matrix are proportional under $\mathcal{H}_0$, but the total covariance matrices are not identical.

- Scenario 3 : Under $\mathcal{H}_0$ and $\mathcal{H}_1$, $\tau_i \sim \Gamma(\nu, 1/\nu)$ (with $\nu = 1$) and $\mathbf{U}_i = \mathbf{U}\mathbf{Q}_i$, $\forall i \in [\![1, I]\!]$ where $\mathbf{U} \in \mathcal{U}_N^R$ is built from the R first elements of the canonical basis and $\{\mathbf{Q}_i\}$ is a set of $R \times R$ rotation matrices. We note that, under $\mathcal{H}_0$, we have (2.31) while the relation (2.32) is also satisfied. The covariance matrix of the tested set is $\mathbf{\Sigma}_0 = \mathbf{U}_0\mathbf{\Lambda}\mathbf{U}_0^H + \sigma^2\mathbf{I}$, under $\mathcal{H}_1$, which corresponds to a change in its principal subspace where $\mathcal{R}_R\{\mathbf{\Sigma}_0\} \neq \mathcal{R}_R\{\mathbf{\Sigma}_i\}$, $\forall i$ as in the other scenarios. Scenario 3 aims to test a specific physical phenomenon as discussed in Remark 2.4.1.

3.1.2 Results

Figure 1 presents the ROC of the different detectors under various scenarios and SNRs:

- The first column (left) displays the ROC curves of the detectors for Scenario 1 for respectively SNR=0dB (top) and SNR=5dB (bottom). Under this setting, t_P and t_E appear to have identical performance. The proposed detector t_{sub} outperforms t_P and t_E since it exploits the low rank structure information. However, the detectors t_P^{LR} and t_E^{LR} outperform all the others. This result was to be expected since these two detectors are specifically suited to this scenario (t_E^{LR} corresponds to the GLRT for this exact setting). We note that at SNR= 5dB, the change detection problem is not especially challenging in this setting, thus all detectors show high performance, i.e., probability detection (PD)$\approx$ 1 for probability false alarm (PFA)$\approx 10^{-4}$.

- The second column (central) displays the ROC curves of detectors for Scenario 2 for respectively SNR=0dB (top) and SNR=5dB (bottom). In this setting low rank proportionality detector $t_{\mathrm{P}}^{\mathrm{LR}}$ corresponds to the most appropriate GLRT, thus it outperforms the other detectors. An explanation is that even under $\mathcal{H}_0$, the total covariance matrices of the different batches verify (2.31) (non-equal and non-proportional). Thus, t_{E}, t_{P}, and $t_{\mathrm{E}}^{\mathrm{LR}}$ require a high detection threshold to ensure a low PFA. It is worth mentioning that we are mostly interested in the performance of the detectors for low PFA (below 0.1). In this range, t_{sub} also offers interesting performance since it is, as $t_{\mathrm{P}}^{\mathrm{LR}}$, designed to be insensitive to power fluctuations of the low rank signal covariance matrix.

- The third (right) column displays the ROC curves of detectors for Scenario 3 for respectively SNR=0dB (top) and SNR=5dB (bottom). In this setting, the proposed detector t_{sub} outperforms t_{E}, $t_{\mathrm{E}}^{\mathrm{LR}}$, t_{P} and $t_{\mathrm{P}}^{\mathrm{LR}}$, since equality, low rank equality, proportionality and low rank proportionality detectors do not directly infer on the equality of signal subspace in the context of low rank structure covariance matrices. For the considered scenario, the correlation between the signal components change w.r.t. i even under $\mathcal{H}_0$. However, (2.32) is satisfied so t_{sub} is not sensitive to this type of heterogeneity. Therefore, this detector allows to reduce the PFA as discussed in Remark 2.4.1.

3.2 Applications: change detection on real data

In this section, the performance of the proposed detectors is illustrated for change detection on a UAVSAR dataset. This application has been done in collaboration with Ammar Mian from SONDRA.

3.2.1 Setup

The considered dataset is SanAnd_26524_03 Segment 4, of coordinates (top left pixel) [2891, 28891], with 2 acquisition dates : April 23, 2009 and May 11, 2015. The ground truth for change detection is taken from [102] and presented in Figure 3.2. For one acquisition, the initial data cube size is $2360 \times 600 \times 3$ and is is pre-processed using the wavelet-decomposition transform presented in [103]. This transformation, which allows to decompose a SAR image into canals corresponding to a physical behaviour of the scatterers, have been shown to increase the detection performance [103]. This transformation increases the size of a sample from $N = 3$ to $N = 12$. The detectors are applied using 5×5 pixel patches.

3.2.2 Compared detectors

We compare the same detectors as in the previous section. The proposed LR detection methods are using a rank $R = 1$, and the white Gaussian noise power is adaptively estimated locally by using the mean of the $(N - R)$ last eigenvalues of the SCM of all samples in the patch. To show the benefits of the multivariate setting, we also compare the results with a standard monovariate detector applied to the summed entries of each pixel, denoted t_{m}.

Remark 3.2.1 *We note that for $R = 1$, the detector for the subspace equality t_{sub} coincides with the low rank proportionality detector t_P^{LR} so t_{sub} is not displayed.*

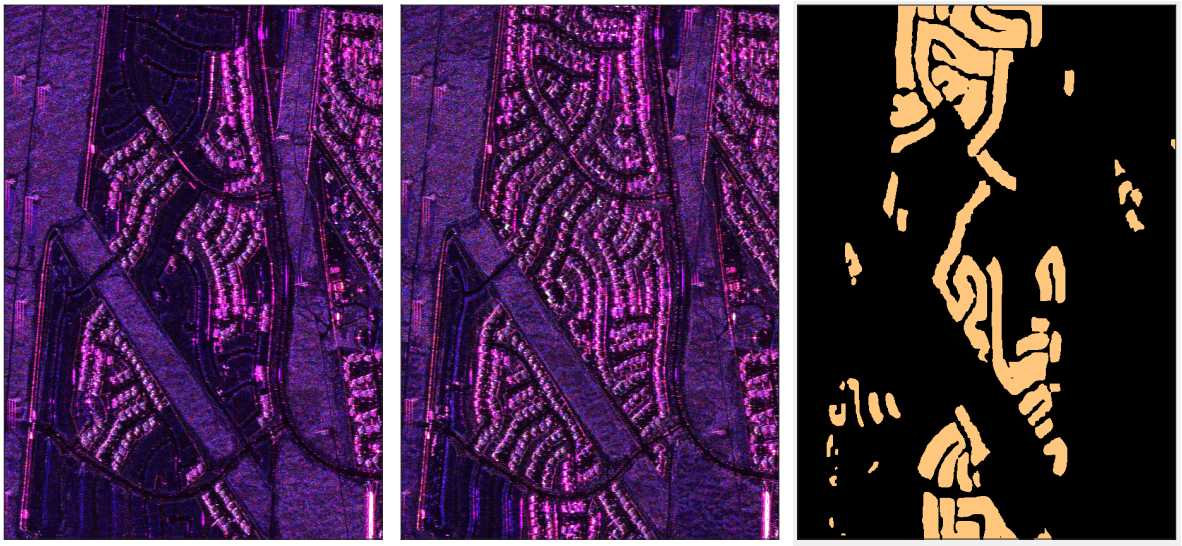


Figure 3.2: UAVSAR Dataset in Pauli representation. Left: April 23, 2009. Middle: May 15, 2011. Right: Ground Truth for change detection.

3.2.3 Results

Figure 3.3 displays the detectors output. The ROC curves of the different detectors is displayed in Figure 3.4. For this example, $t_{\mathrm{E}}^{\mathrm{LR}}$ offers an improvement of the detection performance compared to the standard equality testing t_{E}. This result underlines the interest of the proposed LR formulation. $t_{\mathrm{P}}^{\mathrm{LR}}$ offers similar performance, but only for low false alarm rate. Intuitively, a power fluctuation seems interesting to capture when it comes to change detection. Therefore, proportionality testing (not sensitive to this change) may not be the most appropriate for this application, as illustrated by the performance of t_{P}. However, this detector is still interesting for other purposes, such as local anomaly detection [91].

For each detector, Figure 3.5 displays the PD w.r.t. the spatial window size $\sqrt{K}$ with fixed PFA $= 5\%$. Up to a reasonable window size, the detection performance increases w.r.t. K. However, this is at the detriment of the spatial resolution of the process. This figure hence illustrates the interest of the proposed LR methods, as they offer an improvement of the performance/resolution trade-off. Notably, the proposed methods allow for $K < N$, where other standard covariance based detectors are not defined (due to non-invertible SCMs).

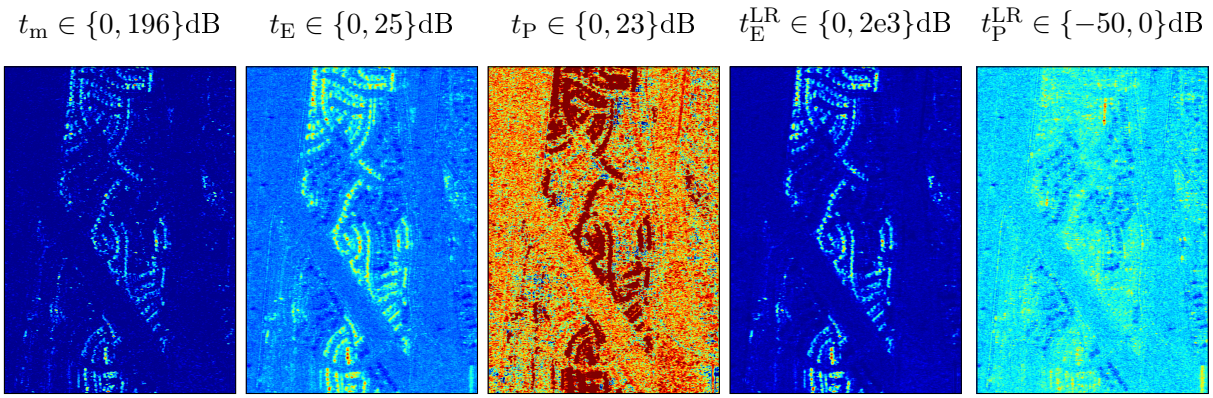


Figure 3.3: Output (dynamic range in dB indicated on top) obtained for the different detectors on the UAVSAR dataset.

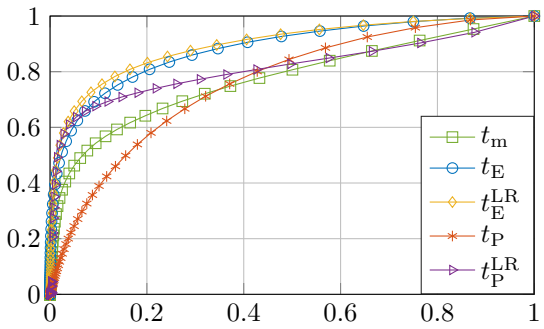


Figure 3.4: ROC curves (PD versus PFA) of the different detectors on the UAVSAR dataset.

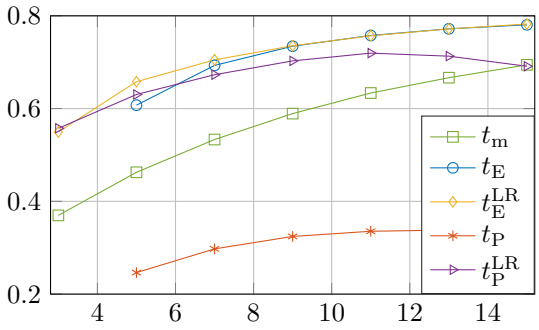


Figure 3.5: PD versus spatial window size $\sqrt{K}$ for fixed PFA $= 5\%$ of the different detectors on the UAVSAR dataset.

Chapter 4

Conclusion and perspectives

We proposed detectors in the context of low rank structured covariance matrices. The newly proposed detectors are only sensitive to the low rank principal component. To this end, we derived the GLRT to infer on the equality/proportionality of the low rank signal component and on the equality of the principal subspace of signal of interest. The GLRT of detectors namely the low rank proportionality and the low rank subspace detectors involved non-trivial optimization problem. In order to evaluate these GLRTs, we designed in the following block-coordinate descent algorithms to compute the aforementioned detectors. Specifically, we made use of the block MM algorithm for these problems. Finally, the performance detection of the proposed methods were illustrated on simulated data. Furthermore, the interest of these detectors were also assessed for a change detection application on a SAR dataset.

In the following, we list some possible future prospects related to this works:

- In this part of thesis, we focused on GLRT formulations, it would be interesting to integrate low rank structures in other statistics (e.g. Rao, Wald, or Gradient), or in other detection paradigms (e.g. matrix distances based, or Information Theory based).
- The use of local adaptive rank selection is also left as potential extension. For more details we refer to the Section 1.1.2 in Part II.
- We may adopt the data model in Chapter 2 in Part II (i.e., adding Bayesian priors) to derive a detection scheme by testing common properties between low rank structured covariance matrices.

Appendices

Appendix A

Generation of complex generalized Bingham Langevin (CGBL) distribution

In this appendix, we show how to sample a unitary random matrix $\mathbf{U} \in \mathcal{U}_N^R$ from the CGBL, i.e., $\mathbf{U} \sim \text{CGBL}(\mathbf{C}, \{\mathbf{A}_r\})$, where, $\mathbf{C} \in \mathbb{C}^{N \times R}$ and $\{\mathbf{A}_r\}$ is a set of Hermitian matrices. The p.d.f. of $\mathbf{U}$ is given by:

$$p_{\text{CGBL}}(\mathbf{U}) \propto \exp\left\{\sum_{r=1}^{R} \text{Re}\{\mathbf{c}_r^H \mathbf{u}_r\} + \mathbf{u}_r^H \mathbf{A}_r \mathbf{u}_r\right\} \tag{A.1}$$

where, $\mathbf{u}_r$ and $\mathbf{c}_r$ stand for the r-th column of respectively $\mathbf{U}$ and $\mathbf{C}$.

A.1 The real vector Bingham Langevin (vBL) distribution

The vBL distribution [40] is a probability distribution on the set of unitary real vectors which combines linear and quadratic terms. For instance, a given real unitary vector $\tilde{\mathbf{u}}$ follows the vBL distribution, i.e., $\tilde{\mathbf{u}} \sim \text{vBL}(\tilde{\mathbf{c}}, \tilde{\mathbf{A}})$. Its p.d.f. reads as

$$p_{\text{vBL}}(\tilde{\mathbf{u}}) \propto \exp\{\tilde{\mathbf{c}}^T \tilde{\mathbf{u}} + \tilde{\mathbf{u}}^T \tilde{\mathbf{A}} \tilde{\mathbf{u}}\} \tag{A.2}$$

where $\tilde{\mathbf{A}} \in \mathcal{S}_N^+$ and $\tilde{\mathbf{c}}$ is a real vector. An acceptance rejection method is proposed and investigated in [41] to generate random vector following real Bingham distribution. Inspired from this scheme, we derive a new sampling technique for the vBL distribution as detailed below. First, let us define the acceptance rejection scheme.

A.1.1 Acceptance rejection method

Consider two known p.d.f. $f(\mathbf{u})$ and $g(\mathbf{u})$. Suppose it is possible to simulate from g and it is desired to simulate random variable from f. The key requirement is that there is a known bound of the form

$$f(\mathbf{u}) \leq M g(\mathbf{u}) \tag{A.3}$$

where M is a known scalar and $\mathbf{u} \in \mathbb{R}^N$. The acceptance rejection algorithm proceeds as follows:
- Step 1: Simulate $\mathbf{u} \sim g$
- Step 2: Simulate $u \sim \text{Unif}(0, 1)$.

• Step 3: if $u < f(\mathbf{u})/(Mg(\mathbf{u}))$, then accept $\mathbf{u}$
• Step 4: Otherwise go back to Step1.
It is worth mentioning that the bound M should satisfy the following condition $M \geq 1$. The efficiency is defined by $1/M$. For high efficiency, the bound M should be as close to 1 as possible.

A.1.2 Acceptance rejection method for the real Bingham distribution

In [41], an acceptance rejection scheme is proposed to sample the real Bingham distribution with an angular central Gaussian (ACG) envelope. This sampling technique is summed up in the box **Algorithm 1** and a brief explanation of the sampling procedure is given below. The p.d.f. of the real Bingham distribution reads as

$$f_{\mathrm{B}}(\tilde{\mathbf{u}}) = c_{\mathrm{B}}\exp\{-\tilde{\mathbf{u}}^T\tilde{\mathbf{A}}\tilde{\mathbf{u}}\},\ f^*_{\mathrm{B}}(\tilde{\mathbf{u}}) = \exp\{-\tilde{\mathbf{u}}^T\tilde{\mathbf{A}}\tilde{\mathbf{u}}\} \tag{A.4}$$

where, c_{B} is a normalizing constant, $\tilde{\mathbf{u}}$ is unit real vector and $\tilde{\mathbf{A}} \in \mathcal{S}_N^+$. The ACG distribution is denoted as $\mathrm{ACG}(\tilde{\mathbf{\Omega}})$ with $\mathbf{\Omega}$ is a symmetric positive definite matrix, its p.d.f. is given below

$$f_{\mathrm{ACG}}(\tilde{\mathbf{u}}) = c_{\mathrm{ACG}}(\tilde{\mathbf{u}}^T\tilde{\mathbf{\Omega}}\tilde{\mathbf{u}})^{-N/2},\ f^*_{\mathrm{ACG}}(\tilde{\mathbf{u}}) = (\tilde{\mathbf{u}}^T\tilde{\mathbf{\Omega}}\tilde{\mathbf{u}})^{-N/2},\ c_{\mathrm{ACG}} = |\tilde{\mathbf{\Omega}}|^{1/2} \tag{A.5}$$

with c_{ACG} is a normalizing constant. Considering $w = \tilde{\mathbf{u}}^T\tilde{\mathbf{A}}\tilde{\mathbf{u}}$ and $\tilde{\mathbf{\Omega}} = \mathbf{I} + 2\tilde{\mathbf{A}}/b$, $b > 0$, the following inequality is achieved

$$\begin{aligned} f^*_{\mathrm{B}}(\tilde{\mathbf{u}}) &= \exp\{-w\} \\ &\leq \exp\{-(N-b)/2\}\left(\frac{N/b}{1+2w/b}\right)^{N/2} \\ &\leq \exp\{-(N-b)/2\}(\tilde{\mathbf{u}}^T\tilde{\mathbf{\Omega}}\tilde{\mathbf{u}})^{-N/2} \\ &\leq \exp\{-(N-b)/2\}(N/b)^{N/2} f^*_{\mathrm{ACG}}(\tilde{\mathbf{u}}) \end{aligned} \tag{A.6}$$

Then,

$$f^*_{\mathrm{B}}(\tilde{\mathbf{u}}) \leq M(b) f^*_{\mathrm{ACG}}(\tilde{\mathbf{u}}) \tag{A.7}$$

with $M(b) = \exp\{-(N-b)/2\}(N/b)^{N/2}$. The function $log(M(b))$ is convex in b admitting a unique minimizer b^* satisfying

$$\sum_{i=1}^{N} \frac{1}{b^* + 2\lambda_i} = 1 \tag{A.8}$$

where the set $\{\lambda_i\}$ denotes the eigenvalues of the matrix $\tilde{\mathbf{A}}$. The generation of random vector $\tilde{\mathbf{u}} \sim \mathrm{ACG}(\tilde{\mathbf{\Omega}})$ [35] is given as follows:
• Step 1: Sample $\tilde{\mathbf{y}} \sim \mathcal{N}(0, \tilde{\mathbf{\Omega}}^{-1})$
• Step 2: Compute $\tilde{\mathbf{u}} = \tilde{\mathbf{y}}/||\tilde{\mathbf{y}}||$ where $\tilde{\mathbf{u}} \sim \mathrm{ACG}(\tilde{\mathbf{\Omega}})$

A.1.3 The proposed sampling method for the vBL distribution

The vBL distribution takes the form:

$$f_{\mathrm{vBL}}(\tilde{\mathbf{u}}) = c_{\mathrm{vBL}}\exp\{\tilde{\mathbf{c}}^T\tilde{\mathbf{u}} + \tilde{\mathbf{u}}^T\tilde{\mathbf{A}}\tilde{\mathbf{u}}\}, f^*_{\mathrm{vBL}}(\tilde{\mathbf{u}}) = \exp\{\tilde{\mathbf{c}}^T\tilde{\mathbf{u}} + \tilde{\mathbf{u}}^T\tilde{\mathbf{A}}\tilde{\mathbf{u}}\}. \tag{A.9}$$

where, c_{vBL} is a normalizing constant. Inspired from [41], we aim to derive a new sampling technique. Let us start first with introducing an inequality between the densities $f^*_{\mathrm{B}}(\tilde{\mathbf{u}})$ and

Algorithm 11: Acceptance rejection sampling scheme for the real Bingham distribution

input : $\tilde{\mathbf{A}}$
output : $\tilde{\mathbf{u}}$
Compute b^* satisfying (A.8)
Compute $\tilde{\mathbf{\Omega}} = \mathbf{I} + 2\tilde{\mathbf{A}}/b^*$
Compute $M(b^*) = \exp\{-(N-b^*)/2\}(N/b^*)^{N/2}$
Repeat
Sample $\tilde{\mathbf{y}} \sim \mathcal{N}(0, \tilde{\mathbf{\Omega}}^{-1})$
Compute $\tilde{\mathbf{u}} = \tilde{\mathbf{y}}/||\tilde{\mathbf{y}}||$
Compute $f^*_{\text{ACG}}(\tilde{\mathbf{u}}) = (\tilde{\mathbf{u}}^T\tilde{\mathbf{\Omega}}\tilde{\mathbf{u}})^{-N/2}$
Compute $f^*_{\text{B}}(\tilde{\mathbf{u}}) = \exp\{-\tilde{\mathbf{u}}^T\tilde{\mathbf{A}}\tilde{\mathbf{u}}\}$
Sample $u \sim \text{Unif}(0,1)$
Until $u < f^*_{\text{B}}(\tilde{\mathbf{u}})/(M(b^*)f^*_{\text{ACG}}(\tilde{\mathbf{u}}))$
Accept $\tilde{\mathbf{u}}$

$f^*_{\text{vBL}}(\tilde{\mathbf{u}})$. By considering $y = \tilde{\mathbf{c}}^T\tilde{\mathbf{u}}$ and $(1-y)^2 \geq 0$, we have

$$\begin{aligned}
f^*_{\text{vBL}}(\tilde{\mathbf{u}}) &= \exp\{y + \tilde{\mathbf{u}}^T\tilde{\mathbf{A}}\tilde{\mathbf{u}}\} \\
&\leq \exp\{1/2(1+y^2) + \tilde{\mathbf{u}}^T\tilde{\mathbf{A}}\tilde{\mathbf{u}}\} = \exp\{1/2(1 + \tilde{\mathbf{u}}^T\tilde{\mathbf{c}}\tilde{\mathbf{c}}^T\tilde{\mathbf{u}}) + \tilde{\mathbf{u}}^T\tilde{\mathbf{A}}\tilde{\mathbf{u}}\} \\
&\leq \exp\{1/2\}\exp\{\tilde{\mathbf{u}}^T(\tilde{\mathbf{A}} + 1/2\tilde{\mathbf{c}}\tilde{\mathbf{c}}^T)\tilde{\mathbf{u}}\} = \exp\{1/2\}\exp\{-\tilde{\mathbf{u}}^T(-\tilde{\mathbf{A}} - 1/2\tilde{\mathbf{c}}\tilde{\mathbf{c}}^T)\tilde{\mathbf{u}}\}\exp\{-\gamma+\gamma\} \\
&\leq \exp\{1/2+\gamma\}\exp\{-\tilde{\mathbf{u}}^T(\gamma\mathbf{I} - \tilde{\mathbf{A}} - 1/2\tilde{\mathbf{c}}\tilde{\mathbf{c}}^T)\tilde{\mathbf{u}}\} \\
&\leq \exp\{1/2+\gamma\}f^*_{\text{B}}(\tilde{\mathbf{u}}) \\
&\leq \exp\{1/2+\gamma\}\exp\{-(N-b)/2\}(N/b)^{N/2}f^*_{\text{ACG}}(\tilde{\mathbf{u}}) \qquad \text{(A.10)}
\end{aligned}$$

with $\gamma = \max(\text{eig}(\tilde{\mathbf{A}} + 1/2\tilde{\mathbf{c}}\tilde{\mathbf{c}}^T))$, $f^*_{\text{ACG}}(\tilde{\mathbf{u}}) = (\tilde{\mathbf{u}}^T\tilde{\mathbf{\Omega}}\tilde{\mathbf{u}})^{-N/2}$, $\tilde{\mathbf{\Omega}} = \mathbf{I} + (2/b)(\gamma\mathbf{I} - \tilde{\mathbf{A}} - 1/2\tilde{\mathbf{c}}\tilde{\mathbf{c}}^T)$. Given that $\tilde{\mathbf{A}} \in \mathcal{S}^+_N$, $\gamma\mathbf{I} - \tilde{\mathbf{A}} - 1/2\tilde{\mathbf{c}}\tilde{\mathbf{c}}^T$ is a symmetric matrix and its eigenvalues are positive. Finally, we have

$$f^*_{\text{vBL}}(\tilde{\mathbf{u}}) \leq M_1(b)f^*_{\text{ACG}}(\tilde{\mathbf{u}}) \qquad \text{(A.11)}$$

where $M_1(b) = \exp\{1/2+\gamma\}\exp\{-(N-b)/2\}(N/b)^{N/2}|\tilde{\mathbf{\Omega}}|^{-1/2}$. In this case, b^* satisfies the following equality

$$\sum_{i=1}^{N} \frac{1}{b^* + 2\tilde{\lambda}_i} = 1 \qquad \text{(A.12)}$$

where the set $\{\tilde{\lambda}_i\}$ defines the eigenvalues of $\gamma\mathbf{I} - \tilde{\mathbf{A}} - 1/2\tilde{\mathbf{c}}\tilde{\mathbf{c}}^T$. The **Algorithm 2** details the sampling technique of the vBL distribution.

A.2 The vector Complex Bingham Langevin (vCBL) distribution

Let us start first with defining the relation between the vCBL distribution and the vBL distribution. Based on [41], for a given complex unitary random vector $\mathbf{u} \in \mathbb{C}^N$ such that $\mathbf{u} \sim \text{vCBL}(\mathbf{c}, \mathbf{A})$ and its p.d.f. reads as

$$p_{\text{vCBL}}(\mathbf{u}) \propto \exp\left\{\text{Re}\{\mathbf{c}^H\mathbf{u}\} + \mathbf{u}^H\mathbf{A}\mathbf{u}\right\}$$

Algorithm 12: Acceptance rejection sampling scheme for the vBL distribution

input : $\tilde{\mathbf{A}}$, $\tilde{\mathbf{c}}$
output : $\tilde{\mathbf{u}}$
Compute $\gamma = \max(\text{eig}(\tilde{\mathbf{A}} + 1/2\tilde{\mathbf{c}}\tilde{\mathbf{c}}^T))$
Compute b^* satisfying (A.12)
Compute $\tilde{\mathbf{\Omega}} = \mathbf{I} + (2/b^*)(\gamma\mathbf{I} - \tilde{\mathbf{A}} - 1/2\tilde{\mathbf{c}}\tilde{\mathbf{c}}^T)$
Compute $M_1(b^*) = \exp\{1/2 + \gamma\}\exp\{-(N - b^*)/2\}(N/b^*)^{N/2}|\tilde{\mathbf{\Omega}}|^{-1/2}$
Repeat
Sample $\tilde{\mathbf{y}} \sim \mathcal{N}_N(0, \tilde{\mathbf{\Omega}}^{-1})$
Compute $\tilde{\mathbf{u}} = \tilde{\mathbf{y}}/||\tilde{\mathbf{y}}||$
Compute $f^*_{\text{ACG}}(\tilde{\mathbf{u}}) = (\tilde{\mathbf{u}}^T\tilde{\mathbf{\Omega}}\tilde{\mathbf{u}})^{-N/2}$
Compute $f^*_{\text{vBL}}(\tilde{\mathbf{u}}) = \exp\{\tilde{\mathbf{c}}^T\tilde{\mathbf{u}} + \tilde{\mathbf{u}}^T\tilde{\mathbf{A}}\tilde{\mathbf{u}}\}$
Sample $u \sim \text{Unif}(0, 1)$
Until $u < f^*_{\text{vBL}}(\tilde{\mathbf{u}})/(M_1(b^*)f^*_{\text{ACG}}(\tilde{\mathbf{u}}))$
Accept $\tilde{\mathbf{u}}$

where $\mathbf{A}$ is an Hermitian matrix and $\mathbf{c}$ is a complex vector. We denote by

$$\begin{aligned}\mathbf{u} &= \tilde{\mathbf{u}}_1 + i\tilde{\mathbf{u}}_2 \\ \mathbf{A} &= \tilde{\mathbf{A}}_1 + i\tilde{\mathbf{A}}_2 \\ \mathbf{c} &= \tilde{\mathbf{c}}_1 + i\tilde{\mathbf{c}}_2\end{aligned} \tag{A.13}$$

where $\tilde{\mathbf{u}}_1$, $\tilde{\mathbf{c}}_1$ are respectively the real parts of $\mathbf{u}$, $\mathbf{c}$ and $\tilde{\mathbf{u}}_2$, $\tilde{\mathbf{c}}_2$ are respectively the imaginary parts of $\mathbf{u}$, $\mathbf{c}$. The matrix $\tilde{\mathbf{A}}_1$ is symmetric and $\tilde{\mathbf{A}}_2$ is a skew-symmetric matrix. In the following, we aim to introduce a relation between the vBL distribution and the vCBL distribution. We notice that,

$$\begin{aligned}p_{\text{vCBL}}(\mathbf{u}) &\propto \exp\left\{\text{Re}\{\mathbf{c}^H\mathbf{u}\} + \mathbf{u}^H\mathbf{A}\mathbf{u}\right\} \\ &\propto \exp\left\{\text{Re}\{(\tilde{\mathbf{c}}_1 + i\tilde{\mathbf{c}}_2)^H(\tilde{\mathbf{u}}_1 + i\tilde{\mathbf{u}}_2)\} + (\tilde{\mathbf{u}}_1 + i\tilde{\mathbf{u}}_2)^H(\tilde{\mathbf{A}}_1 + i\tilde{\mathbf{A}}_2)(\tilde{\mathbf{u}}_1 + i\tilde{\mathbf{u}}_2)\right\} \\ &\propto \exp\left\{\tilde{\mathbf{c}}_1^T\tilde{\mathbf{u}}_1 + \tilde{\mathbf{c}}_2^T\tilde{\mathbf{u}}_2 + \tilde{\mathbf{u}}_1^T\tilde{\mathbf{A}}_1\tilde{\mathbf{u}}_1 + i\tilde{\mathbf{u}}_1^T\tilde{\mathbf{A}}_1\tilde{\mathbf{u}}_2 + i\tilde{\mathbf{u}}_1^T\tilde{\mathbf{A}}_2\tilde{\mathbf{u}}_1 - \tilde{\mathbf{u}}_1^T\tilde{\mathbf{A}}_2\tilde{\mathbf{u}}_2 - i\tilde{\mathbf{u}}_2^T\tilde{\mathbf{A}}_1\tilde{\mathbf{u}}_1\right\} \\ &\exp\left\{\tilde{\mathbf{u}}_2^T\tilde{\mathbf{A}}_1\tilde{\mathbf{u}}_2 + \tilde{\mathbf{u}}_2^T\tilde{\mathbf{A}}_2\tilde{\mathbf{u}}_1 + i\tilde{\mathbf{u}}_2^T\tilde{\mathbf{A}}_2\tilde{\mathbf{u}}_2\right\} \\ &\propto \exp\left\{\tilde{\mathbf{c}}_1^T\tilde{\mathbf{u}}_1 + \tilde{\mathbf{c}}_2^T\tilde{\mathbf{u}}_2 + \tilde{\mathbf{u}}_1^T\tilde{\mathbf{A}}_1\tilde{\mathbf{u}}_1 - \tilde{\mathbf{u}}_1^T\tilde{\mathbf{A}}_2\tilde{\mathbf{u}}_2 + \tilde{\mathbf{u}}_2^T\tilde{\mathbf{A}}_1\tilde{\mathbf{u}}_2 + \tilde{\mathbf{u}}_2^T\tilde{\mathbf{A}}_2\tilde{\mathbf{u}}_1\right\}\end{aligned} \tag{A.14}$$

Given that $\tilde{\mathbf{A}}_1$ a symmetric matrix and $\tilde{\mathbf{A}}_2$ a skew-symmetric matrix, we have:

$$\begin{aligned}\tilde{\mathbf{u}}_2^T\tilde{\mathbf{A}}_2\tilde{\mathbf{u}}_2 &= 0 \\ \tilde{\mathbf{u}}_1^T\tilde{\mathbf{A}}_2\tilde{\mathbf{u}}_1 &= 0 \\ \tilde{\mathbf{u}}_1^T\tilde{\mathbf{A}}_1\tilde{\mathbf{u}}_2 &= \tilde{\mathbf{u}}_2^T\tilde{\mathbf{A}}_1\tilde{\mathbf{u}}_1\end{aligned}$$

Then,

$$p_{\text{vCBL}}(\mathbf{u}) \propto \exp\{\tilde{\mathbf{c}}^T\tilde{\mathbf{u}} + \tilde{\mathbf{u}}^T\tilde{\mathbf{A}}\tilde{\mathbf{u}}\} \tag{A.15}$$

with

$$\tilde{\mathbf{u}}^T = [\tilde{\mathbf{u}}_1^T, \tilde{\mathbf{u}}_2^T]$$
$$\tilde{\mathbf{c}}^T = [\tilde{\mathbf{c}}_1^T, \tilde{\mathbf{c}}_2^T]$$
$$\tilde{\mathbf{A}} = \begin{bmatrix} \tilde{\mathbf{A}}_1 & -\tilde{\mathbf{A}}_2 \\ \tilde{\mathbf{A}}_2 & \tilde{\mathbf{A}}_1 \end{bmatrix}$$

Finally,

$$\mathbf{u} \sim \text{vCBL}(\mathbf{c}, \mathbf{A}) \Rightarrow \tilde{\mathbf{u}} \sim \text{vBL}(\tilde{\mathbf{c}}, \tilde{\mathbf{A}}) \tag{A.16}$$

with $\tilde{\mathbf{u}} \in \mathbb{R}^{2N}$, $\tilde{\mathbf{c}} \in \mathbb{R}^{2N}$ and $\tilde{\mathbf{A}} \in \mathcal{S}_{2N}^{+}$. The **Algorithm 3** details the generation of the unit complex random vector $\mathbf{u} \sim \text{vCBL}(\mathbf{c}, \mathbf{A})$.

Algorithm 13: The generation of the unit complex random vector $\mathbf{u} \sim \text{vCBL}(\mathbf{c}, \mathbf{A})$

input : $\mathbf{c}$, $\mathbf{A}$
output : $\mathbf{u}$
1 Compute the $2N$ real unit vector $\tilde{\mathbf{u}}$ from the vector $\mathbf{u}$
2 Compute the $2N$ real unit vector $\tilde{\mathbf{c}}$ from the vector $\mathbf{c}$
3 Compute the $2N \times 2N$ real symmetric matrix $\tilde{\mathbf{A}}$ from the matrix $\mathbf{A}$
4 Sample the real unit random vector $\tilde{\mathbf{u}} = \text{vBL}(\tilde{\mathbf{c}}, \tilde{\mathbf{A}})$
5 Sample the complex unit random vector $\mathbf{u}$ from $\tilde{\mathbf{u}}$

A.3 The matrix CGBL distribution

Using the same methodology as in [40], a Markov chain is generated which converges to $\text{CGBL}(\mathbf{C}, \{\mathbf{A}_r\})$ in such a way the random unitary matrix $\mathbf{U} \sim \text{CGBL}(\mathbf{C}, \{\mathbf{A}_r\})$. The procedure is detailed in **Algorithm 4**.

Algorithm 14: The generation of the unitary matrix $\mathbf{U} \sim \text{CGBL}(\mathbf{C}, \{\mathbf{A}_r\})$

input : $\mathbf{C}$, $\{\mathbf{A}_r\}$
output : $\mathbf{U}$
initialize: $\mathbf{U}^{(0)} \leftarrow \mathbf{U}_{\text{init}}$ `(a unitary matrix)`
1 **while** *stop criterion unreached* **do**
2 **for** $r \in \{1, ..., R\}$ *in random order* **do**
3 Compute the null space $\mathbf{N}$ of the matrix $\mathbf{U}_{[,-r]}$
4 Compute the unit vector $\mathbf{u} = \mathbf{N}^H \mathbf{U}_{[,r]}$
5 Compute $\bar{\mathbf{c}} = \kappa \mathbf{N}^H \mathbf{c}_r$ and $\bar{\mathbf{A}} = \mathbf{N}^H \mathbf{A}_r \mathbf{N}$
6 Update the complex unit vector $\mathbf{u} = \text{vCBL}(\bar{\mathbf{c}}, \bar{\mathbf{A}})$
7 Update $\mathbf{u}_r = \mathbf{N}\mathbf{u}$
8 **end**
9 **end**

Appendix B

MM algorithm

B.1 Block Majorization-Minimization algorithm

To solve further-coming optimization problems, we adopt the block MM algorithm framework, which is briefly stated below. For more complete information, we refer the reader to [75]. Consider the following problem:

$$\begin{aligned} \underset{\boldsymbol{\theta}}{\text{minimize}} \quad & f\left(\boldsymbol{\theta}\right) \\ \text{subject to} \quad & \boldsymbol{\theta} \in \boldsymbol{\Theta}, \end{aligned} \tag{B.1}$$

where the optimization variable $\boldsymbol{\theta}$ can be partitioned into m blocks as $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m)$, with each n_i-dimensional block $\boldsymbol{\theta}_i \in \boldsymbol{\Theta}_i$ and $\boldsymbol{\Theta} = \prod_{i=1}^{m} \boldsymbol{\Theta}_i$. At the $(t+1)$-th iteration, the i-th block $\boldsymbol{\theta}_i$ is updated by solving the following problem:

$$\begin{aligned} \underset{\boldsymbol{\theta}_i}{\text{minimize}} \quad & g_i\left(\boldsymbol{\theta}_i | \boldsymbol{\theta}^{(t)}\right) \\ \text{subject to} \quad & \boldsymbol{\theta}_i \in \boldsymbol{\Theta}_i, \end{aligned} \tag{B.2}$$

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

$$\begin{aligned} f\left(\boldsymbol{\theta}^{(t)}\right) &= g_i\left(\boldsymbol{\theta}_i^{(t)} | \boldsymbol{\theta}^{(t)}\right), \\ f\left(\boldsymbol{\theta}_1^{(t)}, \ldots, \boldsymbol{\theta}_i, \ldots, \boldsymbol{\theta}_m^{(t)}\right) &\leq g_i\left(\boldsymbol{\theta}_i | \boldsymbol{\theta}^{(t)}\right) \ \forall \boldsymbol{\theta}_i \in \boldsymbol{\Theta}_i, \\ f'\left(\boldsymbol{\theta}^{(t)}; \mathbf{d}_i^0\right) &= g_i'\left(\boldsymbol{\theta}_i^{(t)}; \mathbf{d}_i | \boldsymbol{\theta}^{(t)}\right) \\ & \quad \forall \boldsymbol{\theta}_i^{(t)} + \mathbf{d}_i \in \boldsymbol{\Theta}_i, \\ & \quad \mathbf{d}_i^0 \triangleq (\mathbf{0}; \ldots; \mathbf{d}_i; \ldots; \mathbf{0}), \end{aligned}$$

where $f'(\boldsymbol{\theta}; \mathbf{d})$ stands for the directional derivative at $\boldsymbol{\theta}$ 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 other blocks fixed.

B.2 Surrogates and updates for the algorithms of Part II

In the following, we derive two propositions needed for the algorithms design. These propositions are generic. Specifically, the first proposition will be useful to derive updates w.r.t. positive

scaling factors and eigenvalues of the structured covariance matrices. The second proposition will be useful for the update of eigenvectors.

proposition B.1 *Let* $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_R] \in \mathcal{U}_N^R$, $\mathbf{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_R] \in \mathbb{C}^{N \times R}$ *and* $\{\mathbf{Y}_r\} \subset \mathcal{H}_N^+$. *The function:*

$$f(\mathbf{U}) = \sum_{r=1}^{R} Re\{\mathbf{q}_r^H \mathbf{u}_r\} + \mathbf{u}_r^H \left[\mathbf{Y}_r\right] \mathbf{u}_r \tag{B.3}$$

is lower bounded at $\mathbf{U}^t$ *as:*

$$\begin{aligned} f(\mathbf{U}) &\geq \sum_{r=1}^{R} \mathbf{u}_r^H(\mathbf{Y}_r \mathbf{u}_r^t + 1/2\mathbf{q}_r) + (\mathbf{u}_r^{t^H} \mathbf{Y}_r + 1/2\mathbf{q}_r^H)\mathbf{u}_r + const \\ &\geq Tr\{\mathbf{U}^H \mathbf{H}^t\} + Tr\{\mathbf{H}^{t^H} \mathbf{U}\} + const \\ &\geq -||\mathbf{U} - \mathbf{H}^t||_F^2 + const \end{aligned}$$

with equality when $\mathbf{U} = \mathbf{U}^t = [\mathbf{u}_1^t, \ldots, \mathbf{u}_R^t]$ *and* $\mathbf{H}^t = 1/2\mathbf{Q} + [\mathbf{Y}_1 \mathbf{u}_1^t, \ldots, \mathbf{Y}_R \mathbf{u}_R^t]$. *The surrogate function reads as:*

$$f(\mathbf{U}|\mathbf{U}^t) = -||\mathbf{U} - \mathbf{H}^t||_F^2 \tag{B.4}$$

Maximizing the above function under unitary constraints is equivalent to solve:

$$\begin{aligned} &\underset{\widehat{\mathbf{U}}}{minimize} \quad ||\mathbf{U} - \mathbf{H}^t||_F^2 \\ &subject\ to \quad \widehat{\mathbf{U}}^H \widehat{\mathbf{U}} = \mathbf{I}_R \end{aligned} \tag{B.5}$$

which is an orthogonal Procrustes problem [69] that has a unique solution given as:

$$\widehat{\mathbf{U}}^{t+1} = \mathcal{P}_{Proc}(\mathbf{H}^t) \tag{B.6}$$

where the projection onto the set $\mathcal{U}_N^R$ *is denoted by the operator:*

$$\begin{aligned} \mathcal{P}_{\text{Proc}}: \ \mathbb{C}^{N \times R} &\longrightarrow \mathcal{U}_N^R \\ \mathbf{Z} \overset{\text{TSVD}}{=} \mathbf{U}\mathbf{D}\mathbf{V}^H &\longmapsto \mathcal{P}_{\text{Proc}}\{\mathbf{Z}\} = \mathbf{U}\mathbf{V}^H \end{aligned} \tag{B.7}$$

with $\overset{\text{TSVD}}{=}$ *defines the thin-singular value decomposition of a given matrix.*

proposition B.2 *Let us consider* a, $\{b_p\}$ *and* $\{s_p\}$ *where* $a > 0$, $b_p > 0$ *and* $s_p > 0$, $\forall p \in [\![1, P]\!]$. *The objective function*

$$g(a) = \sum_{p=1}^{P} \left(ln(ab_p + \sigma^2) - \frac{ab_p s_p}{ab_p + \sigma^2} \right) \tag{B.8}$$

is upper bounded by

$$g(a) \leq Aln(Ba + C) - Dln(a) \tag{B.9}$$

with

$$\begin{cases} \theta_p^t = 1 + s_p \dfrac{a^t b_p}{a^t b_p + \sigma^2} & A = \displaystyle\sum_{i=1}^{I} \theta_p^t \\ B = \dfrac{\sum_{p=1}^{P} \frac{\theta_p^t b_p}{b_p a^t + \sigma^2}}{\sum_{p=1}^{P} \theta_p^t} & C = \sigma^2 \dfrac{\sum_{p=1}^{P} \frac{\theta_p^t}{b_p a^t + \sigma^2}}{\sum_{p=1}^{P} \theta_p^t} \\ D = \displaystyle\sum_{p=1}^{P} s_p \frac{a^t b_p}{a^t b_p + \sigma^2} \end{cases}$$

then, the surrogate function reduces to

$$g(a|a^t) = Aln(Ba + C) - Dln(a) \tag{B.10}$$

with equality at $a = a^t$*. The minimizer of the above function under positivity constraint is given as*

$$a^{t+1} = \frac{DC}{B(A-D)} = \frac{1}{P}\frac{\left(\sum\limits_{p=1}^{P} s_p \frac{a^t b_p}{a^t b_p + \sigma^2}\right)\left(\sum\limits_{p=1}^{P} \sigma^2 \frac{\theta_p^t}{a^t b_p + \sigma^2}\right)}{\sum\limits_{p=1}^{P} \frac{\theta_p^t b_p}{a^t b_p + \sigma^2}} \tag{B.11}$$

Proof: The proof of Propositions B.1 and B.2 are similar to [69].

B.3 Surrogates and updrates for the detectors of Part III

In the following, we derive two propositions needed for the algorithms design. Specifically, the first proposition will be useful to derive updates w.r.t. positive scaling factors and eigenvalues of the structured covariance matrices. The second proposition will be useful for the update of eigenvectors.

proposition B.3 *Let* $\tau_i \in \mathbb{R}^+$*, and sets of real parameters* $\{\lambda_r\}$ *with* $\lambda_r \geq 0,\ \forall r \in [\![1, R]\!]$, $\{s_{k,i}^r\}$ *with* $s_{k,i}^r \geq 0,\ \forall i \in [\![1, I]\!],\ \forall r \in [\![1, R]\!],\ \forall k \in [\![1, K_i]\!]$*, and* $\sigma^2 > 0$*. The function*

$$\mathcal{L}_i(\tau_i) = \sum_{r=1}^{R}\left[K_i \ln\left(\tau_i\lambda_r + \sigma^2\right) - \sum_{k=1}^{K_i} \frac{\tau_i\ \lambda_r\ s_{k,i}^r}{\tau_i\lambda_r + \sigma^2}\right] \tag{B.12}$$

is upper bounded at the point $\tau_i^{(t)}$ *as*

$$\mathcal{L}_i(\tau_i|\tau_i^{(t)}) \leq A_i \ln\left(B_i\tau_i + C_i\right) - D_i \ln\left(\tau_i\right) \tag{B.13}$$

where

$$\begin{cases} \gamma_i^r = \left(K_i + \sum\limits_{k=1}^{K_i} s_{k,i}^r \frac{\tau_i^{(t)}\lambda_r}{\tau_i^{(t)}\lambda_r + \sigma^2}\right) & \beta_i^r = \sum\limits_{k=1}^{K_i} s_{k,i}^r \frac{\tau_i^{(t)}\lambda_r}{\tau_i^{(t)}\lambda_r + \sigma^2} & A_i = \sum\limits_{r=1}^{R}\gamma_i^r \\ B_i = \frac{\sum_{r=1}^{R} \frac{\gamma_i^r\lambda_r}{\tau_i^{(t)}\lambda_r + \sigma^2}}{\sum_{r=1}^{R}\gamma_i^r} & C_i = \frac{\sum_{r=1}^{R} \frac{\gamma_i^r\sigma^2}{\tau_i^{(t)}\lambda_r + \sigma^2}}{\sum_{r=1}^{R}\gamma_i^r} & D_i = \sum\limits_{r=1}^{R}\beta_i^r \end{cases} \tag{B.14}$$

with equality at $\tau_i = \tau_i^{(t)}$*. The above surrogate function has a unique non-negative minimum that leads to the following MM update:*

$$\tau_i^{(t+1)} = \frac{D_i C_i}{(A_i - D_i)B_i} \tag{B.15}$$

Proof. The first part of the proof follows the lines of Proposition 1 in [45]. First, we construct an inequality by the first order Taylor expansion of ln as:

$$-\sum_{k=1}^{K_i} \frac{\tau_i\ \lambda_r\ s_{k,i}^r}{\tau_i\lambda_r + \sigma^2} \leq -\sum_{k=1}^{K_i} s_{k,i}^r \frac{\tau_i^{(t)}\lambda_r}{\tau_i^{(t)}\lambda_r + \sigma^2}\left[1 + \ln(\tau_i\lambda_r) - \ln(\tau_i\lambda_r + \sigma^2)\right] + \text{const.} \tag{B.16}$$

The cost function $\mathcal{L}_i$ is therefore majorized by a first surrogate as:

$$\mathcal{L}_i(\tau_i|\tau_i^{(t)}) \leq \sum_{r=1}^{R} \left[\left(K_i + \sum_{k=1}^{K_i} s_{k,i}^r \frac{\tau_i^{(t)}\lambda_r}{\tau_i^{(t)}\lambda_r + \sigma^2} \right) \ln\left(\tau_i\lambda_r + \sigma^2\right) - \left(\sum_{k=1}^{K_i} s_{k,i}^r \frac{\tau_i^{(t)}\lambda_r}{\tau_i^{(t)}\lambda_r + \sigma^2} \right) \ln(\tau_i) \right] \tag{B.17}$$

that reads

$$\mathcal{L}_i(\tau_i) \leq \sum_{r=1}^{R} \left[\gamma_i^r \ \ln\left(\tau_i\lambda_r + \sigma^2\right) \ - \ \beta_i^r \ \ln(\tau_i) \right] \tag{B.18}$$

with γ_i^r and β_i^r in (B.14). Second, thanks to the Jensen's inequality, we have:

$$\sum_{r=1}^{R} \gamma_i^r \ \ln\left(\tau_i\lambda_r + \sigma^2\right) \leq \left(\sum_{r=1}^{R} \gamma_i^r \right) \ln\left(\frac{\sum_{r=1}^{R} \gamma_i^R \frac{\tau_i\lambda_r + \sigma^2}{\tau_i^{(t)}\lambda_r + \sigma^2}}{\sum_{r=1}^{R} \gamma_i^r} \right) + \text{const.} \tag{B.19}$$

that splits into

$$\sum_{r=1}^{R} \gamma_i^r \ \ln\left(\tau_i\lambda_r + \sigma^2\right) \leq \left(\sum_{r=1}^{R} \gamma_i^r \right) \ln\left(\frac{\sum_{r=1}^{R} \frac{\gamma_i^r\lambda_r}{\tau_i^{(t)}\lambda_r + \sigma^2}}{\sum_{r=1}^{R} \gamma_i^r} \tau_i + \frac{\sum_{r=1}^{R} \frac{\gamma_i^r\sigma_n^2}{\tau_i^{(t)}\lambda_r + \sigma^2}}{\sum_{r=1}^{R} \gamma_i^r} \right) + \text{const.} \tag{B.20}$$

Finally, we have the second surrogate function

$$\mathcal{L}_i(\tau_i) \leq A_i \ln\left(B_i\tau_i + C_i\right) - D_i \ln\left(\tau_i\right) \tag{B.21}$$

with A_i, B_i, C_i and D_i in (B.14). The proof concludes by directly applying Proposition 2 of [45], that states that the above majorizer is quasi convex and has a unique minimizer given in (B.15). ■

proposition B.4 *Let $\mathbf{V} \in \mathbb{C}^{M\times R}$ be a unitary matrix, $\{\mathbf{B}_i\}$, with $\mathbf{B}_i \in \mathcal{H}_R^{++}$, $\forall i \in [\![1, I]\!]$ and $\{\mathbf{A}_i\}$, with $\mathbf{A}_i \in \mathcal{H}_M^{++}$, $\forall i \in [\![1, I]\!]$. The function*

$$f(\mathbf{V}) = -\sum_{i=1}^{I} \text{Tr}\{\mathbf{V}^H\mathbf{A}_i\mathbf{V}\mathbf{B}_i\} \tag{B.22}$$

is upper bounded at point $\mathbf{V}^{(t)}$ as

$$f(\mathbf{V}|\mathbf{V}^{(t)}) \leq -\sum_{i=1}^{I} \Big[\text{Tr}\{\mathbf{V}^H\mathbf{A}_i\mathbf{V}^{(t)}\mathbf{B}_i\} + \text{Tr}\{\mathbf{V}^{(t)H}\mathbf{A}_i\mathbf{V}\mathbf{B}_i\} \Big] + \text{const.} \tag{B.23}$$

with equality at $\mathbf{V}^{(t)} = \mathbf{V}$. The above surrogate function has a unique minimum on the set of unitary matrices, that leads to the following MM update

$$\mathbf{V}^{(t+1)} = \mathbf{U}_{\text{left}}\mathbf{U}_{\text{right}}^H \tag{B.24}$$

where $\mathbf{U}_{\text{left}}$ and $\mathbf{U}_{\text{right}}^H$ are the left and right eigenvectors of the thin-SVD

$$\sum_{i=1}^{I} (\mathbf{A}_i\mathbf{V}^{(t)}\mathbf{B}_i) \overset{\text{TSVD}}{=} \mathbf{U}_{\text{left}}\mathbf{D}\mathbf{U}_{\text{right}}^H \tag{B.25}$$

Proof. The function f in (B.22) is concave so it can be majorized by its first order Taylor expansion, which is the surrogate given in (B.23). Minimizing (B.23) under unitary constraints is equivalent to solve

$$\begin{array}{ll} \underset{\mathbf{V}}{\text{minimize}} & \left\| \left(\sum_{i=1}^{I} \mathbf{A}_i \mathbf{V}^{(t)} \mathbf{B}_i \right) - \mathbf{V} \right\|_F^2 \\ \text{subject to} & \mathbf{V}^H \mathbf{V} = \mathbf{I} \end{array} \tag{B.26}$$

which is an orthogonal Procrustes problem [45] that has a unique solution given in (B.24). ■

Appendix C

Résumé étendu

C.1 Introduction

Les systèmes d'estimation/détection de signaux doivent souvent fonctionner dans des environnements difficiles à caractériser avec des modèles statistiques précis. Il est donc intéressant de développer des traitements peu sensibles aux écarts entre les observations et les modèles supposés. Au cours des dernières décennies, l'estimation des paramètres d'intérêt à partir d'un ensemble de données bruitées est resté un sujet de recherche actif, car il soulève encore de nombreuses difficultés:

• Tout d'abord, un des problèmes difficiles est celui de l'estimation des paramètres d'intérêt dans le contexte des données de grande dimension (c'est-à-dire, dans le cas où la dimension des données est proche ou supérieure à la taille de l'échantillon). En guise de solution, l'approche de rang faible offre la possibilité de traiter de manière robuste et efficace des données de grande dimension. En effet, dans de nombreux cas, les informations pertinentes résident souvent dans un sous-espace de dimension beaucoup plus faible que l'espace ambiant. Ainsi, l'objectif de nombreux algorithmes peut être interprété, au sens large, comme d'exploiter cette structure présente dans les données. Un exemple de structure récurrente en traitement du signal est celui d'une matrice de covariance de rang faible, apparaissant lorsque le signal est contenu dans un sous-espace.

• En second lieu se pose le problème du choix de modèle statistique à apposer aux données. L'un des modèle le plus utilisé est le modèle gaussien, qui n'est cependant pas toujours réaliste. Ce choix peut donc nuire aux performances des traitements développés lorsque les données sont de nature impulsives, ou présentent des valeurs aberrantes. Ainsi, le modèle choisi doit être robuste à l'éventualité de signaux non gaussiens. À cette fin, nous considérons l'utilisation du modèle gaussien composé, généralement choisi pour sa flexibilité et pour son adéquation avec des jeux de données réelles. Cette modélisation inclut de nombreux cas particuliers usuels, comme par exemple, les cas gaussiens, Student-t, ou encore K-distribués.

Les contributions de cette thèse se concentrent autour de deux axes:

• Nous examinons d'abord le problème de l'estimation du sous-espace signal dans un contexte bayésien. En effet, l'ajout d'une information a priori sur le sous-espace signal peut améliorer le processus d'estimation dans certains cas critiques, comme par exemple, un faible rapport signal sur bruit et/ou un faible support d'échantillons. Nous nous concentrons sur le cas de signaux gaussiens composés noyés dans le bruit blanc gaussien. Cette modélisation a été largement utilisée dans plusieurs applications de traitement du signal pour sa robustesse à la présence de valeurs

aberrantes, éventuellement causée par un bruit impulsif. Suivant le cadre bayésien, nous dérivons deux algorithmes pour calculer respectivement l'estimateur maximum a posteriori et l'estimateur dit du *minimum de distance quadratique moyenne*. Dans ce contexte, nous introduisons également une version généralisée de la distribution de Bingham-Langevin afin de modéliser l'a priori sur la base orthonormée du sous-espace signal.

• En second lieu, nous étudions le problème de savoir si un jeu de données sous test partage des propriétés communes avec un ensemble de données secondaires. À cette fin, nous proposons de nouvelles statistiques basées sur le test du similarité entre les composantes principales des jeux de données. Dans ce contexte, nous formalisions plusieurs tests de rapport de vraisemblance généralisé basés sur des matrice de covariance à structure rang faible. Les intérêts de cette approche dans des applications telles que la détection de changement sont multiples: *i*) Incorporer une structure de rang faible un test d'égalité de matrice de covariance peut offrir une amélioration performances de détection. *ii*) L'hypothèse de rang faible permet d'augmenter la résolution spatiale du processus de détection car ce dernier peut alors être mis en œuvre avec moins d'échantillons.

C.2 Estimation bayésienne de sous-espace signal en présence de sources gaussiennes composées

C.2.1 Introduction

L'estimation de sous-espace est un problème omniprésent en traitement du signal. Le but étant d'estimer le sous-espace, de généralement de rang faible, où se trouve l'information générée par les signaux d'intérêt. Ceci représente une étape clé dans de nombreuses applications, comme l'estimation des directions d'arrivée [1], l'annulation d'interférence [3, 48–50], etc.

Nous considérons un modèle linéaire largement utilisé dans la littérature : la matrice de données est modélisée sous la forme $\mathbf{Z} = \mathbf{US} + \mathbf{N}$, où $\mathbf{U}$ représente la base du sous-espace d'intérêt, $\mathbf{S}$ est le signal d'intérêt et $\mathbf{N}$ est un bruit additif. Le problème adressé est d'estimer $\mathbf{U}$ à partir des observations $\mathbf{Z}$. La méthode la plus courante est d'estimer le sous-espace $\mathbf{U}$ par décomposition en valeurs singulières (DVS) de la *sample covariance matrix* (SCM), c'est-à-dire au travers des vecteurs propres prédominants de la SCM. Ceci fournit généralement un estimateur précis de l'espace $\mathcal{R}(\mathbf{U})$ (généré par $\mathbf{U}$) pour un rapport signal/bruit (RSB) élevé et/ou pour un grand nombre d'échantillons. Cependant, cette méthode montre ses limites en dehors de ces régimes asymptotiques. Une solution serait d'intégrer certaines connaissances a priori dans le processus d'estimation, comme proposé dans [66] et [68]. Par exemple, le cadre bayésien permet de dériver des estimateurs tels que l'EMDM, qui vise à minimiser la distance entre la vraie matrice de projection $\mathbf{UU}^H$ et son estimée $\widehat{\mathbf{U}}\widehat{\mathbf{U}}^H$ [66, 68]. Plus particulièrement, dans [66], l'EMDM est proposé pour différents modèles où l'a priori sur $\mathbf{S}$ est uniforme. Nous étendons le travail de [66] pour un modèle linéaire où $\mathbf{S}$ suit une loi gaussienne composée et $\mathbf{N}$ est un bruit blanc gaussien. Ce choix est motivé par le fait que la distribution gaussienne composée a été adoptée dans de nombreuses applications de traitement de signal robustes, car elle peut prendre en compte les fluctuations de puissance locales et présente un bon accord avec plusieurs ensembles de données réelles [16, 17]. Notons que cette famille couvre un large panel de distributions telles que Student *t*-, K- et Weibull (cf. [16] et références dans celles-ci). Par conséquent, le modèle considéré peut décrire avec précision un fouillis (où l'on tient compte des fluctuations de puissance des sources) perturbé par un bruit thermique. A titre d'exemple, ce modèle a été utilisé pour la détection en environnement hétérogène [32, 104] et pour l'estimation de matrice de covariance dans [30, 34].

En particulier, concernant le problème d'estimation de sous-espace, [19,69] a proposé des algorithmes maximum de vraisemblance (MV) pour ce contexte, et des bornes limites sur la précision d'estimation ont été dérivées dans [71]. Cependant, ces études n'ont jamais été conduites dans un contexte bayésien, en ce sens que ils n'ont pas apposé d'a priori sur le sous-espace d'intérêt.

Afin de combler cette lacune, nous dérivons de nouveaux estimateurs bayésiens dans le contexte des sources suivant la loi gaussienne composée, noyées dans un bruit blanc additif gaussien. Tout d'abord, notre développement nécessite d'étendre les distributions utilisées dans [66] pour des données avec des entrées complexes. Dans ce but, nous introduisons une généralisation de la distribution réelle de Bingham-Langevin (également appelée Bingham-von-Mises-Fisher). Nous proposons également une méthode de génération des matrices semi-unitaires adaptée cette distribution (en appendice A du manuscrit complet). Deuxièmement, nous développons un algorithme pour calculer l'estimateur du maximum a posteriori (MAP) pour le modèle considéré, basé sur l'algorithme *maximization minimization* (MM) [45]. Troisièmement, nous dérivons un algorithme d'échantillonneur de Gibbs pour calculer l'estimateur EMDM, qui suit la formalisation de [66]. Dans le chapitre II.4 du manuscrit complet, nous étudions numériquement les performances des l'estimateurs proposé, et illustrons leur applicabilité sur une problématique de détection STAP (par un radar aéroporté) avec des données réelles.

C.2.2 Ètat de l'art

C.2.2.1 La loi gaussienne composée

La distribution gaussienne composée est un outil reconnu dans la littérature de traitement de signal robuste [12]. Ce modèle est polyvalent, car il englobe de nombreuses distributions usuelles telles que les distributions gaussienne, Student t-, K- et Weibull. Une observation qui suit une loi gaussienne composée (CG) de dimension N est représentée comme un produit de deux composants statistiquement indépendants. Plus précisément, si $\mathbf{s} \in \mathbb{C}^N$ suit une distribution CG centrée, notée $\mathbf{s} \sim \mathcal{CG}(\mathbf{0}, \mathbf{\Sigma}, f_\tau)$, il a la représentation stochastique suivante :

$$\mathbf{s} \overset{d}{=} \sqrt{\tau}\mathbf{d}, \tag{C.1}$$

où

i) τ est un scalaire positif, appelé texture, ayant comme une fonction de densité de probabilité f_τ. Ce paramètre est statistiquement indépendant de $\mathbf{d}$. En fonction de f_τ, nous pouvons obtenir diverses distributions multivariées standard pour $\mathbf{s}$ [12]. Afin de concevoir des algorithmes robustes à toutes ces distributions, nous considérons ici ce paramètre comme déterministe inconnu pour chaque réalisation. Cette distribution sera donc notée $\mathbf{s}_k \sim \mathcal{CG}(0, \mathbf{\Sigma}, \tau_k)$ pour chaque observation $k \in [\![1, K]\!]$. Nous notons également $\boldsymbol{\tau}$ le vecteur qui agrège les paramètres $\{\tau_k\}$ pour un ensemble donné d'observations $\{\mathbf{s}_k\}$.

ii) $\mathbf{d}$ suit une distribution gaussienne complexe multivariée de moyenne nulle et de matrice de covariance $\mathbf{\Sigma}$, notée, $\mathbf{d} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma})$. Le paramètre $\mathbf{\Sigma} \in \mathcal{H}_N^+$ est appelé matrice de dispersion. Notons que si $\mathbb{E}\{\tau\} < \infty$, la matrice de covariance de $\mathbf{s}$ existe et est proportionnelle à la matrice de dispersion, c'est-à-dire $\mathbb{E}\{\mathbf{s}\mathbf{s}^H\} = \mathbb{E}\{\tau\}\mathbf{\Sigma}$.

C.2.2.2 La distribution complexe Bingham-Langevin généralisée (CBLG)

Afin de modéliser les a priori sur les sous-espaces, nous nous intéressons dans la suite à la distribution par rapport à l'ensemble $\mathcal{U}_N^R$. Parmi les distributions les plus utilisées sur $\mathcal{U}_N^R$ figurent les distributions de Bingham et de Langevin [35,40,41]. Nous présentons la distribution CBLG comme

une généralisation des statistiques directionnelles habituelles susmentionnées au cas de variables matricielles à entrées complexes. La CBLG est une distribution de probabilité sur l'ensemble des matrices semi-unitaires qui combine des termes linéaires et quadratiques, paramétrés par un ensemble de matrices $\{\mathbf{A}_r\} \subset \mathcal{H}_N^+$ et la matrice $\mathbf{C}$. Nous notons $\mathbf{U} \sim \text{CBLG}(\mathbf{C}, \{\mathbf{A}_r\}) \in \mathbb{C}^{N \times R}$ lorsque la densité de probabilité de $\mathbf{U}$ sur $\mathcal{U}_N^R$ est donnée par

$$p_{\text{CBLG}}(\mathbf{U}) \propto \exp\left\{ \sum_{r=1}^{R} \text{Re}\{\mathbf{c}_r^H \mathbf{u}_r\} + \mathbf{u}_r^H \mathbf{A}_r \mathbf{u}_r \right\}, \tag{C.2}$$

où $\mathbf{c}_r$ et $\mathbf{u}_r$ représentent le r^{me} vecteur colonne de, respectivement, $\mathbf{C}$ et $\mathbf{U}$.

Remarque C.2.1 *Dans* (C.2), p_{CBLG} *favorise la concentration de chaque vecteur* $\mathbf{u}_r$ *autour de* $\mathbf{c}_r$ *et* $\mathbf{u}_r\mathbf{u}_r^H$ *autour du sous-espace associé aux valeurs propres les plus fortes de la matrice Hermitienne* $\mathbf{A}_r$*. En guise d'exemple, si* $\mathbf{A}_r = \mathbf{A}$ *et* $\mathbf{c}_r = \mathbf{0}$*,* $\forall r \in [\![1, R]\!]$*, l'espace engendré par* $\mathbf{U}\mathbf{U}^H$ *tend à être proche de l'espace propre dominant de* $\mathbf{A}$ *pour chaque réalisation.*

C.2.3 Modèle des observations

Les données sont modélisées comme une somme de sources gaussienne composées, résidant dans un sous-espace de rang faible, incorporées dans un bruit gaussien blanc. Cette formulation est utile pour modéliser le fouillis ainsi que le bruit thermique dans plusieurs applications de traitement de signal, telles que radar [19,32,33,69,71]. Pour ce modèle, les échantillons $\{\mathbf{z}_k\}_{k=1}^K$ (les colonnes de $\mathbf{Z}$) sont représentés sous la forme suivante

$$\mathbf{z}_k = \mathbf{s}_k + \mathbf{n}_k \tag{C.3}$$

où

• $\mathbf{s}_k \sim \mathcal{CG}(\mathbf{0}, \mathbf{\Sigma}, \tau_k)$ réside dans un sous-espace de rang $R \ll N$. De plus, la matrice de dispersion est paramétrée par sa DVS

$$\mathbf{\Sigma} \overset{\text{DVS}}{=} \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H \tag{C.4}$$

i) L'ensemble des scalaires $\{\tau_k\}$ représente les textures qui sont supposées être positives et déterministes inconnues.

ii) $\mathbf{\Lambda} = \text{diag}(\{\lambda_r\}) \in \mathbb{R}^{R \times R}$ est une matrice diagonale contenant les valeurs propres de la matrice de dispersion, supposées positives et déterministes inconnues.

iii) $\mathbf{U} \in \mathcal{U}_N^R$ sont les vecteurs propres de la matrice de dispersion, dont les colonnes représentent une base orthonormée du sous-espace du signal. Cette base suit la distribution $\mathbf{U} \sim \text{CBLG}(\mathbf{C}, \{\mathbf{A}_r\})$.

• $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_N)$ est un bruit blanc additif Gaussien de variance connue (ou pré-estimée) σ^2.

Remarque C.2.2 *Dans ces travaux, nous considérons un modèle hybride bayésien car notre principal intérêt est d'intégrer une connaissance préalable du sous-espace du signal dans le processus d'estimation. Inversement, nous choisissons de ne pas spécifier la densité de probabilité des paramètres de texture* $\{\tau_k\}$ *(et des valeurs propres* $\{\lambda_r\}$*), supposés ici inconnus et déterministes. Ce faisant, nous garantissons une robustesse du processus pour toute distribution sous-jacente de ces paramètres. De plus, cette hypothèse permet également de proposer des méthodes tractables : l'inclusion d'une distribution préalable sur* $\{\tau_k\}$ *dans le modèle considéré conduit en effet à des fonctions intégrales complexes à manipuler et calculer en pratique [72]. Dans ce qui suit, par souci de concision, l'estimateur hybride bayésien MV-EMDM (respectivement MV-MAP) sera simplement appelé EMDM (respectivement MAP) par abus de langage.*

En désignant

$$\boldsymbol{\Sigma}_k = \tau_k \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H + \sigma^2\mathbf{I}_N \quad \forall k \in [\![1, K]\!] \tag{C.5}$$

nous avons pour chaque échantillon la représentation conditionnelle $\mathbf{z}_k|\mathbf{U}, \boldsymbol{\Lambda}, \tau_k \sim \mathcal{CN}(0, \boldsymbol{\Sigma}_k)$, conduisant à la densité de probabilité conditionnelle de la série d'échantillons $\mathbf{Z}$

$$p(\mathbf{Z}|\mathbf{U}, \{\lambda_r\}, \{\tau_k\}) = \prod_{k=1}^{K} p(\mathbf{z}_k|\mathbf{U}, \{\lambda_r\}, \tau_k) \propto \prod_{k=1}^{K} \frac{\exp\left\{-\mathbf{z}_k^H \boldsymbol{\Sigma}_k^{-1}\mathbf{z}_k\right\}}{\det(\boldsymbol{\Sigma}_k)} \tag{C.6}$$

Grâce au lemme de Sherman-Morrison-Woodbury [73, 74], l'expression de $\boldsymbol{\Sigma}_k^{-1}$ est simplifiée comme

$$\boldsymbol{\Sigma}_k^{-1} = \sigma^{-2}\mathbf{I} - \mathbf{U}\boldsymbol{\Gamma}_k\mathbf{U}^H \tag{C.7}$$

où $\boldsymbol{\Gamma}_k = \sigma^{-2}\mathbf{I}_R - (\tau_k\boldsymbol{\Lambda} + \sigma^2\mathbf{I}_R)^{-1}$ est une matrice diagonale d'entrées

$$[\boldsymbol{\Gamma}_k]_{r,r} = \gamma_{k,r} = \frac{\tau_k\lambda_r}{\sigma^2(\tau_k\lambda_r + \sigma^2)} \tag{C.8}$$

À partir de (C.7) et (C.6), l'expression de la probabilité postérieure de $\mathbf{U}$ est redonnée par

$$\begin{aligned}
p(\mathbf{U}|\mathbf{Z}, \{\tau_k\}, \{\lambda_r\}) &\propto p(\mathbf{Z}|\mathbf{U}, \{\tau_k\}, \{\lambda_r\}) p_{\mathrm{CBLG}}(\mathbf{U}) \propto \prod_{k=1}^{K} \frac{\exp\left\{-\mathbf{z}_k^H \left(\boldsymbol{\Sigma}_k\right)^{-1} \mathbf{z}_k\right\}}{\det(\boldsymbol{\Sigma}_k)} p_{\mathrm{CBLG}}(\mathbf{U}) \\
&\propto \prod_{k=1}^{K}\left(\prod_{r=1}^{R} \frac{1}{\tau_k\lambda_r + \sigma^2}\right) \exp\left\{-\mathbf{z}_k^H(-\mathbf{U}\boldsymbol{\Gamma}_k\mathbf{U}^H + \sigma^{-2}\mathbf{I}_N)\mathbf{z}_k\right\} p_{\mathrm{CBLG}}(\mathbf{U}) \\
&\propto \prod_{k=1}^{K}\left(\prod_{r=1}^{R} \frac{1}{\tau_k\lambda_r + \sigma^2}\right) \exp\left\{\sum_{k=1}^{K} \mathbf{z}_k^H\mathbf{U}\boldsymbol{\Gamma}_k\mathbf{U}^H\mathbf{z}_k\right\} p_{\mathrm{CBLG}}(\mathbf{U}) \\
&\propto \left(\prod_{k=1}^{K}\prod_{r=1}^{R} \frac{1}{\tau_k\lambda_r + \sigma^2}\right) \exp\left\{\sum_{r=1}^{R} \mathbf{u}_r^H\mathbf{M}_r\mathbf{u}_r\right\} p_{\mathrm{CBLG}}(\mathbf{U}) \\
&\propto \left(\prod_{k=1}^{K}\prod_{r=1}^{R} \frac{1}{\tau_k\lambda_r + \sigma^2}\right) \exp\left\{\sum_{r=1}^{R} \mathrm{Re}\{\mathbf{c}_r^H\mathbf{u}_r\} + \mathbf{u}_r^H\left[\mathbf{A}_r + \mathbf{M}_r\right]\mathbf{u}_r\right\}
\end{aligned} \tag{C.9}$$

avec

$$\mathbf{M}_r = \sum_{k=1}^{K} \gamma_{k,r}\mathbf{z}_k\mathbf{z}_k^H \tag{C.10}$$

où l'expression de $\gamma_{k,p}$ est donnée dans (C.8).

Notre objectif est de développer des estimateurs bayésiens de la base orthonormale de sous-espace $\mathbf{U}$ pour le modèle de données (C.3). La première proposition est un algorithme majoration-minimisation (MM) pour calculer l'estimateur MAP. Le second est un algorithme permettant d'évaluer l'EMDM au moyen d'itérations MM et d'un échantillonneur de Gibbs. De plus, nous présentons un cas spécial, appelé "modèle simplifié", pour lequel les estimateurs MAP et EMDM coïncident et peuvent être calculés avec un faible cout. Les propriétés de chaque méthode sont énumérées ci-dessous :

- Théoriquement, l'approche EMDM offre les meilleures performances en termes de distance moyenne entre les matrices de projection des sous-espace estimé et réel. Néanmoins, l'évaluation de l'estimateur EMDM nécessite généralement un échantillonneur de Gibbs qui peut être coûteux en temps de calcul.

- Le MAP est théoriquement sous-optimal (par rapport à l'EMDM), mais peut généralement atteindre de bonnes performances en pratique. De plus, l'algorithme proposé pour calculer cet estimateur n'implique que des mises à jour explicites, ce qui réduit considérablement le temps de calcul.

- L'EMDM du modèle simplifié est intéressant car il ne nécessite pas d'échantillonneur de Gibbs. La simplification du modèle n'est pas nécessairement réaliste et introduit un décalage au vrai modèle. Cependant, des simulations numériques (dans le chapitre II.4 du manuscrit complet) illustrent la robustesse et l'intérêt de l'approche en pratique.

C.2.4 L'estimateur maximum a posteriori (MAP)

Dans cette section, nous dérivons un algorithme pour calculer l'estimateur MAP dans le contexte de modèle donné (C.3). Le calcul du MAP requiert de résoudre le problème suivant :

$$\begin{aligned} \underset{\widehat{\mathbf{U}},\{\tau_k\},\{\lambda_r\}}{\text{maximize}} \quad & p(\widehat{\mathbf{U}}|\mathbf{Z},\{\tau_k\},\{\lambda_r\}) \\ \text{subject to} \quad & \tau_k \geq 0 \ \forall k,\ \lambda_r \geq 0 \ \forall r \\ & \widehat{\mathbf{U}}^H\widehat{\mathbf{U}} = \mathbf{I}_R \end{aligned} \tag{C.11}$$

À partir de (C.9) et (C.11), ce problème est exprimé comme suit

$$\begin{aligned} \underset{\{\widehat{\mathbf{u}}_r\},\{\tau_k\},\{\lambda_r\}}{\text{maximize}} \quad & \sum_{r=1}^{R} \mathrm{Re}\{\mathbf{c}_r^H\widehat{\mathbf{u}}_r\} + \widehat{\mathbf{u}}_r^H\left[\mathbf{A}_r + \mathbf{M}_r\right]\widehat{\mathbf{u}}_r - \sum_{k=1}^{K}\ln(\tau_k\lambda_r + \sigma^2) \\ \text{subject to} \quad & \tau_k \geq 0 \ \forall k,\ \lambda_r \geq 0 \ \forall r \\ & \widehat{\mathbf{U}}^H\widehat{\mathbf{U}} = \mathbf{I}_R \ \text{ with } \ \widehat{\mathbf{U}} = [\widehat{\mathbf{u}}_1|\ldots|\widehat{\mathbf{u}}_R] \\ & \mathbf{M}_r = \sum_{k=1}^{K}\frac{\tau_k\lambda_r}{\sigma^2(\tau_k\lambda_r+\sigma^2)}\mathbf{z}_k\mathbf{z}_k^H \end{aligned} \tag{C.12}$$

Pour résoudre ce problème, nous dérivons un algorithme itératif en appendice B.2 du manuscrit complet. Cet algorithme se base sur la méthode MM [75,76], qui effectue une mise à jour des blocs de paramètres en minimisant une fonction de substitution (majoration) de l'objectif. L'intérêt de cette méthode réside dans la possibilité d'obtenir des mises à jour sous forme explicite et d'assurer un décrément monotone de la valeur de l'objective à chaque étape.

C.2.5 L'estimateur minimisant la distance moyenne (EMDM)

C.2.5.1 Définition

L'EMDM vise à minimiser la distance euclidienne moyenne entre la vraie matrice de projection $\mathbf{U}\mathbf{U}^H$ et son estimateur en supposant une distribution préalable sur la base du sous-espace. En

étendant directement la formulation de [66] au cas complexe, l'EMDM est exprimé par :

$$\begin{aligned}\widehat{\mathbf{U}}_{\text{EMDM}} &= \arg\min_{\widehat{\mathbf{U}}} \ \mathbb{E}_{\mathbf{U},\mathbf{Z}}\left\{\| \ \widehat{\mathbf{U}}\widehat{\mathbf{U}}^H - \mathbf{U}\mathbf{U}^H \ \|_F^2\right\} \\ &= \arg\max_{\widehat{\mathbf{U}}} \ \mathbb{E}_{\mathbf{U},\mathbf{Z}}\{\text{Tr}\{\widehat{\mathbf{U}}^H\mathbf{U}\mathbf{U}^H\widehat{\mathbf{U}}\}\} \\ &= \arg\max_{\widehat{\mathbf{U}}} \ \int\left[\int \text{Tr}\{\widehat{\mathbf{U}}^H\mathbf{U}\mathbf{U}^H\widehat{\mathbf{U}}\}p(\mathbf{U}|\mathbf{Y})d\mathbf{U}\right]p(\mathbf{Y})d\mathbf{Y} \\ &= \arg\max_{\widehat{\mathbf{U}}} \ \int \text{Tr}\{\widehat{\mathbf{U}}^H\mathbf{U}\mathbf{U}^H\widehat{\mathbf{U}}\}p(\mathbf{U}|\mathbf{Y})d\mathbf{U} \\ &= \arg\max_{\widehat{\mathbf{U}}} \ \text{Tr}\left\{\widehat{\mathbf{U}}^H\left[\int \mathbf{U}\mathbf{U}^H p(\mathbf{U}|\mathbf{Y})d\mathbf{U}\right]\widehat{\mathbf{U}}\right\}\end{aligned} \tag{C.13}$$

qui peut être obtenu en tant que

$$\widehat{\mathbf{U}}_{\text{EMDM}} = \mathcal{P}_R\left\{\int \mathbf{U}\mathbf{U}^H p(\mathbf{U}|\mathbf{Y})d\mathbf{U}\right\} = \mathcal{P}_R\left\{\mathbf{M}(p(\mathbf{U}|\mathbf{Z}))\right\} \tag{C.14}$$

Dans lequel

$$\mathbf{M}(p(\mathbf{U}|\mathbf{Z})) = \int \mathbf{U}\mathbf{U}^H p(\mathbf{U}|\mathbf{Z})d\mathbf{U} \tag{C.15}$$

Remarque C.2.3 *L'expression l'EMDM dépend de* $p(\mathbf{U}|\mathbf{Z})$*, qui est spécifié en fonction du modèle de données et de la distribution a priori attribuée aux paramètres. Généralement, il n'existe pas de solution de forme explicite pour calculer* $\mathbf{M}(p(\mathbf{U}|\mathbf{Z}))$*. Cependant,* (C.14) *peut toujours être évalué à l'aide de la moyenne arithmétique induite (IAM) [66], comme*

$$\widehat{\mathbf{U}} \approx \mathcal{P}_R\left\{\frac{1}{N_r}\sum_{n=N_{bi}+1}^{N_{bi}+N_r}\mathbf{U}_{(n)}{\mathbf{U}_{(n)}}^H\right\} \tag{C.16}$$

où $\mathbf{U}_{(n)}$ *sont générées suivant* $p(\mathbf{U}|\mathbf{Z})$*. Une méthode de génération de cette loi est proposée en appendice A du manuscrit complet. Ici,* N_{bi} *représente le nombre d'échantillons nécessaires à initier la chaîne de Markov, et* N_r *est le nombre d'échantillons utilisés pour évaluer l'intégrale.*

C.2.5.2 EMDM dans le contexte des sources gaussienne composée

Nous rappelons que, selon le modèle de données de (C.3), $\mathbf{U} \sim \text{CGBL}(\mathbf{C}, \{\mathbf{A}_r\})$ et $\mathbf{z}_k \sim \mathcal{CN}(0, \tau_k\mathbf{U}\mathbf{\Lambda}\mathbf{U}^H + \sigma^2\mathbf{I}_N)$. En se basant sur (C.13) et (C.5), l'estimateur EMDM de $\mathbf{U}$ définit la solution de problème d'optimisation suivant

$$\begin{aligned}\underset{\widehat{\mathbf{U}},\{\tau_k\},\{\lambda_r\}}{\text{minimize}} \quad & \mathbb{E}_{\mathbf{U},\mathbf{Z}}\left\{\| \ \widehat{\mathbf{U}}\widehat{\mathbf{U}}^H - \mathbf{U}\mathbf{U}^H \ \|_F^2\right\} \\ \text{subject to} \quad & \tau_k \geq 0 \ \forall k, \ \lambda_r \geq 0 \ \forall r \\ & \widehat{\mathbf{U}}^H\widehat{\mathbf{U}} = \mathbf{I}_R\end{aligned} \tag{C.17}$$

Pour évaluer une solution de problème, nous dérivons dans le chapitre II.2.3.2 un algorithme itératif qui met à jour séquentiellement les variables $\widehat{\mathbf{U}}$, $\{\tau_k\}$ et $\{\lambda_r\}$. La mise à jour $\widehat{\mathbf{U}}$ nécessite un échantillonneur de Gibbs. Les mises a jour des des textures $\{\tau_k\}$ et les valeurs propres $\{\lambda_r\}$ mettent en œuvre une procédure MM identique à celle de l'estimateur map.

C.2.6 Modèle simplifié

Dans cette section, nous nous concentrons sur un cas particulier appelé modèle simplifié, dans lequel $\mathbf{\Sigma} = \lambda \mathbf{U}\mathbf{U}^H$. Cette relaxation du vrai modèle, par exemple utilisé dans [72], permet des simplifications intéressantes qui réduisent considérablement le temps de calcul de la procédure d'estimation. Notons que toute mise à l'échelle peut ici être absorbée dans les paramètres de texture comme $\tilde{\tau} = \lambda\tau$. On peut donc supposer $\mathbf{\Sigma} = \mathbf{U}\mathbf{U}^H$. Pour ce modèle, $\mathbf{z}_k|\mathbf{U}, \tau_k \sim \mathcal{CN}(0, \mathbf{\Sigma}_k)$, ainsi, la matrice de covariance est donnée par

$$\mathbf{\Sigma}_k = \tau_k \mathbf{U}\mathbf{U}^H + \sigma^2 \mathbf{I}_N \tag{C.18}$$

En utilisant le lemme Sherman-Morrison-Woodbury, $\mathbf{\Sigma}_k^{-1}$ est exprimée comme

$$\mathbf{\Sigma}_k^{-1} = (\tau_k \mathbf{U}\mathbf{U}^H + \sigma^2 \mathbf{I})^{-1} = \sigma^{-2}\mathbf{I} - \frac{\tau_k}{\sigma^2(\tau_k + \sigma^2)} \mathbf{U}\mathbf{U}^H, \ \forall k \tag{C.19}$$

Puis, la densité de probabilité $p(\mathbf{Z}|\mathbf{U}, \{\tau_k\})$ est réduite à

$$\begin{aligned} p(\mathbf{Z}|\mathbf{U}, \{\tau_k\}) &\propto \prod_{k=1}^{K} p(\mathbf{z}_k|\mathbf{U}, \tau_k) \propto \prod_{k=1}^{K} \frac{\exp\left\{-\mathbf{z}_k^H \mathbf{\Sigma}_k^{-1} \mathbf{z}_k\right\}}{\det(\mathbf{\Sigma}_k)} \\ &\propto \exp\left\{ \sum_{k=1}^{K} \frac{\tau_k}{\sigma^2(\tau_k + \sigma^2)} \mathbf{U}^H \mathbf{z}_k \mathbf{z}_k \mathbf{U} \right\} \left(\prod_{k=1}^{K} (\tau_k + \sigma^2)^{-R} \right) \\ &\propto \operatorname{etr}\left\{ \mathbf{U}^H \mathbf{W} \mathbf{U} \right\} \left(\prod_{k=1}^{K} (\tau_k + \sigma^2)^{-R} \right) \end{aligned} \tag{C.20}$$

où $\mathbf{W} = \mathbf{Z}\mathbf{B}\mathbf{Z}^H$ et $\mathbf{B} = \operatorname{diag}\left(\left\{ \frac{\tau_1}{\sigma^2(\tau_1+\sigma^2)} \cdots \frac{\tau_K}{\sigma^2(\tau_K+\sigma^2)} \right\} \right)$. Pour obtenir une expression sous forme explicite des solutions, nous affectons à $\mathbf{U}$ la distribution complexe Bingham invariant, i.e., $\mathbf{U} \sim \mathrm{CIB}(\kappa, \mathbf{A})$, ainsi sa densité de probabilité est $p_{\mathrm{CIB}}(\mathbf{U}) \propto \operatorname{etr}\{\kappa \mathbf{U}^H \mathbf{A} \mathbf{U}\}$.

C.2.6.1 EMDM et MAP dans le contexte de modèle simplifié

Dans ce cas, l'EMDM de la base $\widehat{\mathbf{U}}$ est défini comme le minimiseur du problème suivant

$$\begin{aligned} \underset{\widehat{\mathbf{U}}, \{\tau_k\}}{\text{minimize}} \quad & \mathbb{E}_{\mathbf{U},\mathbf{Z}} \left\{ \| \ \widehat{\mathbf{U}}\widehat{\mathbf{U}}^H - \mathbf{U}\mathbf{U}^H \ \|_F^2 \right\} \\ \text{subject to} \quad & \tau_k \geq 0, \ \forall k \\ & \widehat{\mathbf{U}}^H \widehat{\mathbf{U}} = \mathbf{I}_R \end{aligned} \tag{C.21}$$

Dans la section II.2.4.1, nous développons une méthode d'estimation itérative en résolvant (C.21) par rapport à $\widehat{\mathbf{U}}$ et $\{\tau_k\}$ séquentiellement. Il est ensuite montré en section II.2.4.2 que le MAP coïncide avec l'EMDM pour ce cas. Notons aussi que ce modèle spécifique fournit une solution avec une interprétation intéressante : l'EMDM apparaît naturellement comme le sous-espace principal d'une SCM utilisant des échantillons normalisés selon leur distance au sous-espace recherché.

C.2.7 Estimateurs bayésiens de sous-espace en présence de valeurs aberrantes

Les nouveaux estimateurs de sous espace proposés ont été dérivés dans le contexte des sources gaussiennes composées intégrées dans un bruit blanc gaussien. Ces estimateurs peuvent se révéler peu performants face à l'introduction de valeurs aberrantes dans les données secondaires. Ce, au même titre que des échantillons altérés par des valeurs aberrantes faussent le sous-espace dominant de la SCM [77]. Nous avons donc souhaité proposer des estimateurs robustes à la présence de telles données aberrantes en intégrant ces dernières dans le modèle de signal. Nous proposons ci-après des estimateurs MAP et EMDM dans le contexte de sources gaussienne composées noyées dans un bruit hétérogène (i.e. des données aberrantes qui suivent la loi gaussienne composée plus un bruit blanc gaussien).

C.2.7.1 Ajout de valeurs aberrantes au modèle gaussienne composé plus un bruit blanc additif gaussien

Les données seront modélisées comme une somme de sources gaussiennes composées $\mathbf{s}_k$, un bruit blanc gaussien $\mathbf{n}_k$ et des valeurs aberrantes gaussiennes composées $\mathbf{c}_k$. Pour ce modèle, les échantillons $\{\mathbf{z}_k\}$ est donnée par

$$\mathbf{z}_k = \mathbf{s}_k + \mathbf{c}_k + \mathbf{n}_k, \ \forall k \in [\![1, K]\!] \tag{C.22}$$

Pour chacun des paramètres, on supposera les distributions suivantes:

- Le signal $\mathbf{s}_k \sim \mathcal{CG}(0, \mathbf{\Sigma}_R, \tau_k)$ réside dans un sous espace de rang faible R qui est supposé pré-estimé. La DVS de la matrice de dispersion est $\mathbf{\Sigma}_R \overset{\text{DVS}}{=} [\mathbf{U}|\mathbf{U}^\perp] \begin{bmatrix} \mathbf{I}_R & 0 \\ 0 & 0 \end{bmatrix} [\mathbf{U}|\mathbf{U}^\perp]^H$. Notons que, comme cela a été fait dans [30, 72] et le modèle simplifié précédemment considéré, les valeurs propres non nulles de $\mathbf{\Sigma}_R$ sont supposées être égales à 1. L'hypothèse de l'égalité des valeurs propres est une relaxation réalisée à des fins de simplification de calcul, mais offre néanmoins des performances intéressantes dans la pratique [30].

- La base orthonormale de sous-espace de signal est distribuée comme suit: $\mathbf{U} \sim \text{CGBL}(\kappa, \bar{\mathbf{C}}, \{\mathbf{A}_r\})$.

- La valeur aberrante potentielle est distribuée comme $\mathbf{c}_k \sim \mathcal{CG}(0, \mathbf{\Sigma}_c, \beta_k)$, où sa matrice de covariance est définie par $\mathbf{\Sigma}_c = \mathbf{I}_N - \mathbf{U}\mathbf{U}^H = \mathbf{U}^\perp \mathbf{U}^{\perp H}$, où $\mathbf{U}^\perp$ est un complémentaire orthonormé de la base du sous espace source. Ce modèle de valeurs aberrantes est intéressant en termes de robustesse, car il permet de considérer des valeurs aberrante pour, défavorable en termes d'estimation de sous espace signal.

- Le bruit additif $\mathbf{n}_k$ est distribué comme $\mathbf{n}_k \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_N)$. Par souci de clarté, la variance σ^2 est supposée pré-estimée.

Ces hypothèses conduisent à $\mathbf{z}_k|\tau_k, \beta_k, \mathbf{U} \sim \mathcal{CN}(0, \mathbf{\Sigma}_k)$ avec

$$\mathbf{\Sigma}_k = \tau_k \mathbf{U}\mathbf{U}^H + \beta_k \mathbf{U}^\perp \mathbf{U}^{\perp H} + \sigma^2 \mathbf{I}_N \tag{C.23}$$

Ainsi, la densité de probabilité $\mathbf{Z}$ conditionnelle à $\{\tau_k\}, \{\beta_k\}, \mathbf{U}$ avec $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_K]$, est

$$p(\mathbf{Z}|\{\tau_k\}, \{\beta_k\}, \mathbf{U}) = \prod_{k=1}^{K} p(\mathbf{z}_k|\tau_k, \beta_k, \mathbf{U}) \propto \prod_{k=1}^{K} \frac{\exp\{-\mathbf{z}_k^H \mathbf{\Sigma}_k^{-1} \mathbf{z}_k\}}{\det(\mathbf{\Sigma}_k)} \tag{C.24}$$

Grâce au lemme the Sherman-Morisson-Woodbury, l'expression de $\mathbf{\Sigma}_k^{-1}$ est donnée par

$$\mathbf{\Sigma}_k^{-1} = \frac{1}{\sigma^2+\beta_k}\mathbf{I} - \frac{\tau_k-\beta_k}{(\sigma^2+\tau_k)(\sigma^2+\beta_k)}\mathbf{U}\mathbf{U}^H \tag{C.25}$$

De (C.24), la probabilité a posteriori $\mathbf{U}|\mathbf{Z},\{\tau_k\},\{\beta_k\}$ est exprimée comme

$$\begin{aligned}
p\left(\mathbf{U}|\mathbf{Z},\{\tau_k\},\{\beta_k\}\right) &\propto p\left(\mathbf{Z}|\mathbf{U},\{\tau_k\},\{\beta_k\}\right) p_{\mathrm{CGBL}}(\mathbf{U}) \\
&\propto \prod_{k=1}^{K} \frac{\exp\left\{-\mathbf{z}_k^H\mathbf{\Sigma}_k^{-1}\mathbf{z}_k\right\}}{\det(\mathbf{\Sigma}_k)} p_{\mathrm{CGBL}}(\mathbf{U}) \\
&\propto \prod_{k=1}^{K} h_k \exp\left\{\frac{\tau_k}{\sigma^2(\sigma^2+\tau_k)}\mathbf{z}_k^H\mathbf{U}\mathbf{U}^H\mathbf{z}_k\right\} \prod_{k=1}^{K}\exp\left\{-\frac{\beta_k}{\sigma^2(\sigma^2+\beta_k)}\mathbf{y}_k^H\mathbf{U}\mathbf{U}^H\mathbf{y}_k\right\} p_{\mathrm{CGBL}}(\mathbf{U}) \\
&\propto \left(\prod_{k=1}^{K} h_k\right) \mathrm{etr}\{\mathbf{U}^H\mathbf{M}\mathbf{U}\} \exp\left\{\sum_{r=1}^{R}\mathrm{Re}\{\mathbf{c}_r^H\mathbf{u}_r\} + \mathbf{u}_r^H\mathbf{A}_r\mathbf{u}_r\right\} \\
&\propto \left(\prod_{k=1}^{K} h_k\right) \exp\left\{\sum_{r=1}^{R}\mathrm{Re}\{\mathbf{c}_r^H\mathbf{u}_r\} + \mathbf{u}_r^H(\mathbf{A}_r+\mathbf{M})\mathbf{u}_r\right\}
\end{aligned} \tag{C.26}$$

avec

$$\begin{cases} h_k = \dfrac{1}{(\tau_k+\sigma^2)^R(\beta_k+\sigma^2)^{N-R}} \\ \mathbf{M} = \mathbf{Z}\mathrm{diag}\left(\dfrac{\tau_1-\beta_1}{(\sigma^2+\tau_1)(\sigma^2+\beta_1)},\ldots,\dfrac{\tau_K-\beta_K}{(\sigma^2+\tau_K)(\sigma^2+\beta_K)}\right)\mathbf{Z}^H \end{cases}$$

C.2.7.2 Estimateurs

Pour ce modèle d'observations, nous proposons de nouveaux estimateurs

- Un algorithme permettant de calculer le MAP est proposé en section II.3.2..
- Un algorithme permettant de calculer l'EMDM est proposé en section II.3.3..

De plus, une précision est détaillé pour un prior Bingham-invariant, car MAP et l'EMDM coïncident pour ce cas.

C.2.8 Applications

les performances et la robustesse des méthodes proposées sont illustrées via des simulations numériques dans le chapitre II.4. du manuscrit complet. De plus, une application à la détection radar est illustrée sur des données réelles.

C.3 Détection de changement de sous-espace signal de matrices de covariance structurées

C.3.1 Introduction et motivations

Le test statistique de l'égalité (ou de la proportionnalité) de la matrice de covariance suscite un intérêt croissant dans le contexte du traitement du signal et des images [5]. En effet, ce test d'hypothèse bien établi a été étudié et appliqué à la détection de changements d'anomalies et à

la classification pour une pléthore d'applications comme traitement d'images *Synthetic Aperture Radar* (SAR). Des tests d'égalité ont notamment été proposés pour la détection des changements dans les séries temporelles d'images SAR [86–88]. Une vue d'ensemble et une analyse statistique de ce sujet sont proposées dans [5], et le cas du tests de proportionnalité a été considérée dans [91, 92].

dans cette partie de la thèse, nous proposons de nouveaux tests statistiques pour le contexte de matrices de covariance structurées. En effet, la matrice de covariance des mesures radar présente généralement des structures inhérentes. Nous considérons le traitement des matrices de covariance modélisées par la somme d'une composante de rang faible (où se trouve l'information générée par des signaux d'intérêt) plus une matrice d'identité (représentant la contribution d'un bruit blanc additif). Cette structure particulière des matrices de covariance a été exploitée dans le contexte de détection radar [43, 45].

En particulier, nous proposons ci-après trois nouveaux tests du rapport de vraisemblance généralisée (TRVG) qui tiennent compte de cette structure rang faible. Formellement, les deux premiers détecteurs ré-expriment des tests d'égalité et de proportionnalité, mais uniquement sur la composante principale du signal d'intérêt. Le troisième test propose une détection sensible uniquement à une variation du sous-espace du signal. La dérivation des deuxième et troisième tests nécessite de résoudre certains problèmes d'optimisation non triviaux, pour lesquels nous dérivons des algorithmes MM appropriés.

C.4 Modèle des observations

Nous considérons le jeu de données total partitionné en $I+1$ sous ensembles indépendants. Pour chaque sous-ensemble, $i \in [\![0, I]\!]$, nous disposons de plusieurs observations i.i.d., notées $\{\mathbf{z}_k^i\}_{k\in[\![1,K]\!]}$ et chacune de taille N. De manière très générale, ces données sont modélisées par:

$$\mathbf{z}_k^i = \mathbf{s}_k^i + \mathbf{n}_k^i \tag{C.27}$$

• Le signal d'intérêt $\mathbf{s}_k^i \sim \mathcal{CN}(0, \mathbf{\Sigma}_R^i)$, sa matrice de covariance est de rang $R \ll N$ et notée $\mathbf{\Sigma}_R^i = \mathbf{U}_i\mathbf{R}_i\mathbf{U}_i^H$ où $\mathbf{U}_i \in \mathcal{U}_N^R$ représente une base orthonormale engendrée par le sous-espace signal et $\mathbf{R}_i \in \mathbb{C}^{R\times R}$ est la matrice de covariance correspondante dans cet espace.
• $\mathbf{n}_k^i \sim \mathcal{CN}(0, \sigma^2\mathbf{I})$ désigne un bruit blanc additif gaussien.
• $\mathbf{s}_k^i$ et $\mathbf{n}_k^i$ sont deux vecteurs aléatoires indépendants.
En conséquence, la distribution des observations est :

$$\mathbf{z}_k^i \sim \mathcal{CN}(0, \mathbf{\Sigma}_i) \text{ avec } \mathbf{\Sigma}_i = \mathbf{U}_i\mathbf{R}^i\mathbf{U}_i^H + \sigma^2\mathbf{I} \tag{C.28}$$

On notera $\mathcal{L}(\{\mathbf{z}_k^i\}|\boldsymbol{\theta}) = \prod_{i=0}^{I} \frac{\exp\mathrm{tr}\{-\mathbf{S}_i\mathbf{\Sigma}_i^{-1}(\boldsymbol{\theta})\}}{|\mathbf{\Sigma}_i(\boldsymbol{\theta})|^K}$, la vraisemblance des données $\{\mathbf{z}_k^i\}$, où $\mathbf{S}_i$ est la matrice de covariance empirique et $\boldsymbol{\theta}$ le paramètre d'intérêt définissant $\{\mathbf{\Sigma}_i\}$.

C.4.1 État de l'art

Nous présentons deux TRVG standard utilisés pour tester la similarité des matrices de covariance. Le premier test concerne l'égalité des matrices de covariance, tandis que le second est dérivé pour le test de proportionnalité entre les matrices de covariance dans le cas gaussien. Nous considérons $I+1$ ensembles d'échantillons indépendants $\mathbf{z}_k^i \in \mathbb{C}^N$, $i \in [\![0, I]\!]$, $K \in [\![1, K]\!]$, avec K_i échantillons pour chaque ensemble i. Les échantillons $\mathbf{z}_k^i \sim \mathcal{CN}(0, \mathbf{\Sigma}_i)$ sont supposés être i.i.d. et leur SCM correspondant sont notés $\mathbf{S}_i$.

$$\mathbf{S}_i = \frac{1}{K_i}\sum_{k=1}^{K_i} \mathbf{z}_k^i\mathbf{z}_k^{iH} \tag{C.29}$$

C.4.1.1 test d'égalité

Le problème général du test de l'égalité de la matrice de covariance [97] dans le cas gaussien à valeur complexe est analysé, par exemple dans [5]. Ce test d'hypothèse se formule comme

$$\begin{cases} \mathcal{H}_0 : \ \mathbf{\Sigma}_0 = \mathbf{\Sigma}, \ \mathbf{\Sigma}_i = \mathbf{\Sigma} \ \forall i \in [\![1, I]\!] \\ \mathcal{H}_1 : \ \mathbf{\Sigma}_0 \neq \mathbf{\Sigma}, \ \mathbf{\Sigma}_i = \mathbf{\Sigma} \ \forall i \in [\![1, I]\!] \end{cases} \tag{C.30}$$

Le TRVG pour ce test d'hypothèse, noté t_{E}, est donné par

$$\frac{|\hat{\mathbf{\Sigma}}_{\mathcal{H}_0}|}{\left(|\hat{\mathbf{\Sigma}}^0_{\mathcal{H}_1}|^{\rho_0}|\hat{\mathbf{\Sigma}}^\star_{\mathcal{H}_1}|^{\rho_\star}\right)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \delta^{\mathrm{E}}_{\mathrm{glr}} \tag{C.31}$$

avec les quantités

$$\begin{aligned} K &= \sum_{i=0}^{I} K_i \\ K_\star &= K - K_0 \\ \rho_0 &= K_0/K \\ \rho_\star &= K_\star/K \end{aligned}$$

et les SCMs

$$\begin{aligned} \hat{\mathbf{\Sigma}}_{\mathcal{H}_0} &= \sum_{i=0}^{I} \mathbf{S}_i/K \\ \hat{\mathbf{\Sigma}}^0_{\mathcal{H}_1} &= \mathbf{S}_0/K_0 \\ \hat{\mathbf{\Sigma}}^\star_{\mathcal{H}_1} &= \sum_{i=1}^{I} \mathbf{S}_i/K_\star \end{aligned} \tag{C.32}$$

C.4.1.2 Test de proportionnalité

Le problème général du test de la proportionnalité de la matrice de covariance [98] dans le cas gaussien est analysé dans par exemple [91]. Le test d'hypothèse est formulé comme

$$\begin{cases} \mathcal{H}_0 : \ \mathbf{\Sigma}_0 = \beta_0\mathbf{\Sigma}, \ \mathbf{\Sigma}_i = \beta_i\mathbf{\Sigma} \ \forall i \in [\![1, I]\!] \\ \mathcal{H}_1 : \ \mathbf{\Sigma}_0 \neq \beta_0\mathbf{\Sigma}, \ \mathbf{\Sigma}_i = \beta_i\mathbf{\Sigma} \ \forall i \in [\![1, I]\!] \end{cases} \tag{C.33}$$

Le TRVG sur la proportionnalité, noté t_{P}, est donné par

$$\left(\frac{|\hat{\beta}^0_{\mathcal{H}_0}\hat{\mathbf{\Sigma}}^{\mathrm{gfp}}_{\mathcal{H}_0}|}{|\hat{\mathbf{\Sigma}}^0_{\mathcal{H}_1}|}\right)^{K_0} \prod_{i=1}^{I} \left(\frac{|\hat{\beta}^i_{\mathcal{H}_0}\hat{\mathbf{\Sigma}}^{\mathrm{gfp}}_{\mathcal{H}_0}|}{|\hat{\beta}^i_{\mathcal{H}_1}\hat{\mathbf{\Sigma}}^{\mathrm{gfp}}_{\mathcal{H}_1}|}\right)^{K_i} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \delta^{\mathrm{P}}_{\mathrm{glr}}, \tag{C.34}$$

avec

i) $\{\hat{\beta}^i_{\mathcal{H}_0}\}$ et $\hat{\mathbf{\Sigma}}^{\mathrm{gfp}}_{\mathcal{H}_0}$ sont les coefficients de proportionnalité estimés avec le *generalized fixed point estimator* (GFPE) [92] appliqué sur l'ensemble $\{\mathbf{S}_i\}_{i\in[\![0,I]\!]}$

$$\begin{aligned} \hat{\mathbf{\Sigma}}^{\mathrm{gfp}}_{\mathcal{H}_0} &= \frac{N}{K}\sum_{i=0}^{I} \frac{K_i}{\mathrm{Tr}\{\mathbf{S}_i\hat{\mathbf{\Sigma}}^{\mathrm{gfp}^{-1}}_{\mathcal{H}_0}\}}\mathbf{S}_i \\ \hat{\beta}^i_{\mathcal{H}_0} &= \frac{1}{N}\mathrm{Tr}\{\mathbf{S}_i\hat{\mathbf{\Sigma}}^{\mathrm{gfp}^{-1}}_{\mathcal{H}_0}\} \end{aligned} \tag{C.35}$$

ii) $\{\hat{\beta}^i_{\mathcal{H}_1}\}$ et $\hat{\mathbf{\Sigma}}^{\text{gfp}}_{\mathcal{H}_1}$ sont obtenus à partir de l'estimateur du point-fixe généralisé (GFPE) sur l'ensemble $\{\mathbf{S}_i\}_{i\in[\![1,I]\!]}$

$$\begin{aligned}\hat{\mathbf{\Sigma}}^{\text{gfp}}_{\mathcal{H}_1} &= \frac{N}{K}\sum_{i=1}^{I}\frac{K_i}{\text{Tr}\{\mathbf{S}_i\hat{\mathbf{\Sigma}}^{\text{gfp}^{-1}}_{\mathcal{H}_1}\}}\mathbf{S}_i \\ \hat{\beta}^i_{\mathcal{H}_1} &= \frac{1}{N}\text{Tr}\{\mathbf{S}_i\hat{\mathbf{\Sigma}}^{\text{gfp}^{-1}}_{\mathcal{H}_1}\}\end{aligned} \tag{C.36}$$

iii) $\hat{\mathbf{\Sigma}}^0_{\mathcal{H}_1}$ est la SCM donné comme

$$\hat{\mathbf{\Sigma}}^0_{\mathcal{H}_1} = \mathbf{S}_0/K_0 \tag{C.37}$$

C.4.2 Détecteurs proposés pour les modèles structurés

C.4.2.1 Test d'égalité rang faible

Nous développons un TRVG sensible à une variation de tout paramètre de la matrice de covariance du signal à structure rang faible dans l'ensemble $i = 0$. Ainsi, pour le modèle de matrice de covariance structurée, considérons le test d'hypothèse suivant

$$\begin{cases}\mathcal{H}_0: & \mathbf{\Sigma}^i_R = \mathbf{\Sigma}_R \ \forall i \in [\![0,I]\!] \\ \mathcal{H}_1: & \begin{vmatrix}\mathbf{\Sigma}^i_R = \mathbf{\Sigma}_R \ \forall i \in [\![1,I]\!] \\ \mathbf{\Sigma}^0_R \neq \mathbf{\Sigma}_R,\end{vmatrix}\end{cases} \tag{C.38}$$

qui est donc similaire à un test d'égalité standard, sauf que les matrices de covariance appartiennent à $\mathcal{H}^{\text{LR}}_R$. Par conséquent, le test d'hypothèse peut être reformulé comme suit:

$$\begin{cases}\mathcal{H}_0: & \left|\mathbf{\Sigma}_i = \mathbf{\Sigma}_{R,\mathcal{H}_0} + \sigma^2\mathbf{I} \ \triangleq\ \mathbf{\Sigma}_{\mathcal{H}_0} \in \mathcal{H}^{\text{LR}}_R, \ \forall i \in [\![0,I]\!]\right. \\ \mathcal{H}_1: & \begin{vmatrix}\mathbf{\Sigma}_0 = \mathbf{\Sigma}^0_{R,\mathcal{H}_1} + \sigma^2\mathbf{I} \ \triangleq\ \mathbf{\Sigma}^0_{\mathcal{H}_1} \in \mathcal{H}^{\text{LR}}_R \\ \mathbf{\Sigma}_i = \mathbf{\Sigma}^\star_{R,\mathcal{H}_1} + \sigma^2\mathbf{I} \ \triangleq\ \mathbf{\Sigma}^\star_{\mathcal{H}_1} \in \mathcal{H}^{\text{LR}}_R, \ \forall i \in [\![1,I]\!]\end{vmatrix}\end{cases}$$

L'expression de TRVG est donnée par

$$\frac{\max_{\boldsymbol{\theta}^{\text{lrE}}_{\mathcal{H}_1}} \mathcal{L}\left(\{\mathbf{z}^i_k\}|\mathcal{H}_1,\boldsymbol{\theta}^{\text{lrE}}_{\mathcal{H}_1}\right)}{\max_{\boldsymbol{\theta}^{\text{lrE}}_{\mathcal{H}_0}} \mathcal{L}(\{\mathbf{z}^i_k\}|\mathcal{H}_0,\boldsymbol{\theta}^{\text{lrE}}_{\mathcal{H}_0})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \delta^{\text{lrE}}_{\text{glr}}, \tag{C.39}$$

avec les ensembles

$$\begin{aligned}\boldsymbol{\theta}^{\text{lrE}}_{\mathcal{H}_1} &= \left\{\mathbf{\Sigma}^0_{\mathcal{H}_1}, \mathbf{\Sigma}^\star_{\mathcal{H}_1}\right\} \\ \boldsymbol{\theta}^{\text{lrE}}_{\mathcal{H}_0} &= \{\mathbf{\Sigma}_{\mathcal{H}_0}\}\end{aligned}$$

Dans le contexte des données gaussiennes, l'estimateur de maximum de vraisemblance (EMV) de la matrice à structure rang faible est obtenue en seuillant les valeurs propres de la SCM [99]. Par conséquent, on peut montrer que

$$\begin{aligned}\hat{\boldsymbol{\theta}}^{\text{lrE}}_{\mathcal{H}_0} &= \left\{\mathcal{T}_R\{\hat{\mathbf{\Sigma}}_{\mathcal{H}_0}\}\right\} \\ \hat{\boldsymbol{\theta}}^{\text{lrE}}_{\mathcal{H}_1} &= \left\{\mathcal{T}_R\{\hat{\mathbf{\Sigma}}^0_{\mathcal{H}_1}\}, \mathcal{T}_R\{\hat{\mathbf{\Sigma}}^\star_{\mathcal{H}_1}\}\right\}\end{aligned}$$

Enfin, la TRVG pour tester l'égalité des matrices à structure rang faible, notée t^{LR}_{E}, est

$$\frac{\mathcal{L}(\{\mathbf{z}^i_k\}|\mathcal{H}_1,\hat{\boldsymbol{\theta}}^{\text{lrE}}_{\mathcal{H}_1})}{\mathcal{L}(\{\mathbf{z}^i_k\}|\mathcal{H}_0,\hat{\boldsymbol{\theta}}^{\text{lrE}}_{\mathcal{H}_0})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \delta^{\text{lrE}}_{\text{glr}} \tag{C.40}$$

C.4.2.2 Test de proportionnalité rang faible

Nous proposons un TRVG qui pour tester la proportionnalité de la composante signal dans une matrice de covariance à structure rang faible. Notons que ce test diffère d'un test de proportionnalité strict. En effet, les fluctuations d'échelle s'appliquent ici que sur la partie signal rang faible de la matrice de covariance (et non sur la composante identité, liée au bruit thermique). Pour le modèle matriciel de covariance dans (C.28), ceci conduit au test d'hypothèse

$$\begin{cases} \mathcal{H}_0 : & \left| \boldsymbol{\Sigma}_R^i = \tau_i \boldsymbol{\Sigma}_R \ \forall i \in [\![0, I]\!] \right. \\ \mathcal{H}_1 : & \left| \begin{array}{l} \boldsymbol{\Sigma}_R^i = \tau_i \boldsymbol{\Sigma}_R \ \forall i \in [\![1, I]\!] \\ \boldsymbol{\Sigma}_R^0 \neq \tau_0 \boldsymbol{\Sigma}_R \end{array} \right. \end{cases} \tag{C.41}$$

Le test ci-dessus peut être réécrit comme

$$\begin{cases} \mathcal{H}_0 : \left| \boldsymbol{\Sigma}_i = \tau_i \mathbf{U}_{\mathcal{H}_0} \boldsymbol{\Lambda}_{\mathcal{H}_0} (\mathbf{U}_{\mathcal{H}_0})^H + \sigma^2 \mathbf{I}, \ \forall i \in [\![0, I]\!] \right. \\ \mathcal{H}_1 : \left| \begin{array}{l} \boldsymbol{\Sigma}_0 = \ \mathbf{U}_{\mathcal{H}_1}^0 \mathbf{R}_{\mathcal{H}_1}^0 (\mathbf{U}_{\mathcal{H}_1}^0)^H + \sigma^2 \mathbf{I} \\ \boldsymbol{\Sigma}_i = \tau_i \mathbf{U}_{\mathcal{H}_1}^\star \boldsymbol{\Lambda}_{\mathcal{H}_1}^\star (\mathbf{U}_{\mathcal{H}_1}^\star)^H + \sigma^2 \mathbf{I}, \ \forall i \in [\![1, I]\!] \end{array} \right. \end{cases} \tag{C.42}$$

avec l'ensemble de paramètres

$$\begin{array}{l} \boldsymbol{\theta}_{\mathcal{H}_0}^{\text{lrP}} = \left\{ \{\tau_i\}_{i \in [\![0,I]\!]}, \mathbf{U}_{\mathcal{H}_0}, \boldsymbol{\Lambda}_{\mathcal{H}_0} \right\}, \\ \boldsymbol{\theta}_{\mathcal{H}_1}^{\text{lrP}} = \left\{ \{\tau_i\}_{i \in [\![1,I]\!]}, \mathbf{U}_{\mathcal{H}_1}^\star, \boldsymbol{\Lambda}_{\mathcal{H}_1}^\star, \mathbf{U}_{\mathcal{H}_1}^0, \mathbf{R}_{\mathcal{H}_1}^0 \right\}. \end{array} \tag{C.43}$$

Afin d'évaluer ce TRVG, nous utilisons a nouveau l'algorithme MM par blocs [101]. Les algorithmes proposés pour calculer les EMVs des paramètres $\boldsymbol{\theta}_{\mathcal{H}_0}^{\text{lrP}}$ et $\boldsymbol{\theta}_{\mathcal{H}_1}^{\text{lrP}}$ sous chaque hypothèse sont détaillés dans le chapitre III du manuscrit complet.

C.4.2.3 Test d'égalité de sous espace

Nous nous concentrons maintenant sur la propriété d'égalité de sous-espace principal. Nous visons à construire un test uniquement sensible au changement du sous-espace signal des matrices de covariance structurées comme en (C.28). Formellement, nous allons tester l'hypothèse suivante:

$$\begin{cases} \mathcal{H}_0 : \left| \mathcal{R}_R\{\boldsymbol{\Sigma}_i\} = \mathcal{R}_R\{\boldsymbol{\Sigma}\}, \ \forall i \in [\![0, I]\!] \right. \\ \mathcal{H}_1 : \left| \begin{array}{l} \mathcal{R}_R\{\boldsymbol{\Sigma}_i\} = \mathcal{R}_R\{\boldsymbol{\Sigma}\}, \ \forall i \in [\![1, I]\!] \\ \mathcal{R}_R\{\boldsymbol{\Sigma}_0\} \neq \mathcal{R}_R\{\boldsymbol{\Sigma}\} \end{array} \right. \end{cases} \tag{C.44}$$

avec $\mathcal{R}_R\{\}$ l'opérateur défini par:

$$\begin{array}{rlcl} \mathcal{R}_{\text{R}} : & \mathcal{H}_M^+ & \longrightarrow & \mathcal{G}_R \\ & \boldsymbol{\Sigma} \overset{\text{DVS}}{=} [\mathbf{U}_R | \mathbf{U}_R^\perp] \mathbf{D} [\mathbf{U}_R | \mathbf{U}_R^\perp]^H & \longmapsto & \mathbf{U}_R \mathbf{U}_R^H \end{array} \tag{C.45}$$

Remarque: Ce test permet d'étudier un mécanisme physique spécifique. Par exemple, les puissances et les corrélations des sources peuvent fluctuer tout en engendrant le même sous-espace signal. Cela conduit à la fois aux relations:

$$\boldsymbol{\Sigma}_0 \neq \boldsymbol{\Sigma}_i \text{ et } \boldsymbol{\Sigma}_0 \not\propto \boldsymbol{\Sigma}_i, \ \forall i \tag{C.46}$$

et

$$\mathcal{R}_R\{\boldsymbol{\Sigma}_0\} = \mathcal{R}_R\{\boldsymbol{\Sigma}_i\}, \ \forall i \tag{C.47}$$

Notons que (C.46) est considérée comme $\mathcal{H}_1$ pour les tests standards (C.30) et (C.33) tandis que la relation (C.47) donne $\mathcal{H}_0$ pour le test (C.44). Ainsi, le détecteur proposé sera insensible aux variations de corrélations inter-sources et aux fluctuations de leurs puissances. Cette propriété peut s'avérer utile pour diminuer le taux de fausses alarmes dans des applications spécifiques [91].

C.4.3 Dérivation du test

En reprenant (C.44) avec la structure en (C.28), le test d'hypothèse peut être reformulé :

$$\begin{cases} \mathcal{H}_0 : \left| \boldsymbol{\Sigma}_i = \mathbf{U}_{\mathcal{H}_0} \mathbf{R}^i_{\mathcal{H}_0} (\mathbf{U}_{\mathcal{H}_0})^H + \sigma^2 \mathbf{I}, \ \forall i \in [\![0, I]\!] \right. \\ \mathcal{H}_1 : \begin{cases} \boldsymbol{\Sigma}_i = \mathbf{U}^\star_{\mathcal{H}_1} \mathbf{R}^i_{\mathcal{H}_1} (\mathbf{U}^\star_{\mathcal{H}_1})^H + \sigma^2 \mathbf{I}, \ \forall i \in [\![1, I]\!] \\ \boldsymbol{\Sigma}_0 = \mathbf{U}^0_{\mathcal{H}_1} \mathbf{R}^0_{\mathcal{H}_1} (\mathbf{U}^0_{\mathcal{H}_1})^H + \sigma^2 \mathbf{I} \end{cases} \end{cases} \tag{C.48}$$

où $\mathbf{U}_{\mathcal{H}_0}(\mathbf{U}_{\mathcal{H}_0})^H$, $\mathbf{U}^\star_{\mathcal{H}_1}(\mathbf{U}^\star_{\mathcal{H}_1})^H$ et $\mathbf{U}^0_{\mathcal{H}_1}(\mathbf{U}^0_{\mathcal{H}_1})^H$ représentent les sous-espaces engendrés respectivement par les données secondaires sous $\mathcal{H}_0$, sous $\mathcal{H}_1$ et par l'ensemble sous-test sous $\mathcal{H}_1$. Les quantités $\mathbf{R}^i_{\mathcal{H}_0}$, $\mathbf{R}^i_{\mathcal{H}_1}$ et $\mathbf{R}^0_{\mathcal{H}_1}$ désignent les matrices de covariance dans les sous-espaces engendrés respectivement par les données secondaires sous $\mathcal{H}_0$, sous $\mathcal{H}_1$ et par l'ensemble sous-test sous $\mathcal{H}_1$. Le TRVG pour le test considéré est donné par:

$$\frac{\max_{\boldsymbol{\theta}^{\text{sub}}_{\mathcal{H}_1}} \mathcal{L}\left(\{\mathbf{z}^i_k\}|\mathcal{H}_1, \boldsymbol{\theta}^{\text{sub}}_{\mathcal{H}_1}\right)}{\max_{\boldsymbol{\theta}^{\text{sub}}_{\mathcal{H}_0}} \mathcal{L}(\{\mathbf{z}^i_k\}|\mathcal{H}_0, \boldsymbol{\theta}^{\text{sub}}_{\mathcal{H}_0})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \delta^{\text{sub}}_{\text{glr}} \tag{C.49}$$

où on notera $\mathcal{L}\left(\{\mathbf{z}^i_k\}|\mathcal{H}_1, \boldsymbol{\theta}^{\text{sub}}_{\mathcal{H}_1}\right)$ et $\mathcal{L}\left(\{\mathbf{z}^i_k\}|\mathcal{H}_0, \boldsymbol{\theta}^{\text{sub}}_{\mathcal{H}_0}\right)$ les vraisemblances des données $\{\mathbf{z}^i_k\}$ sous respectivement $\mathcal{H}_1$ et $\mathcal{H}_0$. Les paramètres $\boldsymbol{\theta}^{\text{sub}}_{\mathcal{H}_0}$ et $\boldsymbol{\theta}^{\text{sub}}_{\mathcal{H}_1}$ désignent $\boldsymbol{\theta}^{\text{sub}}_{\mathcal{H}_0} = \left\{\{\mathbf{R}^i_{\mathcal{H}_0}\}_{i\in[\![0,I]\!]}, \mathbf{U}_{\mathcal{H}_0}\right\}$ et $\boldsymbol{\theta}^{\text{sub}}_{\mathcal{H}_1} = \left\{\{\mathbf{R}^i_{\mathcal{H}_1}\}_{i\in[\![0,I]\!]}, \mathbf{U}^\star_{\mathcal{H}_1}, \mathbf{U}^0_{\mathcal{H}_1}\right\}$. Afin d'évaluer ce TRVG, nous utilisons une fois de plus un algorithme MM par blocs [101], dont la dérivation est détaillée dans le chapitre III du manuscrit complet.

C.4.4 Applications

les performances et les propriétés des détecteurs proposées sont illustrées via des simulations numériques dans le chapitre III.3. du manuscrit complet. De plus, une application pour la détection de changements dans une série temporelle d'images SAR est conduite sur des données réelles.

C.5 Conclusion

Dans un premier lieu, nous avons proposé des estimateurs MAP et EMDM du sous-espace de signal dans le contexte de sources suivant la loi gaussienne composée et sous-espace suivant la loi CBLG. Ensuite, afin de développer des estimateurs bayésiens de sous-espaces qui soient robustes à la présence de bruit impulsif, nous avons dérivé de nouveaux estimateurs pour le contexte de sources suivant la loi gaussienne composée noyées dans bruit hétérogènes (valeurs aberrantes suivant la loi gaussienne composée plus un bruit blanc gaussien). L'intérêt pratique de ces estimateurs a été illustré en simulation et pour une problématique de détection radar sur des données réelles.

Dans un second lieu, nous avons proposé des détecteurs de changements pour des groupes de matrices de covariance structurées. Ces nouveaux détecteurs ne sont sensibles qu'à des variations dans le sous-espace principal de la matrice de covariance et généralisent les tests de proportionnalité/égalité précédemment établis. L'intérêt pratique de ces détecteurs a été illustré en simulation et pour une problématique de détection de changements dans une série temporelle d'images SAR sur des données réelles.

Bibliography

[1] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transaction Antennas Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[2] I. Jolliffe, *Principal component analysis*. Springer, 2011.

[3] R. C. D. Lamare and R. Sampaio-Neto, "Reduced-rank adaptive filtering based on joint iterative optimization of adaptive filters," *IEEE Signal Processing Letters*, vol. 14, no. 12, pp. 980–983, 2007.

[4] G. Turin, "An introduction to matched filters," *IRE Transactions on Information Theory*, vol. 6, no. 3, pp. 311–329, 1960.

[5] D. Ciuonzo, V. Carotenuto, and A. De Maio, "On multiple covariance equality testing with application to SAR change detection," *IEEE Transactions on Signal Processing*, vol. 65, no. 19, pp. 5078–5091, 2017.

[6] C. Khatri, "Classical statistical analysis based on a certain multivariate complex gaussian distribution," *The Annals of Mathematical Statistics*, pp. 98–114, 1965.

[7] M. Rouaud, "Probability, statistics and estimation," *Propagation of Uncertainties*, pp. 7–18, 2013.

[8] Y. Wu, C. Wang, H. Zhang, X. Wen, and B. Zhang, "Statistical analysis and simulation of high-resolution SAR ground clutter data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp*, vol. 4, 2008, pp. 2182–2184.

[9] E. Conte, A. De Maio, and C. Galdi, "Statistical analysis of real clutter at different range resolutions," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 40, no. 3, pp. 903–918, 2004.

[10] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing: A survey," *Proceedings of the IEEE*, vol. 73, no. 3, pp. 433–481, 1985.

[11] K. Yao, "A representation theorem and its applications to spherically invariant random processes," *IEEE Transactions on Information Theory*, vol. 19, no. 5, p. 600–608, Sept 1973.

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

[13] F. Shi and I. W. Selesnick, "An elliptically contoured exponential mixture model for wavelet based image denoising," *Applied and Computational Harmonic Analysis*, vol. 23, no. 1, pp. 131–151, 2007.

[14] S. Zozor and C. Vignat, "Some results on the denoising problem in the elliptically distributed context," *IEEE Transactions on Signal Processing*, vol. 58, no. 1, pp. 134–150, 2010.

[15] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Transactions Image Processing*, vol. 12, no. 11, 2003.

[16] E. Ollila, D. E. Tyler, V. Koivunen, and H. P. Vincent, "Compound Gaussian clutter modeling with an inverse Gaussian texture distribution," *IEEE Signal Processing Letters*, vol. 19, no. 12, pp. 876–879, 2012.

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

[18] F. Pascal, Y. Chitour, J.-P. Ovarlez, P. Forster, and P. Larzabal, "Covariance structure maximum-likelihood estimates in compound Gaussian noise: Existence and algorithm analysis," *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 34–48, 2008.

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

[20] S. Zozor and C. Vignat, "Robust m-estimators of multivariate location and scatter," *The Annals of Statistics*, vol. 4, no. 1, pp. 51–67, 1976.

[21] V. Ollier, M. N. E. Korso, R. Boyer, P. Larzabal, and M. Pesavento, "Robust calibration of radio interferometers in non-Gaussian environment," *IEEE Transactions on Signal Processing*, vol. 65, no. 21, pp. 5649–5660, Nov 2017.

[22] X. Zhang, M. N. El Korso, and M. Pesavento, "MIMO radar target localization and performance evaluation under SIRP clutter," *Signal Processing*, vol. 130, pp. 217–232, 2017.

[23] B. Meriaux, C. Ren, M. N. El Korso, A. Breloy, and P. Forster, "Asymptotic performance of complex M-estimators for multivariate location and scatter estimation," *IEEE Signal Processing Letters*, vol. 26, no. 2, pp. 367–371, 2019.

[24] V. Ollier, M. N. El Korso, A. Ferrari, R. Boyer, and P. Larzabal, "Robust distributed calibration of radio interferometers with direction dependent distortions," *Signal Processing*, vol. 153, pp. 348–354, 2018.

[25] A. Bouiba, M. N. El Korso, and T. Boukaba, "Robust joint relaxed maximum likelihood calibration and direction of arrivals estimation with sparse arrays," in *The 1st Conference on Electrical Engineering*, 2019.

[26] 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 Transactions on Signal Processing*, vol. 50, pp. 1908–1915, 2002.

[27] V. Ollier, M. N. El Korso, A. Ferrari, R. Boyer, and P. Larzabal, "Bayesian calibration using different prior distributions: an iterative maximum a posteriori approach for radio interferometers," in *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE, 2018, pp. 2673–2677.

[28] A. Breloy, M. N. El Korso, A. Panahi, and H. Krim, "Robust subspace clustering for radar detection," in *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE, 2018, pp. 1602–1606.

[29] V. Ollier, M. N. El Korso, A. Ferrari, R. Boyer, and P. Larzabal, "Robust calibration of radio interferometers in multi-frequency scenario," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 3494–3498.

[30] A. Breloy, Y. Sun, P. Babu, G. Ginolhac, D. P. Palomar, and F. Pascal, "A robust signal subspace estimator," *Statistical Signal Processing Workshop (SSP),IEEE*, pp. 1–4, 2016.

[31] M. Brossard, M. N. El Korso, M. Pesavento, R. Boyer, P. Larzabal, and S. J. Wijnholds, "Parallel multi-wavelength calibration algorithm for radio astronomical arrays," *Signal Processing*, vol. 145, pp. 258–271, 2018.

[32] G. Ginolhac and P. Forster, "Approximate distribution of the low-rank adaptive normalized matched filter test statistic under the null hypothesis," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 4, pp. 2016–2023, 2016.

[33] M. Rangaswamy, F. C. Lin, and K. R. Gerlach, "Robust adaptive signal processing methods for heterogeneous radar clutter scenarios," *Signal Processing*, vol. 84, no. 9, pp. 1653–1665, 2004.

[34] K. Greenewald, E. Zelnio, and A. O. H. III, "Kronecker PCA based robust SAR STAP," *arXiv preprint arXiv:1501.07481*, 2015.

[35] K. V. Mardia and P. E. Jupp, "Distributions on spheres," *Directional Statistics*, 2000.

[36] Y. Chikuse, *Statistics on Special Manifold*. New York: Springer-Verlag, 2003.

[37] C. Khatri and K. V. Mardia, "The von Mises–Fisher matrix distribution in orientation statistics," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 95–106, 1977.

[38] A. Kume and A. T. Wood, "Saddlepoint approximations for the Bingham and Fisher–Bingham normalising constants," *Biometrika*, vol. 92, no. 2, pp. 465–476, 2005.

[39] D. De Waal, "On the normalizing constant for the Bingham-von Mises-Fisher matrix distribution," *South African Statistical Journal*, vol. 13, no. 2, pp. 103–112, 1979.

[40] P. D. Hoff, "Simulation of the matrix Bingham-von Mises-Fisher distribution with applications to multivariate and relational data," *Journal of Computational and Graphical Statistics*, vol. 18, no. 2, pp. 438–456, 2009.

[41] J. T. Kent, A. M. Ganeiber, and K. V. Mardia, "A new method to simulate the Bingham and related distributions in directional data analysis with applications," *arXiv preprint arXiv:1310.8110*, 2013.

[42] O. Besson, N. Dobigeon, and J. Y. Tourneret, "Minimum mean square distance estimation of a subspace," *arXiv:1101.3462v1*, 2011.

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

[44] A. Aubry, A. De Maio, and L. Pallotta, "A geometric approach for structured radar covariance estimation," *Radar Conference (RadarConf), 2017 IEEE*, pp. 0767–0771, 2017.

[45] Y. Sun, A. Breloy, P. Babu, D. P. Palomar, F. Pascal, and G. Ginolhac, "Low-complexity algorithms for low rank clutter parameters estimation in radar systems," *IEEE Transactions on Signal Processing*, vol. 64, no. 8, pp. 1986–1998, April 2016.

[46] T. Bao, M. N. El Korso, and H. H. Ouslimani, "Cramér–rao bound and statistical resolution limit investigation for near-field source localization," *Digital Signal Processing*, vol. 48, pp. 137–147, 2016.

[47] M. N. El Korso, A. Renaux, R. Boyer, and S. Marcos, "Deterministic performance bounds on the mean square error for near field source localization," *IEEE Transactions on Signal Processing*, vol. 61, no. 4, pp. 871–877, 2012.

[48] J. Liu, M. S. Shbat, and V. Tuzlukov, "Interference cancellation and DOA estimation by generalized receiver applying LMS and MUSIC algorithms," *Progress In Electromagnetics Research Symposium Proceedings*, pp. 187–190, 2013.

[49] R. Grover, D. A. Pados, and M. J. Medley, "Subspace direction finding with an auxiliary-vector basis," *IEEE Transactions on Signal Processing*, vol. 55, no. 2, pp. 758–763, 2007.

[50] M. Haardt, M. Pesavento, F. Roemer, and M. N. El Korso, "Subspace methods and exploitation of special array structures," *Academic Press Library in Signal Processing*, vol. 3, pp. 651–717, 2014.

[51] B. Meriaux, A. Breloy, C. Ren, M. N. El Korso, and P. Forster, "Modified sparse subspace clustering for radar detection in non-stationary clutter," in *8th IEEE Interna-tional Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP 2019)*, 2019.

[52] A. Mennad, A. Younsi, M. N. El Korso, and A. M. Zoubir, "Adaptive detection of range-spread target in compound-Gaussian clutter without secondary data," *Digital Signal Processing*, vol. 60, pp. 90–98, 2017.

[53] R. B. Abdallah, A. Breloy, A. Taylor, M. El Korso, and D. Lautru, "Signal subspace change detection in structured covariance matrices," in *2019 27th European Signal Processing Conference (EUSIPCO)*. IEEE, 2019, pp. 1–5.

[54] P. Zarka, M. Tagger, L. Denis, J. Girard, A. Konovalenko, M. Atemkeng, M. Arnaud, S. Azarian, M. Barsuglia, A. Bonafede *et al.*, "Nenufar: Instrument description and science case," in *2015 International Conference on Antenna Theory and Techniques (ICATT)*. IEEE, 2015, pp. 1–6.

[55] S. Kritchman and B. Nadler, "Determining the number of components in a factor model from limited noisy data," *Chemometrics and Intelligent Laboratory Systems*, vol. 94, no. 1, pp. 19–32, 2008.

[56] P. O. Perry and P. J. Wolfe, "Minimax rank estimation for subspace tracking," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 3, pp. 504 – 513, 2010.

[57] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, "On detection of the number of signals in presence of white noise," *Journal of Multivariate Analysis*, vol. 25, no. 1, pp. 276–280, 1986.

[58] N. A. Goodman and J. M. Stiles, "On clutter rank observed by arbitrary arrays," *IEEE Transactions on Signal Processing*, vol. 55, no. 1, pp. 178–186, Jan 2007.

[59] G. Ginolhac and P. Forster, "Approximate distribution of the low-rank adaptive normalized matched filter test statistic under the null hypothesis," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 4, pp. 2016 – 2023, 2016.

[60] L. E. Brennan and F. Staudaher, "Subclutter visibility demonstration," Tech. Rep., 1992.

[61] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 61, no. 3, pp. 611–622, 1999.

[62] 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, ICASSP 2007. IEEE International Conference on*, vol. 3, 2007, pp. 1105–1108.

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

[64] E. Ollila and D. E. Tyler, "Regularized m-estimators of scatter matrix," *IEEE Transactions on Signal Processing*, vol. 62, no. 22, pp. 6059–6070, 2014.

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

[66] O. Besson, N. Dobigeon, and J. Y. Tourneret, "Minimum mean square distance estimation of a subspace," *IEEE Transactions on Signal Processing*, vol. 59, no. 12, pp. 5709–5720, 2011.

[67] O. Besson, N. Dobigeon, and J. Tourneret, "Joint Bayesian estimation of close subspaces from noisy measurements," *IEEE Signal Processing Letters*, vol. 21, pp. 168–171, 2014.

[68] A. Srivastava, "A Bayesian approach to geometric subspace estimation," *IEEE Transaction on Signal Processing*, vol. 48, no. 5, pp. 1390–1400, 2000.

[69] Y. Sun, A. Breloy, P. Babu, D. P. Palomar, F. Pascal, and G. Ginolhac, "Low-complexity algorithms for low rank clutter parameters estimation in radar systems," *IEEE Transactions on Signal Processing*, vol. 64, no. 8, pp. 1986–1998, 2016.

[70] J. Ward, "Space time adaptive processing for airborne radar," MIT, Lexington, Mass., USA, Tech. Rep., December 1994.

[71] O. Besson, "Bounds for a mixture of low-rank compound-Gaussian and white Gaussian noises," *IEEE Transactions on Signal Processing*, vol. 64, no. 21, pp. 5723–5732, 2016.

[72] R. S. Raghavan, "Statistical interpretation of a data adaptive clutter subspace estimation algorithm," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 2, pp. 1370–1384, April 2012.

[73] J. Sherman and W. J. Morrison, "Adjustment of an inverse matrix corresponding to a change in one element of a given matrix," *The Annals of Mathematical Statistics*, vol. 21, no. 1, pp. 124–127, 1950.

[74] M. Woodbury, "Inverting modified matrices (memorandum rept., 42, statistical research group)," 1950.

[75] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing communications and machine learning," *IEEE Transaction on Signal Processing*, vol. 65, pp. 794–816, 2016.

[76] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *The American Statistician*, vol. 58, no. 1, pp. 30–37, 2004.

[77] J. K. Thomas, L. L. Scharf, and D. W. Tufts, "The probability of a subspace swap in the svd," *IEEE Transactions on Signal Processing*, vol. 43, no. 3, pp. 730–736, 1995.

[78] R. B. Abdallah, A. Breloy, M. N. E. Korso, D. Lautru, and H. H. Ouslimani, "Minimum mean square distance estimation of subspaces in presence of Gaussian sources with application to STAP detection," *Journal of Physics: Conference Series, IOP Publishing*, vol. 904, no. 1, p. 012010, 2017.

[79] J. P. Ovarlez, F. L. Chevalier, and S. Bidon, "Les données de club STAP, introduction au STAP," *Traitement de Signal*, vol. 28, 2011.

[80] X. Mestre and L. M. Ángel, "Modified subspace algorithms for DoA estimation with large arrays," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 598–614, 2008.

[81] A. Combernoux, F. Pascal, G. Ginolhac, and M. Lesturgie, "Random matrix theory applied to low rank STAP detection," pp. 1–5, 2013.

[82] A. T. Wood, "Estimation of the concentration parameters of the Fisher matrix distribution on 50 (3) and the Bingham distribution on sq," *Australian Journal of Statistics*, vol. 35, no. 1, pp. 69–79, 1993.

[83] K. Conradsen, A. A. Nielsen, J. Schou, and H. Skriver, "A test statistic in the complex Wishart distribution and its application to change detection in polarimetric SAR data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 1, pp. 4–19, 2003.

[84] L. M. Novak, "Change detection for multi-polarization multi-pass SAR," *Algorithms for Synthetic Aperture Radar Imagery XII*, vol. 5808, no. 1, p. 234–247, 2005.

[85] S. W. Chen, X. S. Wang, and M. Sato, "Polinsar complex coherence estimation based on covariance matrix similarity test," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 11, pp. 4699–4710, Nov 2012.

[86] M. Liu, H. Zhang, C. Wang, and F. Wu, "Change detection of multilook polarimetric SAR images using heterogeneous clutter models," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 12, pp. 7483–7494, Dec 2014.

[87] V. Carotenuto, A. De Maio, C. Clemente, and J. Soraghan, "Unstructured versus structured for multipolarization SAR change detection," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 8, pp. 1665–1669, 2015.

[88] A. A. Nielsen, K. Conradsen, and H. Skriver, "Change detection in full and dual polarization, single- and multifrequency SAR data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 8, pp. 4041–4048, Aug 2015.

[89] ——, "Omnibus test for change detection in a time sequence of polarimetric SAR data," *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2016.

[90] A. De Maio, D. Orlando, L. Pallotta, and C. Clemente, "A multifamily GLRT for oil spill detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 1, p. 63–79, 2017.

[91] A. Taylor, H. Oriot, P. Forster, and F. Daout, "Reducing false alarm rate by testing proportionality of covariance matrices." 2017 International Radar Conference, 2017.

[92] A. Taylor, P. Forster, F. Daout, H. Oriot, and L. Savy, "A generalization of the fixed point estimate for packet-scaled complex covariance matrix estimation," *IEEE Transactions on Signal Processing*, vol. 65, no. 20, pp. 5393–5405, 2017.

[93] G. Vasile, J. P. Ovarlez, F. Pascal, and C. Tison, "Coherency matrix estimation of heterogeneous clutter in high-resolution polarimetric SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 4, pp. 1809–1826, April 2010.

[94] P. Formont, F. Pascal, G. Vasile, J.-P. Ovarlez, and L. Ferro-Famil, "Statistical classification for heterogeneous polarimetric SAR images," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 3, p. 567–576, 2011.

[95] M. Liu, H. Zhang, and C. Wang, "Change detection in urban areas of high-resolution polarization SAR images using heterogeneous clutter models," *3rd International Asia-Pacific Conference on Synthetic Aperture Radar (APSAR)*, 2011.

[96] A. Mian, J.-P. Ovarlez, G. Ginolhac, and A. Atto, "A robust change detector for highly heterogeneous images." 2018 IEEE International Conference on Acoustics, Speech and Signal Processing, 2018.

[97] P. C. O'Brien, "Robust procedures for testing equality of covariance matrices," *Biometrics*, pp. 819–827, 1992.

[98] F. T. Walterand *et al.*, "Testing proportionality of covariance matrices," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 102–106, 1951.

[99] B. Kang, V. Monga, and M. Rangaswamy, "Rank-constrained maximum likelihood estimation of structured covariance matrices," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 1, pp. 501–515, January 2014.

[100] R. B. Abdallah, A. Mian, A. Breloy, A. Taylor, M. N. El Korso, and D. Lautru, "Detection methods based on structured covariance matrices for multivariate SAR images processing," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 7, pp. 1160–1164, 2019.

[101] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Transactions on Signal Processing*, vol. 65, no. 3, pp. 794–816, Feb 2017.

[102] D. Ratha, S. De, T. Celik, and A. Bhattacharya, "Change detection in polarimetric SAR images using a geodesic distance between scattering mechanisms," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 7, pp. 1066–1070, July 2017.

[103] A. Mian, J.-P. Ovarlez, G. Ginolhac, and A. Atto, "Multivariate change detection on high resolution monovariate SAR image using linear time-frequency analysis," *25th European Signal Processing Conference (EUSIPCO)*, pp. 1942–1946, 2017.

[104] M. Rangaswamy, F. C. Lin, and K. R. Gerlach, "Robust adaptive signal processing methods for heterogeneous radar clutter scenarios," *Signal Processing*, vol. 84, no. 9, pp. 1653–1665, 2004.