Robust estimation analysis for signal and image processing

Gordana Draskovic

université PARIS-SACLAY

CentraleSupélec

NNT : 2019SACLC069

Thèse de doctorat

Statistiques des estimateurs robustes pour le traitement du signal et des images

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

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

Spécialité de doctorat : Traitement du signal et des images

Thèse présentée et soutenue à Gif-sur-Yvette, le 27 septembre 2019, par

Gordana DRAŠKOVIĆ

Composition du jury :

Patrick Flandrin DR CNRS, ENS de Lyon	Président
Cédric Richard Professeur des Universités, Université Nice Sophia Antipolis	Rapporteur
Olivier Michel Professeur des Universités, GIPSA Lab, INP de Grenoble	Rapporteur
Jean-Yves Tourneret Professeur des Universités, INP-ENSEEIHT de Toulouse	Examinateur
Michèle Sebag DR CNRS, LRI	Examinateur
Frédéric PASCAL Professeur, CentraleSupélec, L2S	Directeur de thèse
Florence TUPIN Professeur, Télécom Paris, LTCI	Co-directrice de thèse
Matthieu Jonckheere Professeur, UBA, Argentine	Invité

UNIVERSITÉ PARIS-SACLAY

ÉCOLE DOCTORALE STIC
« Sciences et Technologies de l'Information et de la Communication »

THÈSE DE DOCTORAT
Spécialité : Traitement du Signal et des Images

Soutenue le 27 septembre 2019 à Gif-sur-Yvette
par
Gordana DRAŠKOVIĆ

Statistiques des estimateurs robustes pour le traitement du signal et des images

Robust estimation analysis for signal and image processing

Composition du jury :

Président du jury :	Patrick Flandrin	Directeur de recherche CNRS, ENS de Lyon
Rapporteurs :	Cédric Richard	Professeur des Universités Université Nice Sophia Antipolis
	Olivier Michel	Professeur des Universités GIPSA Lab, INP de Grenoble
Examinateurs :	Jean-Yves Tourneret	Professeur des Universités INP-ENSEEIHT de Toulouse
	Michèle Sebag	Directeur de recherche CNRS, LRI
Directeurs de thèse :	Frédéric Pascal	Professeur L2S, CentraleSupélec
	Florence Tupin	Professeur LTCI, Télécom Paris
Invité :	Matthieu Jonckheere	Professeur Université de Buenos Aires

Acknowledgements

First of all, I would like to thank my supervisors Frédéric Pascal and Florence Tupin without whom this thesis would not have been possible. Fred, I cannot express how thankful I am for all suggestions, ideas, and many useful discussions that we have had. My success would not have been imaginable without your guidance and persistent help. Florence, thank you for all constructive comments and encouragement. Without you, the practical part of this thesis would have never been accomplished.

I am deeply grateful to the examination committee. Many thanks to Cédric Richard and Olivier Michel for a detailed pre-examination of the thesis, to Jean-Yves Tourneret, Michèle Sebag and Matthieu Jonckheere for agreeing to be the examiners and to Patrick Flandrin for accepting to be the chair. I am also very grateful to my mid-term committee Guillaume Ginolhac and Sylvie Marcos for their suggestions and support.

I owe my deepest gratitude to Arnaud Breloy: my mentor, my colleague and my friend. Arnaud, your advices and comments were an enormous help to me. Thank you for all nice talks and warm encouragement. I am particularly grateful for all the Palmashow and Kaamelott sessions! :) I have definitely learned a lot from you!

I would also like to express my gratitude to the members of Signals and systems laboratory that made my stay there great. I owe a very important debt to Odette Le Roux, Maryvonne Giron, Antonio Loria and Elena Panteley that welcomed me warmly and made my first days much easier. I would like to thank Silviu Niculescu and Pascal Bondon for their attention and support over the past years. Many thanks to the lab staff for their availability and assistance. My deepest appreciation goes to Stephanie Douesnard and Audrey Bertinet for their unfailing support and all the positive energy they have brought to the lab. Thanks to all my colleagues for all the great moments we spent together. Thank you Violeta, Milan, Walid, Alina, Yassine, Weichao, Ralph, Vincent, Elena, Incha,... Special thanks go to Lucien and Kuba, my PhD students co-delegates. Thank you for a productive exchange of ideas and all the work we have done together. It was great sharing laboratory with all of you during last three years.

I would like to extend my sincere thanks to my colleagues from the Laboratory for communication and processing of information: Clément, Chenguang, Xiangli, David, Weiying, Sylvain. I really appreciated the group meetings and *pauses-café* that we had. I am particularly grateful for the assistance given by Professors Henri Maitre, Jean-Marie Nicolas, and Michel Roux. Thank you all for your help and a true team spirit!

My thanks go also to Claude Delpha, Dima Rodriguez, Patrick Ruiz, Sylvie Dugard, and Cédric Bernez from Polytech Paris-Sud. Thank you for your help and support during my teaching mission. I would also like to thank my colleagues from the working group *Vie des doctorants*, especially to Betty Brès who profoundly believed in my abilities. Thank you for helping me overcome my fears and find myself.

I cannot express my gratitude to Jean-Phi, Fred, Arnaud, Guillaume, Lucien, Alex, Chengfang, Nabil, Gilles, and Ammar for the time we spent together. You have turned every conference into an unforgettable experience!

I had a great pleasure of working with Professor Esa Ollila from Aalto University and his former PhD student Shahab Basiri. I would like to thank them for the great collaboration we have had and their wonderful hospitality in Finland. Another collaboration that I am

deeply grateful for has been done with Matthieu Jonckheere from Buenos Aires University. The time I spent in Argentina was incredibly fruitful for me. Matt, thank you for bringing me into the world of machine learning. Thank you for all discussions, seminars, for all interesting people I met thanks to you, thank you for the time spent in Uruguay. I am also extremely grateful to Roizman family that welcomed me with open arms and made my stay in Argentina even more amazing. Thank you with all my heart!

I would also like to thank my flatmate Bogdan for being there from the very beginning. The thesis will certainly remain an extraordinary experience for both of us.

Thanks to my friends Mara, Bibi, Ivana, Roske, Nidzan, and Aco for their daily messages full of optimism and pure love. A very special gratitude goes to my *Parisian* parents, Jelena and Zoran. There are no words to express how grateful I am for all you have done and you are doing for me. Thank you for being who you are! Thanks to Sneza and her lovely family. Thank you for your innocent love and for reminding us how beautiful our Pozega is.

A tremendous amount of gratitude goes to my parents Vera and Milan, my brother Marko, my sister Branka, and my grandparents Dragomir, Drinka, Vinka, and Branko. You are my greatest support and my strength. This thesis is dedicated to you!

And last but not least, thank you Mauricio. Thank you for coming into my life, thank you for every day we spend together, thank you for all the love and attention you give me. As you once said: *Thanks to you, I returned to innocence...*

Gordana

Résumé

Un des défis majeurs en traitement radar consiste à identifier une cible cachée dans un environnement bruité. Pour ce faire, il est nécessaire de caractériser finement les propriétés statistiques du bruit, en particulier sa matrice de covariance. Sous l'hypothèse gaussienne, cette dernière est estimée par la matrice de covariance empirique (SCM) dont le comportement est parfaitement connu. Cependant, dans de nombreuses applications actuelles, tels les systèmes radar modernes à haute résolution par exemple, les données collectées sont de nature hétérogène, et ne peuvent être proprement décrites par un processus gaussien. Pour pallier ce problème, les distributions symétriques elliptiques complexes, caractérisant mieux ces phénomènes physiques complexes, ont été proposées.

Dans ce cas, les performances de la SCM sont très médiocres et les M-estimateurs apparaissent comme une bonne alternative, principalement en raison de leur flexibilité par rapport au modèle statistique et de leur robustesse aux données aberrantes et/ou aux données manquantes. Cependant, le comportement de tels estimateurs reste encore mal compris. Dans ce contexte, les contributions de cette thèse sont multiples.

D'abord, une approche originale pour analyser les propriétés statistiques des M-estimateurs est proposée, révélant que les propriétés statistiques des M-estimateurs peuvent être bien approximées par une distribution de Wishart.

Grâce à ces résultats, nous analysons la décomposition de la matrice de covariance en éléments propres. Selon l'application, la matrice de covariance peut posséder une structure particulière impliquant valeurs propres multiples contenant les informations d'intérêt. Nous abordons ainsi divers scénarios rencontrés dans la pratique et proposons des procédures robustes basées sur des M-estimateurs.

De plus, nous étudions le problème de la détection robuste du signal. Les propriétés statistiques de diverses statistiques de détection adaptative construites avec des M-estimateurs sont analysées.

Enfin, la dernière partie de ces travaux est consacrée au traitement des images radar à synthèse d'ouverture polarimétriques (PolSAR). En imagerie PolSAR, un effet particulier appelé speckle dégrade considérablement la qualité de l'image. Dans cette thèse, nous montrons comment les nouvelles propriétés statistiques des M-estimateurs peuvent être exploitées afin de construire de nouvelles techniques pour la réduction du speckle.

Mots-clés : Estimation robuste, distributions CES, Loi de Wishart, détection du signal, images PolSAR.

Abstract

One of the main challenges in radar processing is to identify a target hidden in a disturbance environment. To this end, the noise statistical properties, especially the ones of the disturbance covariance matrix, need to be determined. Under the Gaussian assumption, the latter is estimated by the sample covariance matrix (SCM) whose behavior is perfectly known. However, in many applications, such as, for instance, the modern high resolution radar systems, collected data exhibit a heterogeneous nature that cannot be adequately described by a Gaussian process. To overcome this problem, Complex Elliptically Symmetric distributions have been proposed since they can correctly model these data behavior.

In this case, the SCM performs very poorly and M-estimators appear as a good alternative, mainly due to their flexibility to the statistical model and their robustness to outliers and/or missing data. However, the behavior of such estimators still remains unclear and not well understood. In this context, the contributions of this thesis are multiple.

First, an original approach to analyze the statistical properties of M-estimators is proposed, revealing that the statistical properties of M-estimators can be approximately well-described by a Wishart distribution.

Thanks to these results, we go further and analyze the eigendecomposition of the covariance matrix. Depending on the application, the covariance matrix can exhibit a particular structure involving multiple eigenvalues containing the information of interest. We thus address various scenarios met in practice and propose robust procedures based on M-estimators.

Furthermore, we study the robust signal detection problem. The statistical properties of various adaptive detection statistics built with M-estimators are analyzed.

Finally, the last part deals with polarimetric synthetic aperture radar (PolSAR) image processing. In PolSAR imaging, a particular effect called speckle significantly degrades the image quality. In this thesis, we demonstrate how the new statistical properties of M-estimators can be exploited in order to build new despeckling techniques.

Keywords: Robust estimation, CES distributions, Wishart distribution, signal detection, PolSAR imaging.

Contents

List of Figures

List of Tables

Abbreviations

i.e.	id est (that is)
e.g.	exempli gratia (for example)
w.r.t.	with respect to
w.l.o.g.	without loss of generality
r.v.	random vector
r.va.	random variable
i.i.d.	Independent and Identically Distributed
p.d.f.	Probability Density Function
c.d.f.	Cumulative Distribution Function
SA	Standard Asymptotic
CLT	Central Limit Theorem
RN	Real Normal
CN	Complex Normal
GCN	Generalized Complex Normal
CCG	Complex Compound-Gaussian
SIRV	Spherically Invariant Random Vector
CACG	Comlex Angular Central Gaussian
CGG	Complex Generalized Gaussian
RES	Real Elliptically Symmetric
CES	Complex Elliptically Symmetric
DoF	Degrees of Freedom
IG	Inverse Gamma
SCM	Sample Covariance Matrix
MLE	Maximum Likelihood Estimator
GCWE	Gaussian Core Wishart Equivalent
EVD	EigenValue Decomposition
PCA	Principal Component Analysis
MSE	Mean Squared Error
ICRLB	Intrinsic Cramer-Rao Lower Bound
B-ICRLB	Biased ICRLB
aB-ICRLB	approximated B-ICRLB
RSCM	Regularized Sample Covariance Matrix
SD	Standard Deviation
IR	Iteratively Reweighting
IRLS	Iteratively Reweighted Least Squares
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
MDL	Minimum Description Length
STAP	Space-Time Adaptive Process

GLRT	Generalized Likelihood Ratio Test
SNR	Signal to Noise Ratio
PFA	Probability of False Alarm
CFAR	Constant False Alarm Rate
AMF	Adaptive Matched Filter
NMF	Normalized Matched Filter
ANMF	Adaptive NMF
ACE	Adaptive Cosine Estimator
LR	Low-Rank
LR-ANMF	Low-Rank ANMF
AWGN	Additive White Gaussian Noise
GCED	Gaussian Core Equivalent Detector
GCEM	Gaussian Core Equivalent Mahalanobis
SAR	Synthetic Aperture Radar
PolSAR	Polarimetric SAR
LLMMSE	Local Linear Minimum MSE
IDAN	Intensity-Driven Adaptive Neighborhood
PDE	Partial Differential Equation
AD	Anisotropic Diffusion
TV	Total Variation
NL	Non-Local
NLM	Non-Local Means
NLRB	Non-Local Reduced Bias
MRF	Markov Random Field
GP	Gradient Preservation
EP	Edge Preservation

Notation

General notation

$\mathbb{R}$	Field of real numbers
$\mathbb{R}^+$	Field of positive real numbers
$\mathbb{R}_0^+$	Field of non-negative real numbers
$\mathbb{C}$	Field of complex numbers
$\Re(\cdot)$	Real part of a complex number
$\Im(\cdot)$	Imaginary part of a complex number
$\mathbf{a}$	Boldface and lower case letter denotes a vector
$\mathbf{A}$	Boldface and upper case letter denotes a matrix
$\mathbf{I}$	Identity matrix
$\mathbf{K}$	Commutation matrix which transforms $\mathrm{vec}(\mathbf{A})$ into $\mathrm{vec}(\mathbf{A}^T)$, i.e. $\mathbf{K}\,\mathrm{vec}(\mathbf{A}) = \mathrm{vec}(\mathbf{A}^T)$
$(\cdot)^T$	Transpose operator
$(\cdot)^*$	Complex conjugate operator
$(\cdot)^H$	Hermitian (transpose) conjugate
$(\cdot)^+$	Moore–Penrose pseudo-inverse
$(\cdot)^{1/2}$	Hermitian square root
$\|\cdot\|$	Euclidean/Frobenius norm
$\|\cdot\|$	Determinant of a matrix
$\mathrm{Tr}(\cdot)$	Trace operator
$\mathrm{rank}(\cdot)$	Rank of a matrix
$\mathbf{A} \otimes \mathbf{B}$	Kronecker product of two matrices
$\mathrm{diag}(\mathbf{a})$	Matrix with elements of the vector $\mathbf{a}$ on the main diagonal
$\mathbb{1}$	Indicator function
$\mathbf{1}$	Column vector of ones
$\mathbf{0}$	Column vector of zeros
$[\mathbf{A}]_{i,j}$	(i,j) element of the matrix $\mathbf{A}$
$\mathrm{vec}(\cdot)$	Vectorization of a matrix
$\exp(\cdot)$	Exponential function
$\ln(\cdot)$	Natural logarithm
$\log(\cdot)$	Logarithm with base 10
$(\cdot)'$	Derivative of a function
$\Gamma(\cdot)$	Gamma function
$\beta(a,b)$	Beta distribution with shape parameters a and b
$\beta'(a,b)$	Beta prime distribution with shape parameters a and b
$\chi^2(d)$	Chi-squared distribution with d degrees of freedom

$\mathcal{CS}$	Set of complex symmetric matrices
$\mathcal{H}$	Set of positive semidefinite Hermitian matrices
$\mathcal{H}^+$	Set of positive definite Hermitian matrices
$\mathcal{U}_r^p$	Stiefel manifold (set of semi-unitary matrices)
$\mathcal{G}_r^p$	Set of rank r orthogonal projectors of $\mathbb{C}^{p\times p}$

Notation related to statistical models of random vectors

p	Vector dimension
n	Sample size
$\boldsymbol{\mu}$	Mean
$\mathbf{C}$	Covariance matrix
$\widehat{\mathbf{C}}$	Covariance matrix estimator
$\mathbf{P}$	Pseudo-covariance matrix
$\boldsymbol{\Sigma}$	Scatter matrix
$\widehat{\boldsymbol{\Sigma}}$	Scatter matrix estimator
$\mathbb{P}(\cdot)$	Probability
$\mathbb{E}(\cdot)$	Expectation
$\sim$	Is distributed as
$\overset{d}{=}$	Shares the same distribution as
$\overset{d}{\to}$	Convergence in distribution
$\mathcal{N}$	Real Normal distribution
$\mathcal{CN}$	Complex Normal distribution
$\mathcal{GCN}$	Generalized Complex Normal distribution
$\mathcal{CW}_p$	Complex Wishart distribution
$\mathcal{ES}$	Real Elliptically Symmetric distribution
$\mathcal{CES}$	Complex Elliptically Symmetric distribution
$\mathcal{GCES}$	Generalized Complex Elliptically Symmetric distribution
$\mathcal{CCG}$	Complex Compound-Gaussian distribution
$\mathcal{CACG}$	Complex Angular Central Gaussian distribution
$\mathcal{U}(\mathcal{CS}^{k-1})$	Uniform distribution on the unit Complex k-hypersphere
$\mathcal{C}t_\nu$	Complex Student's t-distribution
$\mathcal{CK}_\nu$	Complex K-distribution
$\mathcal{CW}_{s,b}$	Complex W-distribution
$\mathcal{CGG}_{s,b}$	Complex Generalized Gaussian distribution

Notation related to signal detection

H_0	Null hypothesis
H_1	Alternative hypothesis
P_d	Detection probability
P_{fa}	Probability of false alarm
δ	Signal to noise ratio
$\mathbf{p}$	Steering vector
Λ	Detection statistics
$\widehat{\Lambda}$	Adaptive detector
λ	Detection threshold

Notation related to PolSAR imagery

$\mathbf{S}$	Scattering matrix
$\mathbf{k}$	Scattering vector
H	Entropy
A	Anisotropy
L	Number of looks
L_e	Equivalent number of looks
I	Single polarization intensity
$\mathbb{W}$	Search window size
$\mathbb{P}$	Patch size
$\mathbb{S}$	Scale size

Introduction

Main objectives in data processing are either to extract some knowledge (information of interest) from the data set (this is commonly included in the "Data-to-Knowledge" domain) or to help making a decision (included in the "Data-to-Decision" domain). Among the main branches of statistical signal processing, a large amount of research activities are dedicated to estimation, detection, and signal analysis. In estimation problem, one seeks at estimating either the value(s) of one or more parameters, or some functions of interest, from a set of measurements (also called observations or data). This thesis focuses on the first class of problem, referred to as parametric estimation problems. In such cases, these measurements are generally considered as noisy data whose underlying distribution depends on the parameters to determine. One of the widely assumed statistical models is the Gaussian, or normal, one, whose distribution is fully characterized by its first and second-order moments. Usually, for signal processing applications, the first-order moment, or mean, is supposed to be known or equal to zero. Thus, the remaining problem consists in estimating the second-order moments. A complete second-order statistical characterization of complex random Gaussian vectors is given by covariance and pseudo-covariance matrix. Furthermore, it is important to keep in mind that the covariance matrix is involved in many applications beyond the signal processing community: in machine learning (various problems such as data visualization, data clustering, dimension reduction), in finance (e.g. portfolio analysis), in medicine (e.g. genome analysis), etc.

In most applications, it has often been assumed that complex random signals are proper and circular. A proper complex random variable is uncorrelated with its complex conjugate, and a circular complex random variable has a probability distribution that is invariant under rotation in the complex plane. These assumptions are convenient because they simplify computations and, in many aspects, make complex random signals look and behave like real random signals. In the context of complex vectors, the property means that the pseudo-covariance of the observation vectors vanishes. Hence, the only parameter to estimate is the covariance matrix.

Depending on the data model assumptions, many different algorithms to estimate the covariance matrix have been proposed. Under the standard Gaussian assumption, the optimal estimator is the sample covariance matrix (SCM) whose behavior and properties are perfectly known nowadays. In general, the Gaussian assumption is very convenient because it permits to derive explicit formulas for adaptive statistical methods. However, it often happens that real-world data cannot be completely described by normal distributions, i.e. the model fits well the majority of observations, while some observations can be considered as deviations from assumed model. Such atypical data are called outliers and they can be dangerous for many classical statistical procedures. Consequently, the corrupted data need more complex models to characterize their behavior. Moreover, according to the underlying physics, the data may not contain outliers but follow a "more heterogeneous" behavior than the Gaussian distribution. This is the case, for instance, in image processing where it is commonly assumed that wavelets coefficients behave as generalized Gaussian random variables. On can also mention radar (RAdio Detection And Ranging) processing where

the additive disturbance, or clutter, are generally heavy-tailed distributed (K-distribution, Weibull distribution, etc). A solution arises with complex elliptically symmetric (CES) distributions that are able to describe a wide range of heavy-tailed distributions. These distributions can also be useful for the "outliers" case. The CES distributions can be considered as a generalization of Gaussian distribution since they keep a lot of nice Gaussian properties. In this context, the second order characteristics are given by the scatter matrix of the data which is proportional to the covariance matrix, when the latter exists.

In a non-Gaussian framework, even a small amount of aberrations can significantly corrupt the covariance matrix estimation if the latter relies on the SCM. Therefore, in order to ensure a more accurate estimation of the scatter matrix, one needs to reduce the outliers influence or completely reject them during the estimation process. Robust statistics proposes different solutions to this problem. One of them is given by the general framework of M-estimation. Similarly to the CES distributions, that can be viewed as a generalization of the Gaussian model, M-estimation is a generalization of Maximum Likelihood (ML) estimation. These M-estimators are defined by implicit equations, or fixed-point equations, and can be obtained using iterative reweighting algorithms. However, the distribution of M-estimators remains unknown because of their fixed-point design. Therefore, an important challenge is to accurately analyze and better characterize the behavior of M-estimators. This is one of the main goals of this thesis.

The knowledge of M-estimators' distribution is essential in various signal processing applications. One of them is radar signal processing. Radar systems are used for object detection, tracking and analysis. They use electromagnetic waves that can interact with objects and ground surface. After a pulse signal is transmitted, the portion of energy, backscattered to the receiver, is analyzed in order to deduce the presence and the properties of an object of interest. In such applications, the performances of the corresponding methods strongly rely on the quality of the covariance matrix estimation and performance characterization depends on covariance matrix properties. In this thesis, we deal with two common problems in radar processing: radar detection and radar image processing.

The classic radar detection problem consists in identifying a useful complex signal containing the information about the target that is hidden in the clutter or other kind of disturbance. To that end, several detection statistics have been proposed in the literature. All of them require the knowledge of the disturbance (or clutter/noise) covariance matrix that is usually unknown. In order to estimate this matrix, a set of signal-free data referred to as secondary data is usually available. For a long time, these data have been assumed to be Gaussian-distributed. Increasing the resolution capabilities of radar systems caused the high heterogeneity of the data, which brought the need to use robust scatter matrix estimators to build adaptive detection statistics. Also, the secondary data does not always contain only signal-free data, but also some data similar to outliers. In general, the disturbance can be represented as a combination of the clutter caused by the environment and the white noise generated within the radar system. Moreover, the disturbance is often composed of a low-rank clutter. In this case, the number of required secondary data can be significantly reduced. Consequently, an approximate whitening in adaptive detectors is performed using a projector orthogonal onto the clutter subspace. This projector is typically estimated using the eigenvalue decomposition of a scatter matrix estimator. Consequently, in this thesis, we investigate the performances of robust adaptive detectors in order to improve the detection performances.

Similarly to conventional radars, imaging radars use radio waves to construct an image of the target scene. An example of imaging radar is Synthetic Aperture Radar (SAR). Each pixel in a SAR image contains the amplitude and the phase of the backscattered wave

expressed by a complex value. Moreover, SAR systems can transmit and receive signals in different orthogonal polarizations. Thus, instead of a simple scalar value, a pixel in polarimetric SAR (PolSAR) images is described by a complex vector. The value in each pixel is obtained as a sum of backscattered waves for one resolution cell, introducing the common phenomena in SAR imagery, called speckle. Speckle reduction is one of essential steps in SAR image processing. Under the Gaussian assumption, this step has been often performed by simple averaging of surrounding pixels. However, modern high resolution SAR systems produce higher variations between neighboring pixels that cannot be described by the Gaussian model. Consequently, more complex methods are required to deal with heterogeneous PolSAR data. In this thesis, we analyze how M-estimators can be used in order to improve the performances of PolSAR despeckling methods.

This thesis is organized as follows. In Chapter 1, the state-of-the-art methods for covariance matrix estimation, signal detection and PolSAR image despeckling are recapitulated. The first part is dedicated to the used data models and the corresponding covariance matrix estimators. First, the general definitions about random complex signals are recalled. Then, the standard Gaussian context is presented together with the definition and properties of the SCM. This is followed by a brief overview of the CES framework. Finally, the robust M-estimators are described and some particular cases are detailed. The second part is on detection theory, where various proposed detectors adapted to the Gaussian framework are recapitulated. Their properties are analyzed in terms of distribution under several scenarios, "probability of false alarm-threshold" relationships and detection probability versus Signal-to-Noise Ratio. Finally, the third part details the SAR technology, SAR polarimetry principles and reviews the state-of-the-art PolSAR despeckling methods.

The contributions of the thesis are summarized in the remaining four chapters as described in Figure 0.1. Chapter 2 investigates the statistical properties of M-estimators and provides the first important contribution of this work. M-estimators are given by fixed-point equations and only their properties in the standard asymptotic regime are known. In this chapter, we propose to analyze these estimators from another point of view, i.e., by comparing them with the well-known SCM. For this purpose, the stochastic representation of CES distributions is rewritten in such a way as to introduce the Gaussian cores of CES data. Then, a Gaussian-Core Wishart Equivalent (GCWE) of an M-estimator is defined as a theoretical equivalent built with fictive (in the sense that they cannot be observed) Gaussian cores of the measured data. Finally, the relationship between two types of estimators is derived. Generally speaking, the aim of this chapter is to derive a relationship between the standard Gaussian framework and robust procedures applied to corrupted data, in order to show that it is always better to use robust methods that offer better estimation performances, but keep the statistical properties of the classical Gaussian tools. Results advocates for robust methods as a perfect trade-off between robustness and attractive statistical properties.

Chapter 3 focuses on the estimation of the eigenvalue/eigenvector decomposition (EVD) of the scatter matrix. The scatter matrix can have various structures containing multiples eigenvalues and/or multiples blocks of equal eigenvalues. In practice, we can find many examples when this situation occurs, such as already mentioned low-rank detection or number of sources estimation. In this chapter, the EVD parameters, namely the eigenvalues and eigenvectors, of M-estimators are analyzed in various asymptotic regimes. Moreover, this analysis is extended to the principal subspaces obtained from M-estimators. It is shown that the conclusions of Chapter 2 are also valid in this case. Furthermore, an original robust method for eigenvalue estimation is developed. The algorithm is based on the idea to iteratively fuse equal eigenvalues thanks to a robust penalty function. The method is tested on simulated data and compared to a state-of-the-art method, showing an important performance improvement.

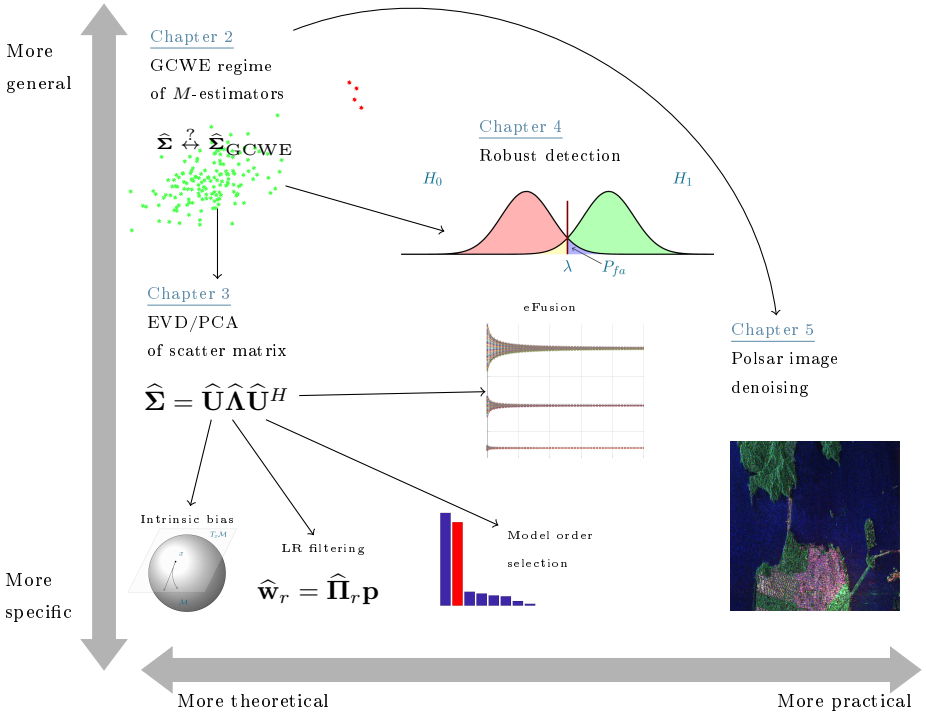


Figure 0.1 – Thesis organization: contributions

Chapter 4 deals with robust signal detection in a non-Gaussian framework. Robust detectors are defined as classical Gaussian-based detection statistics with an M-estimator plugged-in instead of the traditional SCM. Thanks to the analysis done in Chapter 2, we define a Gaussian-Core Equivalent Detector (GCED) of a robust detectors. The GCED of a robust detector corresponds to the same detection statistics built with the GCWE of the used M-estimator. Similarly to the analysis done in Chapter 2, the robust detectors are compared to their GCEDs, revealing that they can be well-described by the statistical properties of their GCEDs. These results are of great interest since they provide new insights in detection threshold computation and illustrate a direct application of the results obtained in Chapter 2.

Another possible application is given in Chapter 5. We address the problem of PolSAR despeckling. Reducing the speckle in PolSAR images correspond to the scatter matrix estimation of the scattering vector containing the information about different polarization channels. In non-local means (NLM) methods, the final estimation of the scatter matrix in a target pixel is obtained as a weighted mean computed with the surrounding pixels. The weights depend on the similarity between the target pixel and its neighbors. Since the PolSAR data can be modeled by different CES distributions, such as K-distribution or Weibull distribution, we propose to apply M-estimators in order to pre-estimate the scatter matrix and search for similar pixels in the image. More precisely, a new statistical method for weight computation in NLM methods is designed. The dissimilarities are computed using M-estimators and their new statistical properties derived in Chapter 2. The benefits and drawbacks of such a method are also discussed.

Chapter 1

State of the art

This chapter introduces the general context of the thesis. It is divided into three parts.

The first part focuses on the covariance matrix estimation. It starts with a brief introduction about complex random signals. It should be noted that, even we consider only the complex case, the same results with slight modifications are valid for the real-valued case. Then, the properties of the well-known SCM under the classical Gaussian assumption are recalled. Some widely used non-Gaussian models encompassed within the class of CES distributions are detailed. Finally, the robust alternatives to the SCM called M-estimators are discussed.

Second part is on signal detection. The problem of detecting a complex signal in a homogeneous environment is studied. Several detection statistics adapted to the Gaussian background are itemized. Their detection performances are also analyzed.

Finally, the last part deals with PolSAR imagery. First, the principles of SAR technology are explained. Then, the interest of radar polarimetry is illustrated. Speckle, common phenomena in (Pol)SAR imagery, is described. Moreover, the relationship between the PolSAR models and CES distributions is illustrated. This part is concluded with a brief overview of speckle reduction techniques.

1.1 Covariance matrix estimation

1.1.1 Introduction

The knowledge of the data second order statistics, usually described by the data covariance matrix, is of crucial importance in many different areas. For instance, in radar processing the knowledge of the data covariance matrix is indispensable for the construction of signal detection statistics. In image processing, where an image is usually assumed as a random process, this matrix can provide the information about the covariance between pixels within the image or between the pixels in different images sampled from the same distribution. The eigenvalue decomposition of the covariance matrix is very useful in image coding, segmentation, image classification, object recognition, etc. In general, in many signal processing applications, it is employed in whitening transformation in order to decorrelate the data. In machine learning, it is often used to represent the data in a more compact way thanks to the principal component analysis. The covariance matrix is also widely employed in financial economics, especially in portfolio theory where its information helps to maximize the expected return for a given level of risk. Moreover, the covariance matrix finds its application in biomedical data processing such as cardiac signal processing and neuroscience.

However, the true value of this parameter is usually unknown and needs to be estimated. In recent years, there has been a growing interest in its estimation in a vast amount of literature on this topic (see e.g., [130, 135, 27, 182, 134, 117, 162, 129] and references therein). In practice, we dispose of a certain amount of the data used for its estimation. The data are usually assumed to be independent and identically distributed (i.i.d.). Generally, in most of signal processing methods the data are modeled by a multivariate zero-mean Gaussian stochastic process. Multivariate Gaussian, also called normal, distribution plays a vital role in the theory of statistical analysis [77]. Very often the multivariate observations are only approximately normally distributed. This approximation is (asymptotically) valid even when the original data is not multivariate normal, due to the central limit theorem (CLT). In that case, the classical covariance matrix estimator is the sample covariance matrix (SCM) whose behavior is perfectly known. Indeed, it follows the Wishart distribution [13] which is the multivariate extension of the gamma distribution. Thanks to its explicit form, the SCM is easy to manipulate and therefore widely used in the signal processing community.

Nevertheless, the complex normality sometimes presents a poor approximation of underlying physics. Data can be corrupted by outliers, missing data or highly heterogeneous patterns that can not be described by the Gaussian distribution. For instance, the noise and interference in communication channels can be spiky and impulsive i.e., have heavier tails than the Gaussian distribution. Another example can be found in radar processing. With high resolution in modern radar technology, the Gaussian model becomes inappropriate for modeling a highly heterogeneous radar clutter. A growing body of literature has examined the fitting of non-Gaussian models to radar clutter [63, 67, 66, 36, 33, 179].

These models belong to the class of Complex Elliptically Symmetric (CES) distributions [130], originally introduced for the real case by Kelker [87]. In the context of CES distributions, the second order statistics are characterized by the scatter matrix which is proportional to the covariance matrix, when the latter exists. In this framework, the classical SCM can perform very poorly and does not provide a satisfying estimation of the scatter matrix.

An alternative is thus, to use more robust estimators, particularly M-estimators. Although Huber introduced robust M-estimators in [82] for the scalar case, Maronna provided the detailed analysis of the corresponding scatter matrix estimators in the multivariate real case in his seminal work [119]. M-estimators correspond to a generalization

of the well-known Maximum Likelihood estimators (MLE), that have been widely studied in the statistics literature [94, 5]. In contrast to MLEs where the estimating equation depends on the probability density function (p.d.f.) of a particular CES distribution, the weight function in the M-estimating equation can be completely independent of the data distribution. Consequently, M-estimators presents a wide class of scatter matrix estimators, including the MLEs, robust to the data model. In [119], it is shown that, under some mild assumptions, the estimator is defined as the unique solution of a fixed-point equation and that the robust estimator converges almost surely (a.s.) to a deterministic matrix, equal to the scatter matrix up to a scale quantity (depending on the true statistical model). Their asymptotic properties have been studied by Tyler in the real case [169]. The complex extension of M-estimators, more useful for signal processing applications, have been recently introduced and analyzed in [130, 117].

In the following, we first provide the essential definitions and the assumptions made in this thesis. Then, we consider the problem of estimating the second-order statistics in the Gaussian and non-Gaussian context.

1.1.2 Complex-valued signals

Let $\mathbf{z} = \mathbf{a} + j\mathbf{b}$ be a p-dimensional complex random vector which consists of a pair of real random vectors $\mathbf{a}$ and $\mathbf{b}$. The distribution of $\mathbf{z}$ on $\mathbb{C}^p$ determines the joint real $2p$-variate distribution of $\mathbf{a}$ and $\mathbf{b}$ on $\mathbb{R}^{2p}$ and conversely. To completely define the second-order moments of $\mathbf{a}$ and $\mathbf{b}$, $\mathbf{z}$ is given by its covariance matrix $\mathbf{C}$ and pseudo-covariance matrix $\mathbf{P}$.

Definition 1.1.1. *Covariance matrix*
The covariance matrix $\mathbf{C} \in \mathcal{H}$ of a complex r.v. $\mathbf{z} = \mathbf{a} + j\mathbf{b}$ is defined as

$$\mathbf{C} = \mathbb{E}\left[\mathbf{z}\mathbf{z}^H\right] = \mathbb{E}\left[\mathbf{a}\mathbf{a}^T\right] + \mathbb{E}\left[\mathbf{b}\mathbf{b}^T\right] + j\left(\mathbb{E}\left[\mathbf{b}\mathbf{a}^T\right] - \mathbb{E}\left[\mathbf{a}\mathbf{b}^T\right]\right). \tag{1.1}$$

Definition 1.1.2. *Pseudo-covariance matrix*
The pseudo-covariance matrix $\mathbf{P} \in \mathcal{CS}$ of a complex r.v. $\mathbf{z} = \mathbf{a} + j\mathbf{b}$ is defined as

$$\mathbf{P} = \mathbb{E}\left[\mathbf{z}\mathbf{z}^T\right] = \mathbb{E}\left[\mathbf{a}\mathbf{a}^T\right] - \mathbb{E}\left[\mathbf{b}\mathbf{b}^T\right] + j\left(\mathbb{E}\left[\mathbf{b}\mathbf{a}^T\right] + \mathbb{E}\left[\mathbf{a}\mathbf{b}^T\right]\right). \tag{1.2}$$

The first assumption that we make in this thesis is that the complex signals are circular. However, it should be pointed out that this assumption is not always true. Some analysis in this case can be found in [159].

Definition 1.1.3. *Circular symmetry*
An r.v. $\mathbf{z}$ is circular if

$$\mathbf{z} \overset{d}{=} e^{j\theta}\mathbf{z}, \quad \forall\theta \in \mathbb{R}. \tag{1.3}$$

If the distribution $f(\cdot)$ of $\mathbf{z}$ exists, it satisfies

$$f\left(e^{j\theta}\mathbf{z}\right) = f\left(\mathbf{z}\right), \quad \forall\theta \in \mathbb{R}. \tag{1.4}$$

If an r.v. $\mathbf{z}$ is circular in the sense defined above, then its pseudo-covariance matrix is equal to zero. This property is used to define the second-order circularity of $\mathbf{z}$.

Definition 1.1.4. *Second-order circularity*
An r.v. $\mathbf{z}$ is said to be second-order circular or proper if its pseudo-covariance matrix vanishes $\mathbf{P} = \mathbf{0}$.

In this thesis, we assume that complex signals possess the second-order circularity which we will refer to as circularity for the sake of simplicity. Consequently, we analyze only the covariance matrix estimators, assuming that the data pseudo-covariance matrix vanishes. In the case of non-circularity, the estimation of the pseudo-covariance matrix should be also considered to obtain full information second-order characteristics.

In order to define asymptotic normality of an estimator that will be highly employed in this work, we first recall the definitions of real and complex normal distribution.

Definition 1.1.5. *Real Normal (RN) distribution*
An r.v. $\mathbf{z}$ is said to have a real Normal (RN) distribution denoted as $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$, if its p.d.f. is given by

$$f(\mathbf{x}) = (2\pi)^{-p/2} |\mathbf{C}|^{-1/2} \exp\left(-\frac{(\mathbf{x}-\boldsymbol{\mu})^T \, \mathbf{C}^{-1} \, (\mathbf{x}-\boldsymbol{\mu})}{2} \right). \tag{1.5}$$

Analogously, one has the following definition.

Definition 1.1.6. *Generalized Complex Normal (GCN) distribution*
An r.v. $\mathbf{z}$ is said to have a generalized complex Normal (GCN) distribution, denoted as $\mathbf{x} \sim \mathcal{GCN}(\boldsymbol{\mu}, \mathbf{C}, \mathbf{P})$, if its p.d.f. is given by

$$f(\mathbf{z}) = \pi^{-p} |\widetilde{\mathbf{C}}|^{-1/2} \exp\left(-\frac{(\widetilde{\mathbf{z}}-\widetilde{\boldsymbol{\mu}})^H \, \widetilde{\mathbf{C}}^{-1} \, (\widetilde{\mathbf{z}}-\widetilde{\boldsymbol{\mu}})}{2} \right), \tag{1.6}$$

where $\widetilde{\mathbf{z}} = \begin{pmatrix} \mathbf{z} \\ \mathbf{z}^* \end{pmatrix}$, $\widetilde{\boldsymbol{\mu}} = \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu}^* \end{pmatrix}$ and $\widetilde{\mathbf{C}} = \begin{pmatrix} \mathbf{C} & \mathbf{P} \\ \mathbf{P}^* & \mathbf{C}^* \end{pmatrix}$.

If the r.v. $\mathbf{z}$ is circular, i.e. $\mathbf{P} = \mathbf{0}$, it has CN distribution. This case will be denoted as $\mathbf{z} \sim \mathcal{CN}(\boldsymbol{\mu}, \mathbf{C})$.

Note that the GCN distribution should not to be confused with the complex generalized Gaussian distribution defined in Definition 1.1.5.

Thanks to the multivariate CLT one has the following definition.

Definition 1.1.7. *Asymptotic Normal distribution*
Let be $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ i.i.d. random vectors with mean $\boldsymbol{\mu}$, finite covariance matrix $\mathbf{C}$ and pseudo-covariance matrix $\mathbf{P}$. Then, an estimator $\widehat{\mathbf{z}}$ has

- an asymptotic RN distribution if

$$\sqrt{n}\,(\widehat{\mathbf{z}} - \boldsymbol{\mu}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{C}), \tag{1.7}$$

- an asymptotic GCN distribution if

$$\sqrt{n}\,(\widehat{\mathbf{z}} - \boldsymbol{\mu}) \xrightarrow{d} \mathcal{GCN}(\mathbf{0}, \mathbf{C}, \mathbf{P}), \tag{1.8}$$

when $n \to \infty$.

Remark 1.1.1. *An important remark is that in most signal applications, such as radar processing or source localization, the mean is known. The preprocessing of the data is then applied to remove the mean. Consequently, we assume without loss of generality (w.l.o.g.) that the mean is equal to zero, i.e.* $\boldsymbol{\mu} = \mathbf{0}$. *However, when the mean is unknown, it can be jointly estimated. This analysis falls outside the scope of this thesis.*

Thanks to the definitions given above, we can now present the classical approach to second order statistics estimation.

1.1.3 Classical approach: Gaussian assumption

Classical statistical approaches in signal processing assume that the received data is complex zero-mean CN-distributed.

Sample Covariance Matrix (SCM)

Applying the maximum likelihood criterion to estimate the covariance matrix of CN-distributed data, one obtains the sample covariance matrix (SCM), that is sufficient statistics for in this case.

Definition 1.1.8. *Sample Covariance Matrix (SCM)*
Let $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ be an n-sample of p-dimensional complex i.i.d. vectors with $\mathbf{z}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$. The SCM is given by

$$\widehat{\mathbf{C}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{z}_i \mathbf{z}_i^H. \tag{1.9}$$

When the number of observation is fixed, this estimator has a complex Wishart distribution (scaled by $1/n$) defined as follows.

Definition 1.1.9. *Complex Wishart (CW) distribution*
A matrix $\mathbf{W} = \sum_{i=1}^{n} \mathbf{z}_i \mathbf{z}_i^H$, where $\mathbf{z}_i$ are i.i.d. r.v.'s with $\mathbf{z}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$, is complex Wishart-distributed, denoted as $\mathbf{W} \sim \mathcal{CW}_p(n, \mathbf{C})$. Its p.d.f. is given by

$$f(\mathbf{W}) = \frac{|\mathbf{W}|^{n-p}}{|\mathbf{C}|^n \tilde{\Gamma}_p(n)} \exp\left[-\mathrm{tr}\left(\mathbf{C}^{-1}\mathbf{W}\right)\right].$$

The SCM is also asymptotically GCN-distributed with covariance and pseudo-covariance matrices given in the following theorem.

Theorem 1.1.1. *Asymptotic distribution of SCM*
Let $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ be an n-sample of p-dimensional complex i.i.d. vectors with $\mathbf{z}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$. Then, the asymptotic distribution of the SCM given by Eq. (1.9) is given by

$$\sqrt{n}\mathrm{vec}\left(\widehat{\mathbf{C}} - \mathbf{C}\right) \xrightarrow{d} \mathcal{GCN}\left(\mathbf{0}, \mathbf{C}_{\mathrm{SCM}}, \mathbf{P}_{\mathrm{SCM}}\right), \tag{1.10}$$

where the asymptotic covariance and pseudo-covariance matrices are

$$\begin{cases} \mathbf{C}_{\mathrm{SCM}} = \mathbf{C}^T \otimes \mathbf{C}, \\ \mathbf{P}_{\mathrm{SCM}} = \left(\mathbf{C}^T \otimes \mathbf{C}\right) \mathbf{K}. \end{cases} \tag{1.11}$$

Consequently, the properties of SCM are completely known in the CN context. It is unbiased, consistent and has an asymptotic GCN distribution when the sample size tends to infinity. Moreover, when the sample size is fixed it has a scaled CW distribution. Since it has an explicit form, it is very convenient for theoretical derivations.

Nevertheless, in practice the data is often corrupted, i.e. contain outliers or samples that follow different patterns. In that case, the CN distribution becomes inadequate and more complex models are needed. An alternative arises with elliptical distributions defined in the following.

1.1.4 Non-Gaussian context

Complex Elliptically Symmetric Distributions

A natural generalization of CN distribution is given by the class of Complex Elliptically Symmetric (CES) distributions, that allow heavier and lighter tails than the CN distribution. In addition to the CN distribution, many other well-known and widely used multivariate distributions belong to the CES class, such as t-distribution applied in financial studies, K-distribution used in radar clutter modeling, generalized Gaussian distribution employed in wavelet modeling and image processing, etc. CES distributions are very convenient for signal modeling since they inherit a lot of nice CN properties. An excellent survey that

highlights importance of CES distributions in engineering and signal processing applications is given in [130].

CES distributions are actually the complex extension of real elliptically symmetric (RES) distributions that were introduced by Kelker[87] and studied by different authors [54, 120, 124, 52, 58]. A generalization of RES distribution was proposed by Frahm in order to allow the asymmetry [57]. For a brief introduction of RES distributions discussion, we refer the reader to Appendix B. For more details about RES distributions and their applications please refer to cited literature. In the following, we present the family of CES distributions that are of the main interest for our study.

Definition 1.1.10. *Complex Elliptically Symmetric (CES) distribution*
An r.v. $\mathbf{z}$ is said to have a Complex Elliptically Symmetric (CES) distribution if its characteristic function has the form

$$\Phi(\mathbf{z}) = \exp\{j\mathrm{Re}(\mathbf{z}^H\boldsymbol{\mu})\}\phi(\mathbf{z}^H\boldsymbol{\Sigma}\mathbf{z}), \tag{1.12}$$

where $\phi : \mathbb{R}^+ \to \mathbb{R}$ is the characteristic generator, positive semi-definite Hermitian matrix $\boldsymbol{\Sigma}$ is the scatter matrix and $\boldsymbol{\mu}$ is the mean, or symmetry center. We shall write $\mathbf{z} \sim \mathcal{CES}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ to denote this property.

In the absolutely continuous case, $\mathbf{z}$ possesses a p.d.f. given by

$$f_{\mathbf{z}}(\mathbf{z}) = C_{p,g}|\boldsymbol{\Sigma}|^{-1}\, g\left((\mathbf{z}-\boldsymbol{\mu})^H\, \boldsymbol{\Sigma}^{-1}\,(\mathbf{z}-\boldsymbol{\mu})\right), \tag{1.13}$$

where $g : \mathbb{R}_0^+ \to \mathbb{R}^+$ is the density generator and $C_{p,g}$ a normalizing constant. Then, we can write $\mathbf{z} \sim \mathcal{CES}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$.

Another definition of CES distributions, that is equivalent to Definition 1.1.10, is given by the following stochastic representation [187].

Theorem 1.1.2. *Stochastic representation theorem*
An r.v. $\mathbf{z} \sim \mathcal{CES}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ *with* $rank(\boldsymbol{\Sigma}) = k \leq p$ *if and only if it can be represented as*

$$\mathbf{z} \overset{d}{=} \boldsymbol{\mu} + \sqrt{\mathcal{Q}}\mathbf{A}\mathbf{u}^{(k)}, \tag{1.14}$$

where the non-negative real r.va. $\mathcal{Q}$*, called the modular variate, is independent of the r.v.* $\mathbf{u}^{(k)}$ *that is uniformly distributed on the unit complex k-hypersphere* $\mathcal{CS}^{k-1}$ *with*

$$\mathcal{CS}^{k-1} = \left\{\mathbf{u} \in \mathbb{R}^k : \|\mathbf{u}\|_2 = 1\right\}, \tag{1.15}$$

i.e. $\mathbf{u}^{(k)} \sim \mathcal{U}(\mathcal{CS}^{k-1})$*, and* $\mathbf{A} \in \mathbb{C}^{p\times k}$ *with* $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^H$.

Remark 1.1.2. *One can note that the couple* $(\boldsymbol{\Sigma}, \phi(\cdot))$ *(or equivalently* $(\boldsymbol{\Sigma}, g(\cdot))$ *or* $(\boldsymbol{\Sigma}, \mathcal{Q})$ *does not uniquely identify the p-variate CES distribution. In order to do so, a scale constraint on* $\phi(\cdot))$ *or on* $\boldsymbol{\Sigma}$ *needs to be imposed in a suitable way. For instance, the constraint* $\mathbb{E}[\mathcal{Q}] = rank(\boldsymbol{\Sigma})$ *produces* $\mathbf{C} = \boldsymbol{\Sigma}$.

When the stochastic representation of a CES-distributed r.v. $\mathbf{z}$ has some particular properties, $\mathbf{z}$ may belong to a particular subclass of CES distribution. We will now recall interesting subclasses of CES distributions.

The first one are spherical distributions, that represent centered CES distribution with $\boldsymbol{\Sigma} = \mathbf{I}$, i.e. $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \mathbf{I}, g)$ which gives $\mathbf{z} \overset{d}{=} \sqrt{\mathcal{Q}}\mathbf{u}^{(p)}$. In this case $\mathbf{z}$ is called to be unitary invariant which offers another definition of spherical distributions.

Definition 1.1.11. *Spherical distribution*
A p-dimensional r.v. $\mathbf{z}$ is said to be spherically distributed if and only if $\mathbf{z} \overset{d}{=} \mathbf{U}\mathbf{z}$ for every p-dimensional unitary matrix $\mathbf{U}$.

Another very important subclass of CES-distributions are Complex Compound-Gaussian (CCG) distributions. They are also referred to as Spherically Invariant Random Vectors (SIRV) in the signal processing applications [35, 70]. These distributions are widely used for modeling radar clutter [37, 38, 65, 178, 174]. In fact, with high resolution the necessity to model a CN distribution with spatially variable power appeared. CCG distributions were proposed as an solution that offers an additional parameter called *texture* that is used to model spatial variations of CN-distributed radar speckle.

Definition 1.1.12. *Complex Compound-Gaussian (CCG) distribution*
An r.v. $\mathbf{z}$ is said to have a Complex Compound-Gaussian (CCG) distribution if it admits the following representation

$$\mathbf{z} \overset{d}{=} \boldsymbol{\mu} + \sqrt{\tau}\mathbf{g}, \tag{1.16}$$

where the positive real r.va. τ with c.d.f. F_τ, called the texture, is independent of the Gaussian speckle $\mathbf{g} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. We denote this case as $\mathbf{z} \sim \mathcal{CCG}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, F_\tau)$.

If $\text{rank}(\boldsymbol{\Sigma}) = p$ and if τ possesses a p.d.f., then the p.d.f. of $\mathbf{z}$ is given by

$$f_{\mathbf{z}}(\mathbf{z}) = \pi^{-p}|\boldsymbol{\Sigma}|^{-1} \int_0^\infty \tau^{-p} \exp\left(-\frac{(\mathbf{z}-\boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu})}{\tau}\right) f_\tau d\tau, \tag{1.17}$$

where $f_\tau(\tau) = F'_\tau(\tau)$ is the density of τ.

Another important distribution class are complex angular Gaussian distributions that represent the distribution of a projection of $\mathbf{z} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ onto the unit complex p-hypersphere.

Definition 1.1.13. *Complex Angular Central Gaussian (CACG) distribution*
An r.v. $\mathbf{z}_a$ is said to have a Complex Angular Central Gaussian (CACG) distribution if it admits a stochastic representation $\mathbf{z}_a = \mathbf{z}/\|\mathbf{z}\|$, where $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$. This case will be denoted $\mathbf{z}_a \sim \mathcal{CACG}(\mathbf{0}, \boldsymbol{\Sigma})$.

For nonsingular $\boldsymbol{\Sigma}$, the p.d.f. of the distribution is given by

$$f_{\mathbf{z}_a}(\mathbf{z}) = C_p|\boldsymbol{\Sigma}|^{-1}(\mathbf{z}^H\boldsymbol{\Sigma}^{-1}\mathbf{z})^{-p}, \tag{1.18}$$

where C_p is the surface area of $\mathcal{CS}^{p-1}$.

One can note that if the CN distribution is replaced by any central CES distribution the resulting ACG distribution is the same. Thus, the more appropriate name would be complex angular elliptical distribution. However, we will keep the used CACG notation. The important remark is that although the p.d.f. of CACG distribution has the form of a CES p.d.f., these distributions do not belong to the class of CES distributions since they can not be represented by the stochastic representation theorem.

In the following we list some examples of CES distributions and discuss their properties.

Examples of CES distributions

Example 1.1.1. *Complex Normal (CN) distribution*
Complex Normal (or Gaussian) distribution is a particular case of CES distributions for which $g(x) = \exp(-x)$ *and* $C_{p,g} = \pi^{-p}$. *The p.d.f. of* $\mathbf{z} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *can also be obtained from Eq. (1.6) for* $\mathbf{P} = \mathbf{0}$. *It admits the corresponding stochastic representation given by Eq. (1.14) for* $\mathcal{Q} \sim \frac{1}{2}\chi^2_{2p}$.

Example 1.1.2. *Complex t-distribution*
The complex multivariate t-distribution with $\nu > 0$ *degrees of freedom (DoF), denoted as* $\mathcal{C}t_\nu(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, *is obtained for*

$$g(x) = (1 + 2x/\nu)^{-(2p+\nu)/2}, \tag{1.19}$$

yielding $C_{p,g} = 2^p\Gamma(\frac{2p+\nu}{2})/[(\pi\nu)^p\Gamma(\frac{\nu}{2})]$ *as the normalizing constant. It admits the corresponding stochastic representation given by Eq.* (1.14) *for* $\mathcal{Q} \sim pF_{2p,\nu}$. *This distribution also belongs to the CCG distributions where* $\tau \sim IG(\nu/2, \nu/2)$, *with IG denoting the inverse Gamma distribution. The distribution has finite 2nd-order moments for* $\nu > 2$. *Note that the case* $\nu \to \infty$ *leads to the CN distribution. The multivariate t-distributions, besides the CN distribution, encompass also multivariate Laplace distribution (for* $\nu = 1/2$*) and the multivariate Cauchy distribution (for* $\nu = 1$*) which are heavy-tailed alternatives to the CN distribution.*

Example 1.1.3. *Complex K-distribution*
The complex K-distribution with shape $\nu > 0$, $\mathcal{CK}_\nu(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, *is obtained for the density generator*

$$g(x) = x^{(\nu-p)/2}K_{\nu-p}(2\sqrt{\nu x}) \tag{1.20}$$

and the normalizing constant $C_{p,g} = 2\nu^{(\nu+p)/2}/(\pi^p\Gamma(p))$, *where* $K_i(\cdot)$ *denotes the modified Bessel function of the second kind of order i. Note that there is no closed form for the p.d.f. of* $\mathcal{Q}$ *in this case. However, it admits the CCG-representation where* $\tau \sim Gam(\nu, 1/\nu)$. *When* $\nu \to \infty$, *the K-distribution yields the CN distribution.*

Example 1.1.4. *Complex W-distribution*
The complex W-distribution with exponent $s > 0$ *and scale* $b > 0$, *denoted as* $\mathcal{CW}_{s,b}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, *is obtained for the density generator of the form*

$$g(x) = x^{s-1}\exp(-x^s/b) \tag{1.21}$$

and the normalizing constant $C_{p,g} = s\Gamma(p)b^{-(s+p-1)/s}/[\pi^p\Gamma(\frac{s+p-1}{s})]$. *It admits the stochastic representation given by Eq.* (1.14) *for* $\mathcal{Q} \overset{d}{=} \mathcal{G}^{1/s}$, *where* $\mathcal{G} \sim Gam(\frac{p+s-1}{s}, b)$. *For proper identifiability of the couple* $(\boldsymbol{\Sigma}, g(\cdot))$ *it is set to* $b = \left[p\Gamma(\frac{p+s-1}{s})/\Gamma(\frac{p+s}{s})\right]^s$ *which yields to* $\mathbf{C} = \boldsymbol{\Sigma}$. *For* $s = 1$ *the CN distribution is obtained. Finally, it should be noted that W-distribution does not belong to the class of CCG distributions.*

Example 1.1.5. *Complex Generalized Gaussian distribution*
The complex generalized Gaussian (CGG) distribution with exponent $s > 0$ *and scale* $b > 0$, *denoted as* $\mathcal{CGG}_{s,b}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, *is obtained for*

$$g(x) = \exp(-x^s/b) \tag{1.22}$$

and the normalizing constant $C_{p,g} = s\Gamma(p)b^{-p/s}/[\pi^p\Gamma(\frac{p}{s})]$. *It admits the stochastic representation given by Eq.* (1.14) *for* $\mathcal{Q} \overset{d}{=} \mathcal{G}^{1/s}$, *where* $\mathcal{G} \sim Gam(\frac{p}{s}, b)$. *For proper identifiability of the couple* $(\boldsymbol{\Sigma}, g(\cdot))$ *it is set to* $b = \left[p\Gamma(\frac{p}{s})/\Gamma(\frac{p+1}{s})\right]^s$ *which yields to* $\mathbf{C} = \boldsymbol{\Sigma}$. *For* $s = 1$ *the CN distribution is obtained, while* $s = 1/2$ *yields the complex Laplace distribution. When* $s < 1$ *distributions with heavier tails are produced.*

All previously itemized examples of CES distributions are recapitulated in Table 1.1.

Maximum Likelihood Estimators

In the context of CES-distributed data, the SCM does not give good estimation of the scatter matrix. Assuming a particular data distribution, one can obtain the optimal scatter matrix estimator maximizing the corresponding likelihood function.

A Maximum Likelihood Estimator (MLE) is obtained maximizing the likelihood function defined as

$$p\left((\mathbf{z}_1, \ldots, \mathbf{z}_n)\,|\boldsymbol{\Sigma}\right) = \prod_{i=1}^{n} p\left(\mathbf{z}_i|\boldsymbol{\Sigma}\right) \tag{1.23}$$

$\mathcal{CES}$	$C_{p,g}$	$g(x)$	$f(\mathcal{Q})$	$f(\tau)$
$\mathcal{CN}$	π^{-p}	$\exp(-x)$	$1/2\chi^2_{2p}$	$\delta_a(\tau), a \geq 0$
$\mathcal{C}t_\nu$	$\frac{2^p\Gamma((2p+\nu)/2)}{(\pi\nu)^p\Gamma(\nu/2)}$	$(1+2x/\nu)^{-(2p+\nu)/2}$	$pF_{2p,\nu}$	$\text{IG}\left(\frac{\nu}{2},\frac{\nu}{2}\right)$
$\mathcal{CK}_\nu$	$\frac{2\nu^{(\nu+p)/2}}{\pi^p\Gamma(p)}$	$x^{(\nu-p)/2}K_{\nu-p}(2\sqrt{\nu x})$	no closed form	$\text{Gam}\left(\nu,\frac{1}{\nu}\right)$
$\mathcal{CW}_{s,b}$	$\frac{s\Gamma(p)b^{-(s+p-1)/s}}{\pi^p\Gamma((s+p-1)/s)}$	$x^{s-1}\exp(-x^s/2)$	$\text{Gam}^{1/s}\left(\frac{p+s-1}{s},b\right)$	$\times$
$\mathcal{CGG}_{s,b}$	$\frac{s\Gamma(p)b^{-p/s}}{\pi^p\Gamma(p/s)}$	$\exp(-x^s/2)$	$\text{Gam}^{1/s}\left(\frac{p}{s},b\right)$	$\times$

Table 1.1 – Examples of CES distributions; $\times$ means that the CES distribution does not belong to the class of CCG distributions

or equivalently, the log-likelihood function

$$\mathcal{L}\left(\boldsymbol{\Sigma}|\mathbf{z}\right) = \ln\left(\left(p\left(\mathbf{z}_1,\ldots,\mathbf{z}_n\right)|\boldsymbol{\Sigma}\right)\right) = \sum_{i=1}^{n}\ln\left(p\left(\mathbf{z}_i|\boldsymbol{\Sigma}\right)\right). \tag{1.24}$$

Supposing that we have an i.i.d. $(\mathbf{z}_1,\ldots,\mathbf{z}_n)$ sample where $\mathbf{z}_i \sim \mathcal{CES}(\mathbf{0},\boldsymbol{\Sigma},g)$ the MLE of the scatter matrix $\boldsymbol{\Sigma}$ is the matrix that maximizes the following quantity

$$\mathcal{L}\left(\boldsymbol{\Sigma}|\mathbf{z}\right) = \sum_{i=1}^{n}\ln\left(f_{\mathbf{z}}\left(\mathbf{z}_i\right)\right) = n\ln\left(C_{p,g}\right) - n\ln\left(|\boldsymbol{\Sigma}|\right) + \sum_{i=1}^{n}\ln\left(g\left(\mathbf{z}_i^H\boldsymbol{\Sigma}^{-1}\mathbf{z}\right)\right). \tag{1.25}$$

Verifying that the derivative with respect to (w.r.t.) $\boldsymbol{\Sigma}$ is equal to zero

$$-n\frac{\partial\ln|\boldsymbol{\Sigma}|}{\partial\boldsymbol{\Sigma}} + \sum_{i=1}^{n}\frac{1}{g(x_i)}\left(\frac{\partial g(x_i)}{\partial x_i}\right)\left(\frac{\partial x_i}{\partial\boldsymbol{\Sigma}}\right) = \mathbf{0} \tag{1.26}$$

one obtains the MLE

$$\widehat{\boldsymbol{\Sigma}}_{\text{ML}} = \sum_{i=1}^{n}\frac{-g'(x_i)}{g(x_i)}\mathbf{z}_i\mathbf{z}_i^H, \tag{1.27}$$

where $x_i = \mathbf{z}_i^H\boldsymbol{\Sigma}^{-1}\mathbf{z}$. The weight function $u(x) = -g'(x)/g(x)$ depends on the density generator $g(\cdot)$ of the underlying CES distribution.

Robust M-estimators

MLEs perform very well when the data follow an assumed model. However, the data distribution is sometimes only approximately described by the proposed assumption. Choosing an appropriate estimator when we do not know the underlying statistical model is even more challenging problem.

This question gave rise to a particular statistics branch called robust theory. Although the first works in this domain date from the end of nineteenth century, the great expansion happened in sixties and seventies of the last century with fundamental work of Tukey [166, 167], Huber [82, 83] and Hampel [80, 81]. The theoretical approach to robust statistics introduced by Huber supposes that the true data distribution is situated in a neighbourhood of a stochastic model. In this spirit, robust estimators represent estimators that are

insensitive to small deviations from the assumed model (usually the CN distribution) and stay optimal under this setup.

In this section we present complex M-estimators of scatter, where M stands for *maximum likelihood-type*. They were first introduced in real case by Maronna [119]. The complex extension was introduced and analyzed in [130, 117].

Definition 1.1.14. *M-estimator*
Let $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ be an n-sample of p-dimensional complex i.i.d. vectors with $\mathbf{z}_i \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}})$. An M-estimator, denoted by $\widehat{\mathbf{\Sigma}}$, is defined by the solution of the following M-estimating equation

$$\widehat{\mathbf{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n} u(\mathbf{z}_i^H \widehat{\mathbf{\Sigma}}^{-1}\mathbf{z}_i)\mathbf{z}_i\mathbf{z}_i^H, \tag{1.28}$$

where u is any real-valued weight function on $[0, \infty)$ that respects Maronna's conditions (introduced by Maronna for real case), ensuring existence and uniqueness of Eq. (1.28). The conditions have been extended to the complex case by Ollila in [130] and are listed here:

- u is non-negative, non-increasing and continuous on $[0, \infty)$.
- Let $\Psi(x) = xu(x)$ and $K = \sup_{x\geq 0}\Psi(x)$. $p < K < \infty$, Ψ is non-decreasing, and strictly increasing on the interval where $\Psi(x) < K$.
- Let $P_{2n}(\cdot)$ denote the empirical distribution of $(\mathbf{v}_1, \ldots, \mathbf{v}_{2n})$ where $\mathbf{v}_i = \left(\Re(\mathbf{z}_i)^T, \Im(\mathbf{z}_i)^T\right)^T$ and $\mathbf{v}_{n+i} = \left(-\Im(\mathbf{z}_i)^T, \Re(\mathbf{z}_i)^T\right)^T$. Then for all linear subspaces $V \in \mathbb{R}^{2p}$, with $\dim(V) = 2p-1, P_{2n} < 1 - (2p-1)/2K$.

It is important to note that the weight function is not necessary related to a p.d.f. of any particular CES distribution or to any p.d.f. Hence, M-estimators constitute a wide class of scatter matrix estimators that includes the MLEs. Thus, M-estimators can be considered as a natural generalization of MLEs which are not necessary robust.

We denote $\mathbf{\Sigma}_\sigma$ the theoretical scatter matrix M-functional, which is defined as a solution of

$$\mathbb{E}\left[u\left(\mathbf{z}^H\mathbf{\Sigma}_\sigma^{-1}\mathbf{z}\right)\mathbf{z}\mathbf{z}^H\right] = \mathbf{\Sigma}_\sigma. \tag{1.29}$$

The M-functional is proportional to the true scatter matrix $\mathbf{\Sigma}$ as

$$\mathbf{\Sigma}_\sigma = \sigma^{-1}\mathbf{\Sigma}, \tag{1.30}$$

where the scalar factor $\sigma > 0$ can be found by solving

$$\mathbb{E}[\Psi(\sigma\mathcal{Q})] = p \tag{1.31}$$

with $\mathcal{Q} \overset{d}{=} \mathbf{z}^H\widehat{\mathbf{\Sigma}}^{-1}\mathbf{z}$.

In [130, 117] it was shown that an M-estimator has the asymptotic GCN distribution centered around the M-functional given by Eq. (1.30).

Theorem 1.1.3. *Asymptotic distribution of M-estimators [117, 130]*
Let $\widehat{\mathbf{\Sigma}}$ be an M-estimator as in Eq. (1.28) built from n samples drawn as $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}})$. The asymptotic distribution of $\widehat{\mathbf{\Sigma}}$ is given by as

$$\sqrt{n}\mathrm{vec}\left(\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}_\sigma\right) \overset{d}{\to} \mathcal{GCN}(\mathbf{0}, \mathbf{C}_M, \mathbf{P}_M),$$

where the asymptotic covariance and pseudo-covariance matrices are

$$\begin{cases} \mathbf{C}_M = \vartheta_1\mathbf{\Sigma}_\sigma^T \otimes \mathbf{\Sigma}_\sigma + \vartheta_2\mathrm{vec}(\mathbf{\Sigma}_\sigma)\mathrm{vec}(\mathbf{\Sigma}_\sigma)^H, \\ \mathbf{P}_M = \vartheta_1\left(\mathbf{\Sigma}_\sigma^T \otimes \mathbf{\Sigma}_\sigma\right)\mathbf{K} + \vartheta_2\mathrm{vec}(\mathbf{\Sigma}_\sigma)\mathrm{vec}(\mathbf{\Sigma}_\sigma)^T. \end{cases} \tag{1.32}$$

The constants $\vartheta_1 > 0$ and $\vartheta_2 > -\vartheta_1/p$ are given by

$$\begin{aligned}\vartheta_1 &= c_M^{-2} a_M p(p+1),\\ \vartheta_2 &= (c_M - p^2)^{-2}(a_M - p^2) - c_M^{-2} a_M (p+1),\end{aligned} \tag{1.33}$$

where

$$\begin{aligned}a_M &= \mathbb{E}\left[\Psi^2(\sigma\mathcal{Q})\right],\\ c_M &= \mathbb{E}\left[\Psi'(\sigma\mathcal{Q})\sigma\mathcal{Q}\right] + p^2.\end{aligned} \tag{1.34}$$

Remark 1.1.3. *Note that for the SCM, under the CN assumption, $\sigma = 1$, $\vartheta_1 = 1$ and $\vartheta_2 = 0$. However, the SCM can not be classified as an M-estimator since the function $u(x) = 1$ does not satisfy Maronna's conditions.*

Examples of M-estimators

Hereafter, we present some of widely used M-estimators both in practical and theoretical analysis.

Example 1.1.6. *Tyler's M-estimator [170, 135, 137]*
Tyler's M-estimator is given as the solution of the following equation

$$\widehat{\mathbf{\Sigma}}_{\mathrm{T}} = \frac{p}{n}\sum_{i=1}^{n}\frac{\mathbf{z}_i\mathbf{z}_i^H}{\mathbf{z}_i^H\widehat{\mathbf{\Sigma}}_{\mathrm{T}}^{-1}\mathbf{z}_i} \tag{1.35}$$

or equivalently as the solution of Eq. (1.28) for $u(x) = p/x$. Nevertheless, the Tyler's M-estimator is not exactly an M-estimator since it does not satisfy the second Maronna's condition. However, it is very useful because of rare property that any CES distribution with the same scatter matrix leads to the same result (hence "distribution-free"). Also, in order to obtained a unique solution of Eq. (1.35) one needs to normalize $\widehat{\mathbf{\Sigma}}_{\mathrm{T}}$ in a suitable way, e.g. $\mathrm{Tr}(\widehat{\mathbf{\Sigma}}_{\mathrm{T}}) = p$ which gives the estimator of the shape matrix. Note that another common normalization for the shape matrix estimator is to set $|\widehat{\mathbf{\Sigma}}_{\mathrm{T}}| = 1$, but we will keep the first one throughout this thesis.

Consequently, the complex Tyler's M-estimator is an MLE under the assumption that $\mathbf{z}_i \sim \mathcal{CN}(\mathbf{0}, \tau_i\mathbf{V})$, $i = 1, \ldots, n$, which means that each sample has the same shape matrix $\mathbf{V}$, but different power (scale) [34, 64]. Moreover, this result is valid, not only for the CN, but for any CES distribution [128]. Interestingly, the complex Tyler's M-estimator is also the MLE of the scatter matrix of the CACG distribution [93].

The asymptotic distribution of this estimator can not be directly obtained from Theorem 1.1.3. However, it was analyzed in [137] where the following properties were derived.

Theorem 1.1.4. *Asymptotic distribution of the Tyler's M-estimator*
Let $\widehat{\mathbf{\Sigma}}_{\mathrm{T}}$ be defined by Eq. (1.35). The asymptotic distribution of $\mathbf{\Sigma}_{\mathrm{T}}$ is given by

$$\sqrt{n}\mathrm{vec}\left(\widehat{\mathbf{\Sigma}}_{\mathrm{T}} - \mathbf{\Sigma}\right) \xrightarrow{d} \mathcal{CN}\left(\mathbf{0}, \mathbf{C}_{\mathrm{T}}, \mathbf{P}_{\mathrm{T}}\right),$$

where $\mathbf{C}_{\mathrm{T}}$ and $\mathbf{P}_{\mathrm{T}}$ are defined by

$$\begin{aligned}\mathbf{C}_{\mathrm{T}} &= \frac{p+1}{p}\mathbf{\Sigma}^T \otimes \mathbf{\Sigma} - \frac{p+1}{p^2}\mathrm{vec}(\mathbf{\Sigma})\mathrm{vec}(\mathbf{\Sigma})^H,\\ \mathbf{P}_{\mathrm{T}} &= \frac{p+1}{p}\left(\mathbf{\Sigma}^T \otimes \mathbf{\Sigma}\right)\mathbf{K} - \frac{p+1}{p^2}\mathrm{vec}(\mathbf{\Sigma})\mathrm{vec}(\mathbf{\Sigma})^T.\end{aligned} \tag{1.36}$$

Remark 1.1.4. *Note that this result has been obtained for the theoretical normalization $\mathrm{Tr}\left(\mathbf{\Sigma}^{-1}\widehat{\mathbf{\Sigma}}_{\mathrm{T}}\right) = p$ that is not possible in practice since $\mathbf{\Sigma}$ is unknown. This normalization is equivalent to $\mathrm{Tr}(\widehat{\mathbf{\Sigma}}_{\mathrm{T}}) = p$ only when $\mathbf{\Sigma} = \mathbf{I}$. However, when the normalization $\mathrm{Tr}(\widehat{\mathbf{\Sigma}}_{\mathrm{T}}) = p$ is employed one has $\widehat{\mathbf{\Sigma}}_{\mathrm{T}} \to p\mathbf{\Sigma}/\mathrm{Tr}(\mathbf{\Sigma})$. The asymptotic distribution in this case can be found in [130].*

Example 1.1.7. *Huber's M-estimator*
The complex extension of Huber's M-estimator is defined by

$$\widehat{\mathbf{\Sigma}}_{\mathrm{H}} = \frac{1}{n\beta}\sum_{i=1}^{n}\left[\mathbf{z}_i\mathbf{z}_i^H \mathbb{1}_{\mathbf{z}_i^H\widehat{\mathbf{\Sigma}}_{\mathrm{H}}^{-1}\mathbf{z}_i\leq\lambda}\right] + \frac{\lambda}{n\beta}\sum_{i=1}^{n}\left[\frac{\mathbf{z}_i\mathbf{z}_i^H}{\mathbf{z}_i^H\widehat{\mathbf{\Sigma}}_{\mathrm{H}}^{-1}\mathbf{z}_i}\mathbb{1}_{\mathbf{z}_i^H\widehat{\mathbf{\Sigma}}_{\mathrm{H}}^{-1}\mathbf{z}_i>\lambda}\right], \tag{1.37}$$

where $\mathbb{1}$ is the indicator function and λ and β depend on a single parameter $0<q<1$, according to

$$\begin{aligned} q &= F_{2p}(2\lambda), \\ \beta &= F_{2p+2}(2\lambda) + \lambda\frac{1-q}{p}, \end{aligned} \tag{1.38}$$

where $F_p(\cdot)$ is the cumulative distribution function of a χ^2 distribution with p DoF. The parameters are chosen so that the resulting M-estimator is consistent for the covariance matrix at the CN distribution. The parameter q represents the percentage of the data to be treated as homogeneous, i.e. CN-distributed, while $1-q$ percentage of the data are treated as outliers. Indeed, one can note that $q=1$ yields the SCM, whereas $q=0$ yields the Tyler's M-estimator. Hence, the Huber's M-estimator can be interpreted as a random weighted combination between the SCM and the Tyler's estimator. In [116] it was shown that the asymptotic distribution of $\widehat{\mathbf{\Sigma}}_{\mathrm{H}}$ is given by Theorem 1.1.3 for

$$\begin{cases} a_M = \left(p(p+1)F_{2p+4}(2\lambda) + \lambda^2(1-q)\right)/\beta^2, \\ c_M = pF_{2p+2}(2\lambda)/\beta + p^2. \end{cases} \tag{1.39}$$

Example 1.1.8. *Student's M-estimator*
Student's M-estimator is the MLE for the Student's t-distribution. It is given as the solution of the following equation

$$\widehat{\mathbf{\Sigma}}_t = \frac{p+\nu/2}{n}\sum_{i=1}^{n}\frac{\mathbf{z}_i\mathbf{z}_i^H}{\mathbf{z}_i^H\widehat{\mathbf{\Sigma}}_t^{-1}\mathbf{z}_i+\nu/2}, \tag{1.40}$$

where $d>0$ is the DoF parameter, i.e. $u(x)=(p+\nu/2)/(x+\nu/2)$. The motivation to analyze this estimator arises from the fact that it presents a trade-off between the SCM and Tyler's estimator, but in a different way than the Huber's M-estimator. Indeed, when $\nu\to\infty$ the Student's t-distribution leads the Gaussian distribution and the Student's M-estimator tends to the SCM ($u(x)\to 1$). On the other hand, for $\nu=0$ Student's M-estimator is equivalent to the Tyler's one. Finally, $\widehat{\mathbf{\Sigma}}_t$ is widely used both in theory (as a benchmark) and in practice which presents strong motivation for understanding its behavior. Note also that as others M-estimators, it is not always used as a MLE for the t-distribution.

Example 1.1.9. *K M-estimator*
K M-estimator is the MLE for the K-distribution with shape ν. It is given as the solution of the following equation

$$\widehat{\mathbf{\Sigma}}_K = \frac{\sqrt{\nu}}{n}\sum_{i=1}^{n}\frac{K_{\nu-p-1}\left(2\sqrt{\nu\mathbf{z}_i^H\widehat{\mathbf{\Sigma}}_K^{-1}\mathbf{z}_i}\right)}{K_{\nu-p}\left(2\sqrt{\nu\mathbf{z}_i^H\widehat{\mathbf{\Sigma}}_K^{-1}\mathbf{z}_i}\right)}\frac{\mathbf{z}_i\mathbf{z}_i^H}{\sqrt{\mathbf{z}_i^H\widehat{\mathbf{\Sigma}}_K^{-1}\mathbf{z}_i}}, \tag{1.41}$$

where $u_K(x)=\frac{\sqrt{\nu}K_{\nu-p-1}\left(2\sqrt{\nu x}\right)}{xK_{\nu-p}\left(2\sqrt{\nu x}\right)}$ satisfies the Maronna's conditions for $\nu>0$.

Example 1.1.10. *Weibull M-estimator*
Weibull M-estimator is the MLE for the complex W-distribution with parameters s and b.

M-estimator	$u(x)$	$\Psi(x)$	MLE
Tyler's M-estimator	$\frac{p}{x}$	p	$\mathcal{CN}(\mathbf{0}, \tau_i \mathbf{V})$, $\mathcal{ACG}(\mathbf{0}, \mathbf{\Sigma})$
Huber's M-estimator	$\frac{1}{\beta}\min\left(1, \frac{\lambda}{x}\right)$	$\frac{1}{\beta}\min(x, \lambda)$	$\times$
Student's M-estimator	$\frac{2p+\nu}{2x+\nu}$	$\frac{2p+\nu}{2x+\nu}x$	$\mathcal{C}t_\nu(\mathbf{0}, \mathbf{\Sigma})$
K M-estimator	$\sqrt{\frac{\nu}{x}}\frac{K_{\nu-p-1}(2\sqrt{\nu x})}{K_{\nu-p}(2\sqrt{\nu x})}$	$\frac{\sqrt{\nu x}K_{\nu-p-1}(2\sqrt{\nu x})}{K_{\nu-p}(2\sqrt{\nu x})}$	$\mathcal{CK}_\nu(\mathbf{0}, \mathbf{\Sigma})$
Weibull M-estimator	$\frac{sx^s/b - s + 1}{x}$	$sx^s/b - s + 1$	$\mathcal{CW}_{s,b}(\mathbf{0}, \mathbf{\Sigma})$
GG M-estimator	$\frac{s}{b}x^{s-1}$	$\frac{s}{b}x^s$	$\mathcal{CGG}_{s,b}(\mathbf{0}, \mathbf{\Sigma})$

Table 1.2 – Examples of M-estimators; $\times$ means that the M-estimator can not be obtained as MLE of any distribution

It is given as the solution of the following equation

$$\widehat{\mathbf{\Sigma}}_W = \frac{1}{n}\sum_{i=1}^{n} \frac{s/b\left(\mathbf{z}_i^H \widehat{\mathbf{\Sigma}}_W^{-1}\mathbf{z}_i\right)^s - s + 1}{\mathbf{z}_i^H \widehat{\mathbf{\Sigma}}_W^{-1}\mathbf{z}_i}\mathbf{z}_i\mathbf{z}_i^H, \tag{1.42}$$

where $\widehat{\mathbf{\Sigma}}_W$ satisfies the Maronna's conditions for $0 < s < 1$. When $s = 1$ and $b = 1$ Weibull M-estimator yields the SCM as expected since the W-distribution for these parameters is equivalent to the CN one. For $s = 0$ Weibull M-estimator is equivalent to the Tyler's one. Consequently, this estimator can also be presented as a trade-off between the SCM and Tyler's M-estimator.

Example 1.1.11. *GG M-estimator [134]*
GG M-estimator is the MLE for the CGG distribution with parameters s and b. It is given as the solution of the following equation

$$\widehat{\mathbf{\Sigma}}_{\mathrm{GG}} = \frac{s/b}{n}\sum_{i=1}^{n} \frac{\mathbf{z}_i\mathbf{z}_i^H}{\left(\mathbf{z}_i^H \widehat{\mathbf{\Sigma}}_{\mathrm{GG}}^{-1}\mathbf{z}_i\right)^{1-s}}, \tag{1.43}$$

where $\widehat{\mathbf{\Sigma}}_{\mathrm{GG}}$ satisfies the Maronna's conditions for $0 < s < 1$. When $s = 1$ and $b = 1$ one can note that GG M-estimator yields the SCM as expected since the CCG distribution for these parameters is equivalent to the CN one. For smaller values of s the CGG distribution has heavier tails and the corresponding MLE becomes more robust.

Table 1.2 details the weight functions $u(x)$ and objective functions $\Psi(x)$ for the itemized special cases of M-estimators. Moreover, the last column indicates if the M-estimator can be obtained as an MLE.

Figure 1.1 plots the corresponding weight function for each M-estimator from Table 1.2. The data dimension p is set to 10.

First, the weight function for Tyler's M-estimator is plotted in Figure 1.1a. Figure 1.1b depicts the weight function for Huber's M-estimator for different values of the parameter q.

One can note that for small values of q, the Huber's weight function is close to the Tyler's one, while for high values of q it tends to the SCM weight function, i.e., $u(x) \to 1$.

Figure 1.1c displays the weight function for the Student's M-estimator for various values of the parameter ν. The figure reveals that for small values of ν the weights tend to the Tyler's ones, while when $\nu \to \infty$, $u(x)$ tends to 1. The same conclusions can be drawn for the weight function of K M-estimator plotted in Figure 1.1d.

Finally, the weight functions for Weibull and GG M-estimators are plotted in Figures 1.1e and 1.1f, respectively. From both figures one can note that for low values of the parameter s, the weights are close to the ones of the Tyler's M-estimator, while when the parameter s increases, the weights tend to the ones of the SCM.

From the figures, we can conclude that each of the studied M-estimators can be considered as a particular trade-off between the Tyler's M-estimator and the SCM.

1.1.5 Conclusion

In this section we have the presented general framework of covariance matrix estimation. When the data is CN-distributed the well-known SCM is used. So far, its behaviour has been broadly studied offering nice theoretical results for computing different parameters in a huge range of applications. The main drawback of this estimator is that it gives a poor covariance matrix estimation even when the data is corrupted with a negligible amount of outliers. In that case, an alternative is to use M-estimators that are robust to the data model deviations. However, these estimators are given by fixed-point equations and consequently, it is quite difficult to directly analyze their statistical properties.

With this in mind, in Chapter 2, we propose a new approach to analyze the behaviour of M-estimators. The originality of the results comes from a new CES representation introducing *Gaussian cores* of the data. This representation allows us to define a *Gaussian Core Wishart Equivalent (GCWE)* of an M-estimator, which corresponds to the SCM built with CN-distributed data. The main contribution lies in the correlation analysis between an M-estimator and the corresponding GCWE establishing a sort of a distance between them, which finally offers new insights into the statistical properties of M-estimators.

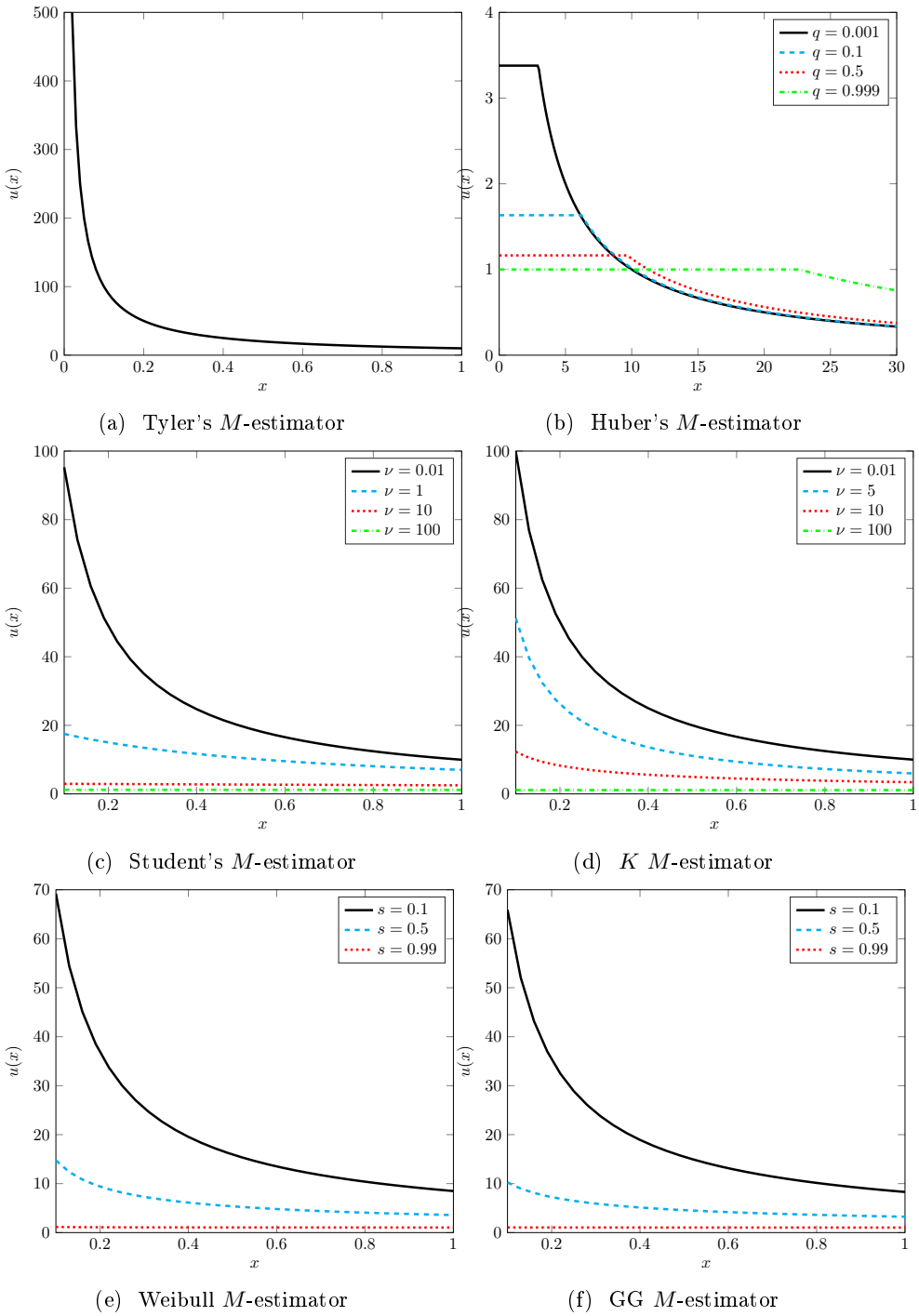
In Chapter 3, the results are extended to the eigenvalues, eigenvectors and principal subspace of M-estimators. Different applications are considered in order to highlight the importance of the theoretical results.

Generally speaking, the main goal of the analysis done in this thesis is to show that it is always better to use robust M-estimators instead of the traditional SCM, since they offer high gain in term of scatter matrix estimation in a non-Gaussian framework. On the other hand, we will see that the loss in term of statistical properties when compared to the SCM in the Gaussian framework is mostly insignificant.

1.2 Signal detection

1.2.1 Introduction

Adaptive detection of signals embedded in Gaussian disturbance is a ubiquitous problem in statistical signal processing [157]. There is a considerable amount of works dealing with the design and performance analysis of various detectors adapted to specific detection problems. In general, the problem consists in deciding whether or not the collected target responses, called primary data, contain useful signals. The target response can have different forms thanks to specific scenarios. For instance, depending on the range resolution and target size, the response can be contained in one range cell or spread over multiples cells [173]. The sizes of the collected data depend on the number of sensors and transmitted pulse signals, resulting in space-time or only temporal/spatial detection. The direction or steering vector depends on the particular set-up and this parameter is usually assumed to



(a) Tyler's M-estimator (b) Huber's M-estimator

(c) Student's M-estimator (d) K M-estimator

(e) Weibull M-estimator (f) GG M-estimator

Figure 1.1 – Weight functions for M-estimators. From left to right and from top to bottom: Tyler's M-estimator, Huber's M-estimator, Student's M-estimator, K M-estimator, Weibull M-estimator, and GG M-estimator, $p = 10$.

be perfectly known. However, when the true direction is different from the nominal one (mismatch scenario), an alternative is to model the target as a linear combination of known basis vectors. This case is known as subspace detection [157, 97, 99]. Moreover, various

assumptions on the complex amplitude of the primary data can be made. It can be modeled as an unknown deterministic parameter [90, 92] or random with unknown deterministic or random module [8].

Various detection statistics adapted to model assumptions have been proposed. These statistics depend on disturbance parameters that are unknown and need to be estimated. Two main approaches have been employed depending on the assumed nature of the disturbance parameters. The first one models these parameters as deterministic quantities. The second one, *Bayesian* approach [43, 44, 10], models these quantities as unknown random values, and thus provides detectors that depend on a chosen a priori distribution for the unknown parameters.

Among the disturbance parameters, the covariance matrix is an essential quantity used for the whitening transformation during the detection process. Moreover, when the covariance matrix is unknown the performance of the adaptive detectors strongly relies on its estimation accuracy. Under the Gaussian assumption, the classical widely used covariance estimator is the well-known SCM built out of some signal-free samples, called secondary data. This estimator permits to obtain nice detection performances and to derive the statistical properties of the detection statistics. However, it has been shown that the required sample size n of secondary data that provides correct performances of detecting statistics is $n \approx 2p$ [149]. Since for large values of p such a number of secondary data is not available, one needs to use detection techniques for reduced value of n. The methods such as autoregressive models [151, 133] or random matrix theory techniques [86, 40, 176] can be used. Another case when the number n can be significantly reduced is when the disturbance is composed of a Low-Rank (LR) clutter and an Additive White Gaussian Noise (AWGN) [95, 148, 69, 79]. In this situation, instead of the SCM, the projector orthogonal onto the clutter subspace constructed with the SCM eigenvectors is usually used in order to remove the clutter from the data [69]. Some other techniques can be found in [146, 18].

In the following, we recall the full and LR detection problems in homogeneous Gaussian disturbance together with proposed detectors and their properties.

1.2.2 Problem formulation

The considered signal detection problem is equivalent to the binary hypothesis testing. The issue is to determine if the useful signal is present in a received measurement. Under the hypothesis H_0 the signal contains only the disturbance. Under the hypothesis H_1, in addition to this disturbance, the signal contains the target to be detected. The objective of a detection process is to determine which of these two assumptions is most likely minimizing the probabilities of two types of possible errors:

- Type I error - error of the first kind presenting the incorrect rejection of a true null hypothesis, called *false alarm*,
- Type II error - error of the second kind presenting the failure to reject a false null hypothesis, called *non-detection*.

According to Neyman–Pearson criterion, one aims at maximizing the detection probability (DP), under the constraint that the probability of false alarm (PFA) is not greater than a pre-specified value. Thus, in order to detect the signal, the value of a detection statistic is compared to a pre-computed threshold value that is obtained for a given PFA.

This problem has been extensively studied in the context of Gaussian distributed disturbance [89, 90, 91, 92, 88]. Several decision statistics have been proposed, such as the Generalized Likelihood Ratio Test (GLRT) (Kelly's detector) [90], the Adaptive Matched Filter (AMF or 2-step GLRT) [150], its normalized counterpart (Adaptive Normalized Matched Filter-ANMF or Adaptive Cosine Estimator-ACE) [39, 97], and the Rao test [118]. The

associated detectors have been characterized in terms of DP and PFA, as well as constant false alarm rate (CFAR) properties (see e.g., [132] and references therein) and performance in mismatched scenarios [68]. In order to detect a signal these detectors require to know the second-order characteristics of signal-free data. When the data covariance matrix is unknown, these statistics involve the SCM, built from the CN-distributed noise, as a core component.

1.2.3 Full rank detection

We consider the problem of detecting a known complex signal vector $\mathbf{p}$, called steering vector, from the received data $\mathbf{z} = \alpha\mathbf{p} + \mathbf{c}$, called primary data, where $\mathbf{c}$ is the unobserved complex noise random vector and α is a complex amplitude modeled as an unknown deterministic parameter. The problem of detecting the signal $\mathbf{p}$ can then be expressed as the following binary hypothesis test

$$\begin{cases} H_0 : \mathbf{z} = \mathbf{c} & \mathbf{z}_i = \mathbf{c}_i, \quad i = 1, \ldots, n, \\ H_1 : \mathbf{z} = \alpha\mathbf{p} + \mathbf{c} & \mathbf{z}_i = \mathbf{c}_i, \quad i = 1, \ldots, n, \end{cases} \tag{1.44}$$

where the $\mathbf{c}_i$ are n signal-free i.i.d. measurements, traditionally referred to as the secondary data, used to estimate the unknown covariance matrix.

Let us recall the main detectors used in the literature. We will focus on their expression and theoretical analysis. Their statistical properties are analyzed in terms of DP and PFA. The DP is computed for fixed values of the detection threshold and signal-to-noise ratio (SNR). By setting SNR to zero, one obtains the value of PFA.

In practice, the detection threshold is obtained for a fixed value of PFA. Then, the DP is obtained for a given value of SNR. However, in the following we choose the theoretical approach that is, to first provide the expression for the DP of the detector followed by the corresponding PFA.

Kelly's GLRT

Assuming that the primary and secondary data are CN-distributed with an unknown covariance matrix $\mathbf{C}$, Kelly [90] has proposed a GLRT

$$\Lambda_{\text{Kelly}}\left(\widehat{\mathbf{C}}\right) = \frac{\left|\mathbf{p}^H\widehat{\mathbf{C}}^{-1}\mathbf{z}\right|^2}{\left(\mathbf{p}^H\widehat{\mathbf{C}}^{-1}\mathbf{p}\right)\left(n + \mathbf{z}^H\widehat{\mathbf{C}}^{-1}\mathbf{z}\right)} \underset{H_1}{\overset{H_0}{\lessgtr}} \lambda_{\text{Kelly}}, \tag{1.45}$$

where $\widehat{\mathbf{C}}$ is the SCM built with secondary data using Eq. (1.9). In [99] it was shown that $\Lambda_{\text{Kelly}} \overset{d}{=} \dfrac{\mathcal{K}}{\mathcal{K}+1}$ where $\mathcal{K} \overset{d}{=} \dfrac{\chi^2_2(2\delta b)}{\chi^2_{2n-2p+2}(0)}$ with $b \sim \beta(n-p+2, p-1)$, where $\beta(a,b)$ denotes the following beta distribution

$$f_\beta(x; a, b) = \frac{a+b-1!}{(a-1)!(b-1)!} x^{a-1}(1-x)^{b-1}\mathbb{1}_{[0,1]}(x). \tag{1.46}$$

After some derivations, we can obtain the following P_d for Λ_{Kelly}

$$\begin{aligned} P_d &= \mathbb{P}\left(\Lambda_{\text{Kelly}}\left(\widehat{\mathbf{C}}\right) > \lambda_{\text{Kelly}} | H_1\right) \\ &= 1 - \int_0^1 \frac{e^{-\delta u}}{N} \int_0^{\lambda_{\text{Kelly}}} u^{n-p+1}(1-u)^{p-2}(1-x)^{n-p}\, {}_1F_1\left(n-p+2, 1; \delta u x\right) dx du, \end{aligned} \tag{1.47}$$

where ${}_1F_1(\cdot)$ is the complex confluent hypergeometric function [3], $N = \frac{\Gamma(n-p+1)\Gamma(p-1)}{\Gamma(n+1)}$ and $\delta = \alpha^2 \mathbf{p}^H \mathbf{C}^{-1}\mathbf{p}$ stands for the SNR. Under H_0, the previous distribution for $\delta = 0$ leads to the $P_{fa} - \lambda_{\text{Kelly}}$ relationship

$$P_{fa} = \mathbb{P}\left(\Lambda_{\text{Kelly}}\left(\widehat{\mathbf{C}}\right) > \lambda_{\text{Kelly}} | H_0\right) = (1-\lambda_{\text{Kelly}})^{n-p+1}, \tag{1.48}$$

revealing that Kelly's GLRT exhibits the CFAR property.

Adaptive Matched Filter

In [150], Robey proposed another GLRT under the same setup, but considering that the covariance matrix $\mathbf{C}$ is known. The adaptive plug-in version of the detector is then obtained with the SCM $\widehat{\mathbf{C}}$ and given by

$$\Lambda_{\text{AMF}}\left(\widehat{\mathbf{C}}\right) = \frac{\left|\mathbf{p}^H \widehat{\mathbf{C}}^{-1}\mathbf{z}\right|^2}{n\left(\mathbf{p}^H \widehat{\mathbf{C}}^{-1}\mathbf{p}\right)} \underset{H_1}{\overset{H_0}{\lessgtr}} \lambda_{\text{AMF}}. \tag{1.49}$$

One can note that this statistic does not contain the factor in parentheses, found in the denominator of the Kelly's GLRT (Eq. (1.45)). This term is computationally demanding for real time systems.

In [99] it was shown that $\Lambda_{\text{AMF}} \overset{d}{=} \frac{\chi^2_2(2\delta b)}{\chi^2_{2n-2p+2}(0)} \frac{1}{b}$ leading to the following P_d for this statistic

$$P_d = 1 - \int_0^1 \frac{e^{-\delta u}}{N} \int_0^{\lambda_{\text{AMF}}} u^{n-p+2} \frac{(1-u)^{p-2}}{(1+ux)^{n-p+2}} {}_1F_1\left(n-p+2, 1; \frac{\delta u^2 x}{1+ux}\right) dx du. \tag{1.50}$$

The $P_{fa} - \lambda_{\text{AMF}}$ relationship for Λ_{AMF} is then given by

$$P_{fa} = {}_2F_1(n-p+1, n-p+2; n+1; -\lambda_{\text{AMF}}), \tag{1.51}$$

where ${}_2F_1(\cdot)$ is the hypergeometric function [3].

Adaptive Normalized Matched Filter

ANMF [39, 97], also called adaptive coherence estimator (ACE) [98], has been derived for partially homogeneous Gaussian noise where the noise of the primary data be scaled by a unknown factor γ^2 relative to the secondary data, i.e. $\mathbf{c} \sim \mathcal{CN}(\mathbf{0}, \gamma^2\mathbf{C})$ and $\mathbf{c}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$,

$$\Lambda_{\text{ANMF}}\left(\widehat{\mathbf{C}}\right) = \frac{\left|\mathbf{p}^H \widehat{\mathbf{C}}^{-1}\mathbf{z}\right|^2}{\left(\mathbf{p}^H \widehat{\mathbf{C}}^{-1}\mathbf{p}\right)\left(\mathbf{z}^H \widehat{\mathbf{C}}^{-1}\mathbf{z}\right)} \underset{H_1}{\overset{H_0}{\lessgtr}} \lambda_{\text{ANMF}}. \tag{1.52}$$

The Normalized Matched Filter (NMF) [39] was originally derived assuming that the noise covariance matrix is known. This statistics is scale-invariant w.r.t. the signal $\mathbf{z}$. Under H_0, this detector is "distribution-free," meaning that is independent of the noise level or a particular p.d.f. of the underlying noise CES distribution. Similarly to the AMF, the adaptive version of the detector is built using the SCM.

The distribution of Λ_{ANMF} is given by $\Lambda_{\text{ANMF}} \overset{d}{=} \frac{\mathcal{F}}{\mathcal{F}+1}$ where $\mathcal{F} \overset{d}{=} \frac{\chi^2_2(2\delta b)}{\chi^2_{2n-2p+2}(0)} \frac{1}{1-b}$ [99], which after some derivations results in the following P_d for a given SNR δ is

$$P_d = 1 - \int_0^1 \frac{e^{-\delta u}}{N} \int_0^{\lambda_{\text{ANMF}}} u^{n-p+1} \frac{(1-u)^{p-1}(1-x)^{n-p}}{(1-ux)^{n-p+2}} {}_1F_1\left(n-p+2, 1; \frac{\delta u x(1-u)}{1-ux}\right) dx du. \tag{1.53}$$

Under H_0 the p.d.f. of $\Lambda_{\text{ANMF}}\left(\widehat{\mathbf{C}}\right)$ is given by [141]

$$f_{\Lambda_{\text{ANMF}}(\widehat{\mathbf{C}})}(u) = M(1-u)^{n-p}{}_2F_1(n-p+2, n-p+2; n+2; u)\mathbf{1}_{[0,1]}(u), \tag{1.54}$$

where $M = \frac{(n-p+1)(p-1)}{n+1}$, resulting in the following PFA

$$P_{fa} = (1-\lambda_{\text{ANMF}})^{n-p+1}{}_2F_1(n-p+2, n-p+1; n+1; \lambda_{\text{ANMF}}). \tag{1.55}$$

If the covariance matrix $\mathbf{C}$ is known, the p.d.f. of Λ_{NMF} (called NMF) is given by

$$f_{\Lambda_{\text{NMF}}}(u) = (p-1)e^{\delta}(1-u)^{p-2}{}_1F_1(p, 1; u\delta) \tag{1.56}$$

and the probability of detection P_d for a given SNR δ and for a fixed value of the detection threshold λ_{NMF} is given by [39]

$$P_d = 1 - (p-1)\int_0^{\lambda_{\text{NMF}}} e^{\delta}(1-u)^{p-2}{}_1F_1(p, 1; u\delta)du. \tag{1.57}$$

Under H_0 Λ_{NMF} follows a beta distribution $\Lambda_{\text{NMF}} \sim \beta(1, p-1)$. This results in the following $P_{fa} - \lambda_{\text{NMF}}$ relationship

$$P_{fa} = (1-\lambda_{\text{NMF}})^{p-1}. \tag{1.58}$$

Rao test

In 2007, De Maio proposed a new detection statistic based on Rao test [118]

$$\Lambda_{\text{Rao}}\left(\widehat{\mathbf{C}}\right) = \frac{\left|\mathbf{p}^H\widehat{\mathbf{C}}^{-1}\mathbf{z}\right|^2 / \left(\mathbf{p}^H\widehat{\mathbf{C}}^{-1}\mathbf{p}\right)}{\left(n + \mathbf{z}^H\widehat{\mathbf{C}}^{-1}\mathbf{z}\right)\left[1 + \frac{1}{n}\mathbf{z}^H\widehat{\mathbf{C}}^{-1}\mathbf{z} - \frac{1}{n}\frac{\left|\mathbf{p}^H\widehat{\mathbf{C}}^{-1}\mathbf{z}\right|^2}{\mathbf{p}^H\widehat{\mathbf{C}}^{-1}\mathbf{p}}\right]} \underset{H_1}{\overset{H_0}{\lessgtr}} \lambda_{\text{Rao}}. \tag{1.59}$$

The distribution of Rao's statistic can be expressed as $\Lambda_{\text{Rao}} \overset{d}{=} \frac{\mathcal{R}b}{\mathcal{R}+1}$ where $\mathcal{R} \overset{d}{=} \frac{\chi^2_2(2\delta b)}{\chi^2_{2n-2p+2}(0)}$. Consequently, we can derive the corresponding detection probability

$$P_d = 1 - \int_0^1 \frac{e^{-\delta u}}{N}\int_0^{\lambda_{\text{Rao}}/(u-\lambda_{\text{Rao}})} u^{n-p+1}\frac{(1-u)^{p-2}}{(1+x)^{n-p+2}}{}_1F_1\left(n-p+2, 1; \frac{\delta u x}{1+x}\right)dxdu. \tag{1.60}$$

This test also ensures the CFAR property with the following PFA

$$P_{fa} = (1-\lambda_{\text{Rao}})^n \tag{1.61}$$

and it is invariant to the set of transformation defined in [118].

1.2.4 Low-rank detection

In LR detection problem the disturbance is composed of a LR Gaussian clutter and an AWGN

$$\begin{cases} H_0 : \mathbf{z} = \mathbf{c} + \mathbf{n} & \mathbf{z}_i = \mathbf{c}_i + \mathbf{n}_i, \quad i = 1, \ldots, n, \\ H_1 : \mathbf{z} = \alpha\mathbf{p} + \mathbf{c} + \mathbf{n} & \mathbf{z}_i = \mathbf{c}_i + \mathbf{n}_i, \quad i = 1, \ldots, n, \end{cases} \tag{1.62}$$

where $\mathbf{c} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_r)$ is the Gaussian clutter complex vector with $\text{rank}(\mathbf{\Sigma}_r) = r \ll p$ and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \gamma^2\mathbf{I}_p)$ is the AWGN . Consequently, the covariance matrix of the secondary data can be written as $\mathbf{\Sigma} = \mathbf{\Sigma}_r + \gamma^2\mathbf{I}_p$.

The eigenvalue decomposition of $\mathbf{\Sigma}_r$ can be written as

$$\mathbf{\Sigma}_r = \mathbf{U}_r \mathbf{\Lambda}_r \mathbf{U}_r^H \quad \text{with} \quad \begin{array}{l} \mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_r] \in \mathcal{U}_r^p, \\ \mathbf{\Lambda}_r = \operatorname{diag}(\boldsymbol{\lambda}_r), \\ \boldsymbol{\lambda}_r = [\lambda_1, \ldots, \lambda_r]\,. \end{array} \tag{1.63}$$

The projector $\mathbf{\Pi}_r$ onto the clutter subspace and the projector $\mathbf{\Pi}_r^\perp$ onto the subspace orthogonal to the clutter subspace are defined as

$$\begin{cases} \mathbf{\Pi}_r = \mathbf{U}_r \mathbf{U}_r^H \\ \mathbf{\Pi}_r^\perp = \mathbf{I} - \mathbf{\Pi}_r = \mathbf{U}_r^\perp \left(\mathbf{U}_r^\perp\right)^H \end{cases} \quad \text{with} \quad \mathbf{U}_r^\perp = [\mathbf{u}_{r+1}, \ldots, \mathbf{u}_p]\,. \tag{1.64}$$

In order to remove the clutter we can perform an approximative whitening

$$\begin{cases} H_0 : \mathbf{y} = \left(\mathbf{U}_r^\perp\right)^H \mathbf{z} = \mathbf{n}_0 & \mathbf{y}_i = \mathbf{n}_{0,i}, \quad i = 1, \ldots, n, \\ H_1 : \mathbf{y} = \left(\mathbf{U}_r^\perp\right)^H \mathbf{z} = \alpha \mathbf{d} + \mathbf{n}_0 & \mathbf{y}_i = \mathbf{n}_{0,i}, \quad i = 1, \ldots, n. \end{cases} \tag{1.65}$$

Among various LR detectors [95, 148, 69, 79, 172, 31], we will focus on the LR-ANMF defined as

$$\Lambda_{\text{LR}}\left(\widehat{\mathbf{\Pi}}_r^\perp\right) = \frac{\left|\mathbf{p}^H \widehat{\mathbf{\Pi}}_r^\perp \mathbf{z}\right|^2}{\left(\mathbf{p}^H \widehat{\mathbf{\Pi}}_r^\perp \mathbf{p}\right)\left(\mathbf{z}^H \widehat{\mathbf{\Pi}}_r^\perp \mathbf{z}\right)}, \tag{1.66}$$

where $\widehat{\mathbf{\Pi}}_r^\perp$ is the estimate of $\mathbf{\Pi}_r^\perp$.

In [69] it has been shown that the distribution of $\Lambda_{\text{LR}}\left(\widehat{\mathbf{\Pi}}_r^\perp\right)$ under the H_0 hypothesis can be approximated with

$$\Lambda_{\text{LR}}\left(\widehat{\mathbf{\Pi}}_r^\perp\right) \approx \frac{\mathcal{A}_1^2}{\mathcal{A}_2} \tag{1.67}$$

with

$$\begin{aligned} \mathcal{A}_1 &= s_1 - \frac{1}{n} x_1 x_2 s_2, \\ \mathcal{A}_2 &= |s_1|^2 + x_2^2 - \frac{2}{n} x_1 x_3 \Re(s_1 s_2 + x_2 s_2), \end{aligned} \tag{1.68}$$

where $x_1 = \sqrt{\frac{1}{2}\chi^2(2r)}$, $x_2 = \sqrt{\frac{1}{2}\chi^2(2(p-r-1))}$, $x_3 = \sqrt{\frac{1}{2}\chi^2(2p)}$, $s_1 \sim \mathcal{CN}(0,1)$, $s_2 \sim \mathcal{CN}(0,1)$ and $s_3 \sim \mathcal{CN}(0,1)$ are i.i.d. variables.

This result has been obtained by a first order approximation and thus, does not provide the exact distribution of LR-ANMF. However, one can note that this distribution does not depend on the structure of the clutter subspace meaning that the LR-ANMF is approximately CFAR.

1.2.5 Conclusion

In this section we have presented the signal detection problem in the case of additive Gaussian disturbance. When the covariance matrix of the secondary data is unknown the SCM is plugged-in instead. Since the SCM is sensitive to heavy-tailed distributed samples, this family of Gaussian detectors can exhibit poor performance in non-Gaussian environments. In this case, a *robust detector* can be built as a classical Gaussian detector where an M-estimator of the scatter is plugged-in instead of the SCM.

In this context, in Chapter 4, we provide an accurate statistical performance analysis of the robust detectors by comparing them to a *Gaussian Core Equivalent Detector (GCED)*. Thanks to the results obtained in Chapter 2, we derive new properties for robust estimators and show how they can be useful in detection threshold derivation.

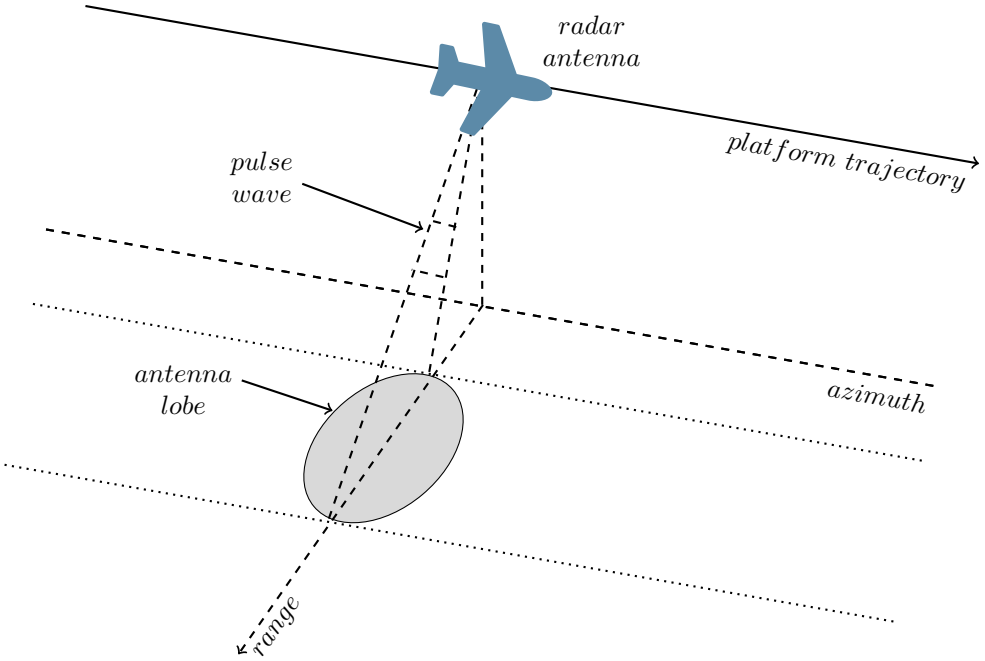


Figure 1.2 – Principle of the construction of SAR images

1.3 Polarimetric SAR images

1.3.1 Introduction

Synthetic Aperture Radar (SAR) has been widely used in Earth remote sensing for several decades. SAR provides high-resolution images that are independent of daylight and weather conditions and thus, are very convenient for a wide range of applications such as topography, geology, forestry, oceanography, environment monitoring including civil infrastructure monitoring, military surveillance, etc. SAR images are created using radio waves that are emitted, backscattered and finally, received and recorded by the radar.

SAR antennas can emit and receive waves with different polarizations, producing polarimetric SAR (PolSAR) images. These images are usually given by three combinations of orthogonally emitted and received polarizations, composing a scattering vector in each pixel, that is then used similarly to the three color channels in a synthesized image. SAR polarimetry is employed in the analysis of physical information of a target scene that is based on the measurements of the polarimetric properties of man-made and natural scatterers.

Because of the coherent processing of the scattered signals, SAR images are susceptible to present speckle noise that needs to be removed. In the context of PolSAR images, despecking is equivalent to the covariance matrix estimation of the scattering vector.

In the following, we give a brief overview of SAR technologies together with the basics of SAR polarimetry. The statistical models for PolSAR images are detailed. Finally, some despeckling methods are discussed.

1.3.2 Synthetic aperture radar (SAR) imagery

Radar is a system generally used to detect objects and measure their range. Another application is observation and analysis of ground surface through SAR imagery. An SAR is an imaging radar fixed on a moving airborne or spaceborne platform. The great ad-

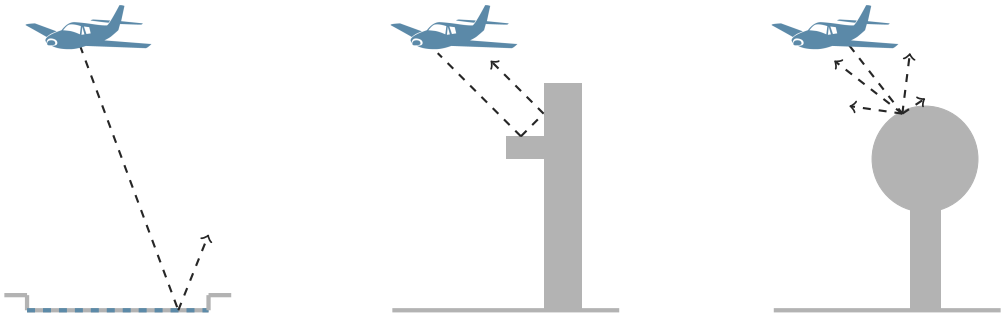


Figure 1.3 – Scattering mechanisms. From left to right: smooth surface (single bounce), a man-made structure (double bounce) and the forest canopy (multiples bounces).

vantage of SAR systems, comparing to optical ones, is that they provide high-resolution two-dimensional images that do not depend on daylight and/or weather conditions. Just as a conventional radar, an SAR sensor transmits electromagnetic radar waves and collects the backscattered echoes. The transmitted pulse interacts with the surface and a fraction of the broadcast energy is received back at the radar antenna. The amplitude and phase of the backscattered signal depends on the properties of the surface and imaged objects such as roughness, moisture, geometry, etc.

Figure 1.2 illustrates the common SAR geometry. The platform moves in the azimuth direction. The direction perpendicular to the radar's moving path is called the slant range. Thanks to side-looking geometry, an electromagnetic pulse (considered as plane) is repeatedly transmitted with a non-null incidence angle in range direction. Consequently, imaging radars have a two-dimensional resolution: the range resolution that depends on time period and is inversely proportional to the system bandwidth and the azimuth resolution presenting the smallest separation between two point targets that can be detected by the radar. The wavelengths are usually chosen in L, C or X bands.

When the signal reaches the ground it is diffused in one or several directions depending on the surface properties. Figure 1.3 portrays different scattering mechanisms achieved for smooth surface, man-made structure and forest cover. In SAR image each pixel corresponds to an area on the ground, called resolution cell. The resulting signal in a resolution cell is obtained as the coherent sum of all backscattered signals returned back to the cell. The strong variations between the resulting amplitudes and phases from pixel to pixel cause a particular effect observed in SAR images called speckle (see Figure 1.4b). The value of speckle is thus no deterministic, but depends on many scatterers with a random distribution within a resolution cell.

The speckle is commonly referred to as noise, but it should be noted that it can not be reduced simply by increasing the transmit signal power. It has a multiplicative character and its variance grows with the signal intensity. A widely used method for speckle reduction is multi-look filtering, which is basically a non-coherent averaging of the intensity image (square of the modulus of the complex backscattered signal). This method can significantly smooth the image, but at the cost of a resolution lost.

1.3.3 Polarimetry

Transmitted SAR waves are polarized and some materials can reflect different polarizations with different intensities. Some others can also convert one polarization into another. SAR systems can then transmit a mixture of polarizations and use receiving antennas with a specific polarization, in order to collect several signals from the same series of pulses.

Traditional SAR systems contain a fixed-polarization antenna for both transmission and reception of radio signals. In such a way, for every resolution cell the SAR sensor measures a single radar reflectivity for a specific transmitting and receiving polarization. A consequence of this implementation is that the reflected wave measures only a couple of transmitted/received polarization quantity and any additional information about the scattering process contained in the polarization properties of the scattered signal is lost. To ensure that all the information of the scattered wave is retained, the polarization of the scattered wave must be measured through a vector measurement process, enabling to use different polarizations to distinguish between various scattering mechanisms.

The basic concept of SAR polarimetry is given by the 2×2 complex scattering matrix that describes the transformation of the transmitted wave vector into the received wave vector performed by the scatterer

$$\mathbf{S} = \begin{bmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{bmatrix}.$$

The elements of $\mathbf{S}$ are the four complex scattering amplitudes where the subscripts horizontal (H) or vertical (V) indicate associated received and transmitted polarization. In PolSAR imagery the absolute phase is in most cases neglected and only the relative phases between the elements are considered.

The scattering matrix can be measured by transmitting two orthogonally polarized waves and measuring the scattered waves in two orthogonal polarizations. Most PolSAR systems operate in the linear H-V basis. By transmitting a H polarized wave and receiving in H and V polarization the S_{HH} and S_{HV} elements are measured. Then, two remaining coefficients S_{VH} and S_{VV} are measured by transmitting a V-polarized wave and receiving in H and V polarization. In monostatic case, i. e. when the same antenna is receiver and transmitter, the cross-correlated elements of $\mathbf{S}$ are assumed to be equal $S_{HV} = S_{VH}$. In this case, the number of independent parameters in $\mathbf{S}$ is equal to five, i.e. three amplitudes and two relative phases. Therefore, there are only three complex coefficients required to characterize the scattering vector

$$\mathbf{k}_{\mathrm{L}} = \begin{bmatrix} S_{HH} & S_{VV} & \sqrt{2}S_{HV} \end{bmatrix}^T. \tag{1.69}$$

Alternatively, the scattering vector is replaced by the linear transformation

$$\mathbf{k}_{\mathrm{P}} = \frac{1}{\sqrt{2}} \begin{bmatrix} S_{HH} + S_{VV} & S_{HH} - S_{VV} & 2S_{HV} \end{bmatrix}^T \tag{1.70}$$

known as the Pauli representation of the scattering vector [105]. Figure 1.5 gives an example of a PolSAR image in Pauli basis compared to an SAR intensity image.

In order to fully characterize distributed scatterers one needs to analyze the second-order characteristics of scattering vectors, that are defined as

$$\mathbf{C}_{\mathrm{L}} = \mathbb{E}\left[\mathbf{k}_{\mathrm{L}} \mathbf{k}_{\mathrm{L}}^H\right] \tag{1.71}$$

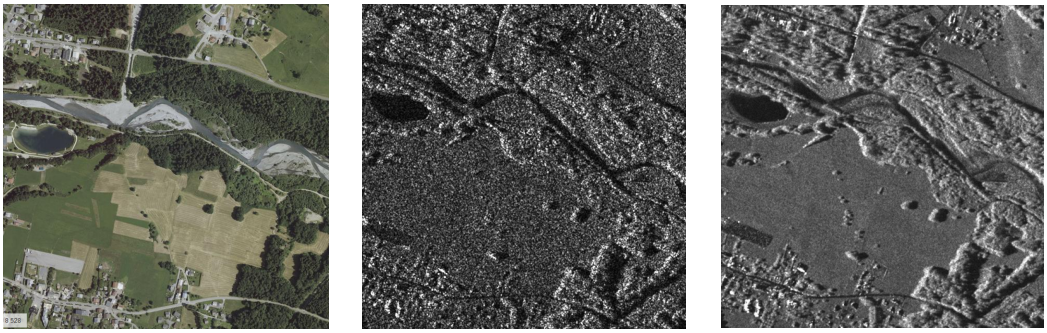
called covariance matrix, and

$$\mathbf{C}_{\mathrm{P}} = \mathbb{E}\left[\mathbf{k}_{\mathrm{P}} \mathbf{k}_{\mathrm{P}}^H\right] \tag{1.72}$$

called coherency matrix.

The scattering vector in Pauli basis can be represented as

$$\mathbf{k}_{\mathrm{P}} = \frac{1}{\sqrt{2}} \begin{pmatrix} S_{HH} + S_{VV} \\ S_{HH} - S_{VV} \\ 2S_{HV} \end{pmatrix} = \frac{S_{HH}}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \frac{S_{VV}}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + \frac{2S_{HV}}{\sqrt{2}} \begin{pmatrix} 0 \\ 0 \\ \sqrt{2} \end{pmatrix}, \tag{1.73}$$



(a) Optical image (b) Single-look image (c) Multi-look image

Figure 1.4 – From left to right: an optical image, the single-look amplitude and the temporally multi-looked amplitude of an SAR image. The temporal multi-looking reduces the speckle while preserving the resolution, but it is only valid for temporally stable areas.

leading to the following relationship between the lexical and the Pauli representation

$$\mathbf{k}_{\mathrm{P}} = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1 & 0\\ 1 & -1 & 0\\ 0 & 0 & \sqrt{2}\end{pmatrix}\mathbf{k}_{\mathrm{L}}, \tag{1.74}$$

$$\mathbf{C}_{\mathrm{P}} = \frac{1}{2}\begin{pmatrix}1 & 1 & 0\\ 1 & -1 & 0\\ 0 & 0 & \sqrt{2}\end{pmatrix}\mathbf{C}_{\mathrm{L}}\begin{pmatrix}1 & 1 & 0\\ 1 & -1 & 0\\ 0 & 0 & \sqrt{2}\end{pmatrix}. \tag{1.75}$$

Another solution to ease the interpretation of the scattering phenomena is to consider a eigenvalue decomposition of $\mathbf{C}$ invariant of the chosen basis proposed by Cloude and Pottier [30]

$$\mathbf{C} = \mathbf{U}\begin{pmatrix}\lambda_1 & 0 & 0\\ 0 & \lambda_2 & 0\\ 0 & 0 & \lambda_3\end{pmatrix}\mathbf{U}^H, \tag{1.76}$$

where $\lambda_1 \geq \lambda_2 \geq \lambda_3$ and $\mathbf{U} = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \mathbf{u}_3]$ the respective eigenvectors that can be represented as

$$\mathbf{u}_i = \begin{pmatrix}\cos\alpha_i e^{j\gamma_i}\\ \sin\alpha_i\cos\beta_i e^{j\gamma_i}\\ \sin\alpha_i\sin\beta_i e^{j\gamma_i}\end{pmatrix}. \tag{1.77}$$

Thanks to these decompositions, we can extract three important physical features describing the underlying physical phenomena

- The entropy $H \in [0,1]$

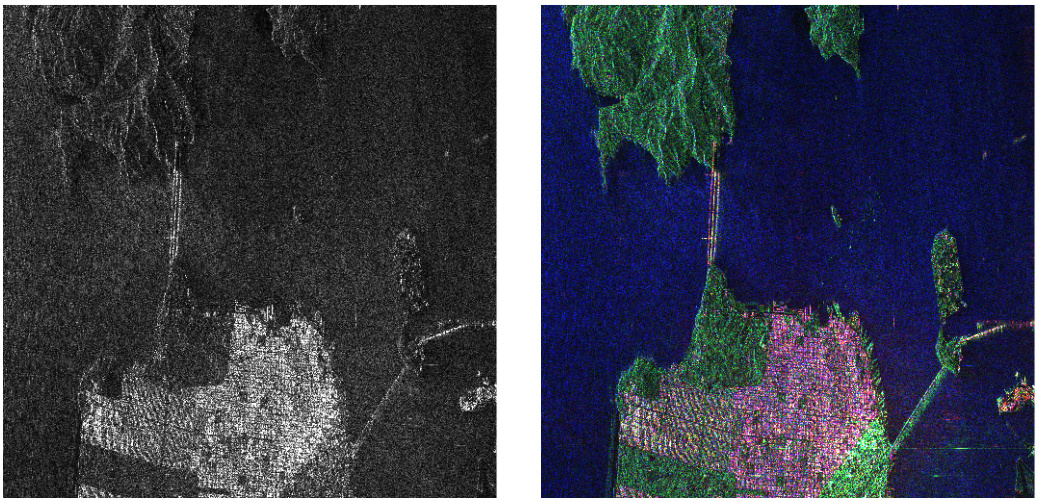
$$H = -\sum_{i=1}^{3} p_i \log_3 p_i \quad \text{with} \quad p_i = \frac{\lambda_i}{\sum_{i=1}^{3}\lambda_i}, \tag{1.78}$$

  where p_i are referred as to the scattering probabilities. The entropy measures the randomness of the scattering process. For instance, when $H = 0$, there is only one single mechanism involved, while for $H = 1$ three pure random mechanisms are involved.

- The anisotropy $A \in [0,1]$

$$A = \frac{\lambda_2 - \lambda_3}{\lambda_2 + \lambda_3}, \tag{1.79}$$

  provides complementary information. When $H > 0$, the anisotropy indicates how many scattering mechanisms are involved.



(a) SAR intensity image (b) PolSAR image

Figure 1.5 – An example of a SAR intensity image and the corresponding PolSAR image. The PolSAR image is displayed using an RGB representation based on the Pauli basis.

- The mean scattering angle $\overline{\alpha} \in (-\pi, \pi]$

$$\overline{\alpha} = \sum_{i=1}^{3} p_i \alpha_i, \tag{1.80}$$

giving the information about the underlying scattering process. When $\alpha = 0$, a single bounce scattering is produced by a rough surface. When $\alpha = 1/4$ volume scattering is produced. Finally, the case $\alpha = 1/2$ corresponds to double bounce scattering.

1.3.4 Modeling

The polarimetric information is extracted through the statistical properties of the data. Therefore, it is very important to have an accurate statistical model to interpret the information about observed area.

Gaussian model for the radar signals have been frequently employed when the spatial resolution of PolSAR images is moderate [71, 61, 112, 109]. The number of scatterers in a resolution cell of low resolution data is large and according to the CLT, Gaussian distribution could give a good approximation of the data distribution, especially in homogeneous areas. For PolSAR data, the mean value of the complex vector is generally assumed to be zero, and all the statistical properties are determined by the covariance matrix under the Gaussian assumption.

In PolSAR imagery, multi-looking despeckling is equivalent to the computation of the SCM built with surrounding pixels

$$\widehat{\mathbf{\Sigma}}_{\mathrm{SCM}} = \frac{1}{L} \sum_{i=1}^{L} \mathbf{k}_i \mathbf{k}_i^H, \tag{1.81}$$

where L is the nominal number of looks (equivalent to the number of pixels used for estimation). Since the pixels are correlated and multi-looked data is usually modeled as an average of independent measurements, the actual number of correlated samples is replaced

by an equivalent number of independent ones, defined as

$$L_e = \frac{\mathbb{E}\left[I\right]^2}{\mathrm{Var}\left[I\right]}, \tag{1.82}$$

where I is a single polarization intensity. As explained in the first part of this chapter, the SCM, under the Gaussian assumption, follows a complex Wishart distribution, which is highly used in literature and applications of PolSAR data [109].

Nowadays, when most of SAR systems operate with considerably higher resolutions than before, Gaussian distribution fails to describe the fluctuations between resolution cells. This is principally apparent in heterogeneous areas where the texture can drastically change between surrounding pixels. Consequently, one needs more complex models in order to describe these scenarios. The scalar texture model appeared as a logical solution to this issue. This model, also referred as to SIRV (see Definition 1.1.12), assume the data is given as a product of a Gaussian random process with the square root of a nonnegative (random) scalar variable that contains the information about the texture variation, i.e.

$$\mathbf{k} = \sqrt{\tau}\mathbf{n} \tag{1.83}$$

where τ is the texture parameter with mean value equal to 1, and $\mathbf{n}$ is the speckle vector, following a multivariate CN distribution. Since the texture parameter is supposed to have unity mean it is only used to model the variation of the radar cross section, while the intensity information is contained in the speckle component.

Few distributions have been proposed for the random parameter τ in the literature, resulting in different multivariate models for the scattering vector $\mathbf{k}$. Hereafter we list some of widely used PolSAR models.

- One of the most used radars texture models is gamma distribution [84, 188] which can be written as
$$f(\tau;\upsilon) = \frac{\upsilon^{\upsilon}}{\Gamma(\upsilon)}\tau^{\upsilon-1}\exp\left(-\upsilon\tau\right) \tag{1.84}$$
in order to ensure the unity mean of τ. This case results in multivarite K-distribution (see Example 1.1.3). This distribution is often used to model forests and the sea surface.
- For the beta distribution [15] of the texture parameter the p.d.f. is given by
$$f(\tau;\upsilon,\varrho) = \frac{\Gamma(\upsilon)}{\Gamma(\varrho)\Gamma(\upsilon-\varrho)}\frac{\upsilon}{\varrho}\left(\frac{\upsilon}{\varrho}\tau\right)^{\upsilon-1}\left(1-\frac{\upsilon}{\varrho}\tau\right)^{\varrho-\upsilon-1}, \quad \tau\in\left[0,\frac{\varrho}{\upsilon}\right] \tag{1.85}$$
the scattering vector is assumed to have the W-distribution [3]. Note that this W-distribution is CCG version of the W-distribution defined in Example 1.1.4 which was introduced particularly for modeling radar data.
- The Normal Inverse Gaussian (NIG) distribution [50] assumes that the texture follows an inverse Gaussian distribution
$$f(\tau;\upsilon) = \left(\frac{\upsilon}{2\pi}\right)^{1/2}\tau^{-3/2}\exp\left(-\frac{1}{2}\left(\frac{\upsilon}{\tau}+\upsilon\tau\right)+\upsilon\right). \tag{1.86}$$
The advantage of this distribution is that is able to captures large distribution shape variation.
- The inverse gamma distribution of the texture parameter [59] which is given by
$$f(\tau;\upsilon) = \frac{(\upsilon-1)^{\upsilon}}{\Gamma(\upsilon)}\tau^{-\upsilon-1}\exp\left(\frac{-\upsilon-1}{\tau}\right), \tag{1.87}$$
is particularly convenient for modeling extremely heterogeneous regions such as urban areas. The IG-distributed texture leads to the t-distribution for scattering vector (see Example 1.1.2).

- If the texture parameter follows a Fisher distribution [165] given by

$$f(\tau;\upsilon,\varrho)=\frac{\Gamma(\upsilon+\varrho)}{\Gamma(\upsilon)\Gamma(\varrho)}\frac{\upsilon}{\varrho-1}\left(\frac{\upsilon}{\varrho-1}\tau\right)^{\upsilon-1}\left(\frac{\upsilon}{\varrho-1}\tau+1\right)^{-\upsilon-\varrho} \tag{1.88}$$

  the scattering vector is Kummer-U distributed. Fisher distribution covers a large range of distributions and thus, is able to model different types of textures.

The aforementioned distributions are only some of the used texture models for PolSAR data. Beside the scalar texture, other representations such as Multi-Texture Models [53], Finite Mixture Models [122] and Copula Based Model [121] have been used to describe PolSAR data. We refer the reader to cited literature for more details about these models. A great survey on PolSAR statistical models is given in [48].

1.3.5 PolSAR image despeckling

PolSAR filtering techniques have received much attention over the last three decades and many different methods have been proposed. These techniques can be generally categorized into several classes basing on basic principles of the methods. Hereafter, we make a brief review of existing methods and refer the reader to the cited literature for more details.

- The first group are local window filters, such as the well-known multi-look or boxcar filter which computes the value of the center pixel in the window as the average of all the neighborhood pixels. As we have already explained in Section 1.3.2, this filter is simple to apply and can significantly reduce speckle in homogeneous areas, but degrades the spatial resolution of the image. In [110] the minimization of the mean square error (MSE) was proposed to estimate the diagonal elements of the scattering matrix while the extension in [74] suggests estimating the whole matrix. In order to better preserve edges and image details Lee et al. [108] proposed to locally select the best window among a few pre-defined windows considering Local Linear Minimum Mean Square Error (LLMMSE). In [175] the authors proposed an intensity-driven adaptive-neighborhood (IDAN) filter which selects similar pixels using a region growing technique. During the estimation, all pixels are given the same weight which causes a selection bias. The sigma filter, first introduced for single-channel SAR data [107], was extended to PolSAR [111] generalizing the idea to select similar pixels in the local window and to filter the center pixel by the LMMSE estimator. In order to avoid the introduction of bias during the filtering process, D'Hondt et al. [49] adapted the idea of bilateral filter to PolSAR data where the value of the center pixel is estimated by the weighted average of all the pixels in a large square window.
- The methods based on partial differential equation (PDE) which are well-suited for Gaussian noise removal have been adapted to PolSAR data in several works. One can cite [143] where an anisotropic diffusion (AD) filter was proposed, based on progressive smoothing of the image starting with the original image and according to the solution of a given PDE. In [55], the authors proposed a trace-based filters, an iterative PDE filter that perform the filtering that corresponds to a local convolution by oriented Gaussian filters.
- The variational methods employ global regularization information to remove noise in the images. The main idea is to construct an energy functional composed of a data fidelity term and a regularization term, that significantly smooth the image while preserve edges. All the pixels are simultaneously processed by minimizing the energy functional. In [125], Nie et al. proposed a PolSAR TV technique based on the Wishart model of covariance matrix. In [126] Nie proposed an improvement of this method by suggesting a new regularization term.
- In recent years a growing interest appeared for machine learning based filtering methods. The method introduced in [158] expresses the filtering problem as an optimization

problem, which is solved using the stochastic relaxation algorithm called simulated annealing. In [76], the authors propose a filtering method based on subspace decomposition, representing a parameter space as a composition of two subspaces: the signal subspace and the orthogonal noise subspace. The polarimetric information is then retrieved from the significant principal component analysis coefficients. Another important group of ML-based methods use sparse representation [185] where a high-dimensional vector can be represented as a linear combination of signals with reduced dimension. Also, in recent years there has been considerable interest in deep learning based methods [28, 186].

- Finally, the last group is composed of non-local means based methods. These approaches rely on patch comparison to select similar samples in the whole image. The idea was born with the first works of Buades *et al.* [22, 21] and then extended and adapted to (Pol)SAR images by several authors such as in [47, 25]. In Chapter 5 we will discuss in details the methodology employed in these techniques and see how it can be used in PolSAR despeckling.

1.3.6 Conclusion

In this section, we have presented the basic principles of PolSAR imagery. The scalar texture statistical models for PolSAR data have been reviewed, followed by a brief survey on despeckling methods. We have seen that PolSAR data can be modeled by the CES distributions defined in Section 1.1. In this context, the despeckling of PolSAR data corresponds to the estimation of the scatter matrix.

In Chapter 5, we will examine in details non-local means methods and a state-of-the-art technique called NL-SAR. We will see how new properties of M-estimators, introduced in Chapter 2, can be exploited in PolSAR image despeckling.

Chapter 2

New statistical properties for M-estimators

This chapter provides an original approach to better understanding the behavior of robust M-estimators of the scatter matrix. To that end, a Gaussian-core representation, that is equivalent to the classical stochastic representation, of CES distributions is established. Thanks to this representation, the Gaussian-Core Wishart Equivalent (GCWE) of an M-estimator is defined as the SCM built with the Gaussian cores of the CES data. Then, the asymptotic distribution between an M-estimator and its GCWE is derived, showing that the behavior of an M-estimator can be better characterized by the Wishart distribution of its GCWE than with its asymptotic normal properties. Moreover, some particular cases of M-estimators are analyzed. Finally, Monte Carlo simulations are provided in order to validate the theoretical results.

Publications associated with this chapter:

[C1] **G. Draškovi**ć and F. Pascal, "New properties for Tyler's covariance matrix estimator," *2016 50th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, 2016, pp. 820-824.

[J1] **G. Draškovi**ć and F. Pascal, "New Insights Into the Statistical Properties of M-Estimators," *IEEE Transactions on Signal Processing*, vol. 66, no. 16, pp. 4253-4263, 2018.

[C3] **G. Draškovi**ć and F. Pascal, "New insights into the statistical properties of M-estimators and application to signal detection," *International Conference on Robust Statistics (ICORS)*, Leuven, Belgium, July 2018.

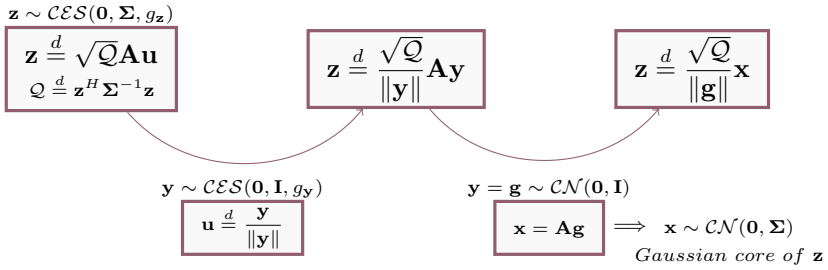


Figure 2.1 – Gaussian-core representation

2.1 Assumed model

In this section, we describe the assumed model of the data and provide definitions required for the analysis done in this thesis.

2.1.1 Gaussian-Core representation

As we have seen in Chapter 1, CES distributions can be represented as affine transforms of a scale mixture of $\mathcal{U}(\mathcal{CS}^{p-1})$. Moreover, the CES distributions belonging to the subclass of CCG distributions can be represented as affine transforms of a scale mixture of $\mathcal{CN}(\mathbf{0}, \mathbf{I})$. This presents the starting point for the analysis done in this thesis.

In order to better explain the context of the work, we will rewrite the stochastic representation given by Eq. (1.14) using the fact that for $\mathbf{u} \sim \mathcal{U}(\mathcal{CS}^{p-1})$ one has $\mathbf{u} \overset{d}{=} \mathbf{g}/\|\mathbf{g}\|$, where $\mathbf{g} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. As it can be seen in Figure 2.1, from the general stochastic representation of a CES-distributed vector one has that $\mathbf{u} \overset{d}{=} \mathbf{y}/\|\mathbf{y}\|$ where $\mathbf{y} \sim \mathcal{CES}(\mathbf{0}, \mathbf{I}, g_{\mathbf{y}})$. Then, since the CN distribution is a particular case of CES distributions, we can set $\mathbf{y} = \mathbf{g} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ and consequently $\mathbf{u} \overset{d}{=} \mathbf{g}/\|\mathbf{g}\|$. Finally, we define the Gaussian core $\mathbf{x}$ of a CES-distributed vector $\mathbf{z}$ as $\mathbf{x} = \mathbf{A}\mathbf{g}$. Hence, one obtains the following representation.

Definition 2.1.1: Gaussian cores of CES

Each $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}})$ has the following stochastic representation

$$\mathbf{z} \overset{d}{=} \frac{\sqrt{\mathcal{Q}}}{\|\mathbf{g}\|}\mathbf{A}\mathbf{g}, \tag{2.1}$$

where $\mathbf{g} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, $\mathcal{Q}$ is a non-negative real random variable, and $\mathbf{\Sigma} = \mathbf{A}\mathbf{A}^H$ is a factorization of $\mathbf{\Sigma}$. We refer to $\mathbf{x} = \mathbf{A}\mathbf{g}$ as the Gaussian core of $\mathbf{z}$.

This representation is completely equivalent to Eq. (1.14), where we choose to represent the vector $\mathbf{u}$ as a ratio of a CN-distributed vector $\mathbf{g}$ and its norm. One can note that if $\sqrt{\mathcal{Q}}/\|\mathbf{g}\|$ is independent of $\mathbf{g}$, the r.v. $\mathbf{z}$ is CCG-distributed, i.e. it can be represented as by Eq. (1.16) where $\tau \overset{d}{=} \dfrac{\mathcal{Q}}{\|\mathbf{g}\|^2}$. In general, the relationship between $\mathcal{Q}$ and $\|\mathbf{g}\|^2$ is more complex and can not be described by a simple linear model. However, for particular parameters of a CES distribution when the CES distribution tends to the CN one, one has $\mathcal{Q} \overset{d}{\to} \|\mathbf{g}\|^2$.

2.1.2 Gaussian-Core Wishart Equivalent

Once we have defined Gaussian cores of CES-distributed data, we can introduce the Gaussian-Core Wishart Equivalent (GCWE) of an M-estimator.

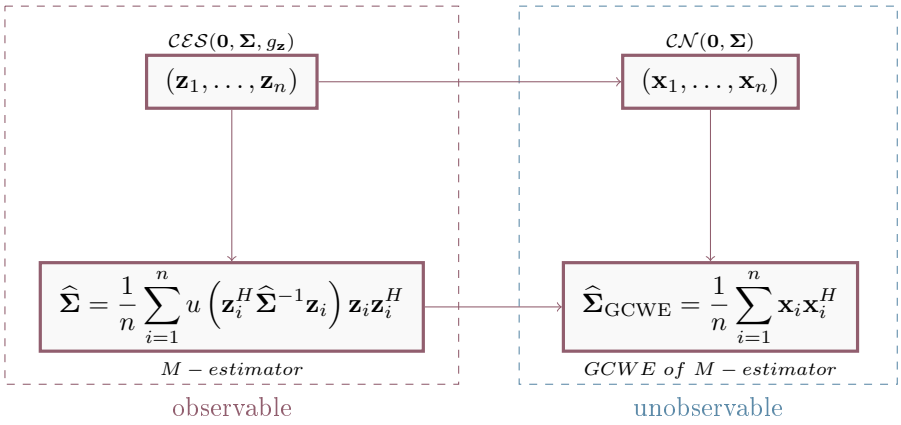


Figure 2.2 – GCWE of an M-estimator

As it can be seen in Figure 2.2, we assume that the measurements $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ are CES-distributed and we apply an M-estimator to estimate the scatter matrix. The Gaussian cores of the observations are then used for a theoretical construction of the SCM that represents the GCWE of the corresponding M-estimator. Although $\widehat{\mathbf{\Sigma}}_{\text{GCWE}}$ cannot be obtained in practice, it is well-known that it follows the complex Wishart distribution for any matrix $\mathbf{A}$ (cf. Definition 1.1.9). We can now state the following definition.

Definition 2.1.2: Gaussian-Core Wishart Equivalent (GCWE)

Let n i.i.d. measurements $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ be drawn as $\mathbf{z}_i \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}})$ and denote $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ their Gaussian cores as $\mathbf{z}_i = \sqrt{Q_i}/\|\mathbf{g}_i\|\mathbf{x}_i$ (cf. Definition 2.1.1), $i = 1, \ldots, n$. Let $\widehat{\mathbf{\Sigma}}$ be an M-estimator built with $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ using Eq. (1.28). The SCM built from the Gaussian cores, i.e.

$$\widehat{\mathbf{\Sigma}}_{\text{GCWE}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^H \tag{2.2}$$

is referred to as the GCWE of $\widehat{\mathbf{\Sigma}}$. It is important to notice that this matrix cannot be computed in practice, but represents a theoretical equivalent.

In the following, we propose to analyze the properties of an M-estimator in the CES framework by comparing it to the corresponding GCWE.

2.2 Convergence towards GCWE

In this section, we introduce new properties for M-estimators. The asymptotic properties of the difference between an M-estimator and its GCWE are derived. Moreover, the results for particular M-estimators are derived and supported by simulation experiments.

2.2.1 Results for the general case

In this section we propose new properties for M-estimators in the CES framework. We consider an n-sample $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ with $\mathbf{z}_i \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}})$, $i = 1, \ldots, n$, and want to estimate the scatter matrix of the population employing an M-estimator defined as a solution of Eq. (1.28). Then, one has the following results.

Theorem 2.2.1: General case [J1]

Let $\widehat{\mathbf{\Sigma}}$ and $\widehat{\mathbf{\Sigma}}_{\text{GCWE}}$ be defined by Eqs. (1.28) and (2.2), respectively. The asymptotic distribution of $\sigma\widehat{\mathbf{\Sigma}} - \widehat{\mathbf{\Sigma}}_{\text{GCWE}}$ is given by

$$\sqrt{n}\text{vec}\left(\sigma\widehat{\mathbf{\Sigma}} - \widehat{\mathbf{\Sigma}}_{\text{GCWE}}\right) \xrightarrow{d} \mathcal{GCN}\left(\mathbf{0}, \mathbf{\Gamma}, \mathbf{\Omega}\right), \tag{2.3}$$

where σ is the solution of Eq. (1.31). The matrices $\mathbf{\Gamma}$ and $\mathbf{\Omega}$ are defined as

$$\begin{cases} \mathbf{\Gamma} = \sigma_1 \mathbf{\Sigma}^T \otimes \mathbf{\Sigma} + \sigma_2 \text{vec}\left(\mathbf{\Sigma}\right) \text{vec}\left(\mathbf{\Sigma}\right)^H, \\ \mathbf{\Omega} = \sigma_1 \left(\mathbf{\Sigma}^T \otimes \mathbf{\Sigma}\right) \mathbf{K} + \sigma_2 \text{vec}\left(\mathbf{\Sigma}\right) \text{vec}\left(\mathbf{\Sigma}\right)^T, \end{cases} \tag{2.4}$$

with

$$\begin{aligned} \sigma_1 &= \frac{a_M p(p+1) + c_M(c_M - 2b_M)}{c_M^2}, \\ \sigma_2 &= \frac{a_M - p^2}{(c_M - p^2)^2} - \frac{a_M(p+1)}{c_M^2} + 2\frac{p(c_M - b_M)}{c_M(c_M - p^2)}, \end{aligned} \tag{2.5}$$

where a_M, c_M are given by Eq. (1.34) and $b_M = \mathbb{E}[\Psi(\sigma\mathcal{Q})\|\mathbf{g}\|^2]$.

Remark 2.2.1

- First, we recall that $\widehat{\mathbf{\Sigma}}$ tends to the M-functional $\mathbf{\Sigma}_\sigma$ that is proportional to the true scatter matrix parameter $\mathbf{\Sigma}$ as $\mathbf{\Sigma}_\sigma = \sigma^{-1}\mathbf{\Sigma}$ (Eq. (1.30)). On the other hand, $\widehat{\mathbf{\Sigma}}_{\text{GCWE}}$ tends to the true scatter matrix $\mathbf{\Sigma}$. Consequently, in order to avoid the bias we observe the quantity $\sigma\widehat{\mathbf{\Sigma}} - \widehat{\mathbf{\Sigma}}_{\text{GCWE}}$.
- Then, notice that the structure of the asymptotic covariance matrix $\mathbf{\Gamma}$ is the same as in classical asymptotic results (Eqs. (1.11) and (1.32)), but the coefficients are different. In the case of the identity matrix as covariance matrix, this very particular structure involves only three non-null elements d_1, d_2 and d_3 at the positions (i, j) which are equal to:
  - $d_1 = \sigma_1 + \sigma_2$ for $k = l = q + p(q-1)$ with $q = 1, \ldots, p$,
  - $d_2 = \sigma_1$ for $k = l = q + p(q'-1)$ with $q \neq q'$ and $p, q = 1, \ldots, p$,
  - $d_3 = \sigma_2$ for $k = q + p(q-1), l = q' + p(q'-1)$ with $q \neq q'$ and $q, q' = 1, \ldots, p$.

  Similar comment with slight modifications is valid for the pseudo-covariance matrix.
- Finally, an important remark is that scale factors σ_1 and σ_2 are smaller than the ones in the standard asymptotic regime (Eq. (1.32)). This will be more profoundly discussed and directly demonstrated on particular cases of M-estimators.

Proof sketch. We provide only a sketch of the proof, while the detailed proof of Theorem 2.2.1 is given in Appendix D. The main idea is to note that the matrix $\mathbf{\Gamma}$ can be represented as

$$\mathbf{\Gamma} = \mathbf{\Gamma}_1\left(\mathbf{\Sigma}\right) - 2\mathbf{\Gamma}_2\left(\mathbf{\Sigma}\right) + \mathbf{\Gamma}_3\left(\mathbf{\Sigma}\right),$$

where $\mathbf{\Gamma}_1\left(\mathbf{\Sigma}\right)$ and $\mathbf{\Gamma}_3\left(\mathbf{\Sigma}\right)$ are given by Eqs. (1.32) and (1.10), respectively, and the matrix $\mathbf{\Gamma}_2\left(\mathbf{\Sigma}\right)$ is the covariance matrix between an M-estimator and the corresponding GCWE. The second important step relies on a decomposition of $\mathbf{\Gamma}_2\left(\mathbf{\Sigma}\right)$:

$$\mathbf{\Gamma}_2\left(\mathbf{\Sigma}\right) = \mathbf{D}_1^{-1}\left(\mathbf{\Sigma}\right)\mathbf{B}\left(\mathbf{\Sigma}\right)\left(\mathbf{D}_2^{-1}\left(\mathbf{\Sigma}\right)\right)^H,$$

where

$$\begin{cases} \mathbf{D}_1(\mathbf{\Sigma}) = \mathbb{E}\left[d\{\mathrm{vec}\left(\Psi_1(\mathbf{\Sigma})\right)\}/d\{\mathrm{vec}\left(\mathbf{\Sigma}\right)\}\right], \\ \mathbf{B}(\mathbf{\Sigma}) = \mathbb{E}\left[\mathrm{vec}\left(\Psi_1(\mathbf{\Sigma})\right)\mathrm{vec}\left(\Psi_2(\mathbf{\Sigma})\right)^H\right], \\ \mathbf{D}_2(\mathbf{\Sigma}) = \mathbb{E}\left[d\{\mathrm{vec}\left(\Psi_2(\mathbf{\Sigma})\right)\}/d\{\mathrm{vec}\left(\mathbf{\Sigma}\right)\}\right], \end{cases} \tag{2.6}$$

with

$$\begin{cases} \Psi_1(\mathbf{\Sigma}) = \sigma u\left(\mathbf{z}^H(\sigma^{-1}\mathbf{\Sigma})^{-1}\mathbf{z}\right)\mathbf{z}\mathbf{z}^H - \mathbf{\Sigma}, \\ \Psi_2(\mathbf{\Sigma}) = \mathbf{x}\mathbf{x}^H - \mathbf{\Sigma}, \end{cases} \tag{2.7}$$

which is a generalization of a result derived in [119]. Finally, using the dependence between the practical data and their Gaussian cores one can derive elements of the matrix $\mathbf{\Gamma}_2(\mathbf{\Sigma})$ and obtain the final result. □

2.2.2 Particular cases

To promote the use of several M-estimators listed in Section 1.1.4, we derive the exact expressions for their covariance and pseudo-covariance matrices in the GCWE regime. In addition, some discussion on the values of the parameters σ_1 and σ_2 for each M-estimator is provided.

The results for each M-estimator are derived for a particular CES distribution. We chose to analyze the cases commonly met in practice and theory. For instance, the Huber's M-estimator was introduced for the CN-distributed data corrupted with a small amount of outliers. Thus, the theoretical results are derived for the CN data under the setting implemented in Example 1.1.7. Other M-estimators are studied in the MLE case, except the Tyler's M-estimator which is "distribution-free" over the class of CES distributions. However, the results for other CES distributions (model error or mismatch scenarios) in each case can be obtained analogously.

Tyler's M-estimator

As already stated in Example 1.1.6, the Tyler's M-estimator does not satisfy all Maronna's conditions and thus, does not belong to the class of M-estimators as defined in Definition 1.1.14. Consequently, it needs to be treated independently.

Theorem 2.2.2: Tyler's M-estimator [C1]

Let $\widehat{\mathbf{\Sigma}}_{\mathrm{T}}$ be defined as in Eq. (1.35). The asymptotic distribution of $\widehat{\mathbf{\Sigma}}_{\mathrm{T}} - \widehat{\mathbf{\Sigma}}_{\mathrm{GCWE}}$ is given by

$$\sqrt{n}\mathrm{vec}\left(\widehat{\mathbf{\Sigma}}_{\mathrm{T}} - \widehat{\mathbf{\Sigma}}_{\mathrm{GCWE}}\right) \xrightarrow{d} \mathcal{GCN}\left(\mathbf{0}, \mathbf{\Gamma}_{\mathrm{T}}, \mathbf{\Omega}_{\mathrm{T}}\right),$$

where $\mathbf{\Gamma}_{\mathrm{T}}$ and $\mathbf{\Omega}_{\mathrm{T}}$ are defined by

$$\begin{cases} \mathbf{\Gamma}_{\mathrm{T}} = \dfrac{1}{p}\mathbf{\Sigma}^T \otimes \mathbf{\Sigma} + \dfrac{p-1}{p^2}\mathrm{vec}\left(\mathbf{\Sigma}\right)\mathrm{vec}\left(\mathbf{\Sigma}\right)^H, \\ \mathbf{\Omega}_{\mathrm{T}} = \dfrac{1}{p}\left(\mathbf{\Sigma}^T \otimes \mathbf{\Sigma}\right)\mathbf{K} + \dfrac{p-1}{p^2}\mathrm{vec}\left(\mathbf{\Sigma}\right)\mathrm{vec}\left(\mathbf{\Sigma}\right)^T. \end{cases} \tag{2.8}$$

Remark 2.2.2

- First, notice that since the Tyler's M-estimator is distribution-free over the class of CES distributions, the results proposed in this section are valid for any CES distribution with the scatter matrix $\mathbf{\Sigma}$.
- Another remark is that the results are derived for the theoretical version of the Tyler's estimator, i.e. when the normalization $\mathrm{Tr}\left(\mathbf{\Sigma}^{-1}\widehat{\mathbf{\Sigma}}_{\mathrm{T}}\right) = p$ is employed.

- An important remark is that the elements of $\mathbf{\Gamma}_{\mathrm{T}}$ and $\mathbf{\Omega}_{\mathrm{T}}$ in Eq. (2.8) are smaller than those of $\mathbf{C}_{\mathrm{T}}$ and $\mathbf{P}_{\mathrm{T}}$ in Eq. (1.36). Moreover, the scale factors σ_1 and σ_2 are inversely proportional to the data dimension p, showing that the elements tend to 0 when p increases. This result is in agreement with the results obtained in [41] using large random matrix theory.
- Finally, note that the result for σ_1 can be obtained directly from Eq. (2.5) for $\Psi(x) = p$, which leads to $a_M = b_M = c_M = p^2$. However, the value of σ_2 can not be deduced from Eq. (2.5) and in the following we provide a separate analysis in order to proof the theorem.

Proof. The proof of Theorem 2.2.2 is given in Appendix E. □

Huber's M-estimator

In this section we analyze the Huber's M-estimator under the setup derived for CN-distributed data as in Example 1.1.7.

Corollary 2.2.3: Huber's M-estimator

Let $\widehat{\mathbf{\Sigma}}_{\mathrm{H}}$ be defined as in Eq. (1.37). The asymptotic distribution of $\widehat{\mathbf{\Sigma}}_{\mathrm{H}} - \widehat{\mathbf{\Sigma}}_{\mathrm{GCWE}}$ is defined by Theorem 2.2.1 for a_M and c_M given by Eq. (1.39) and

$$b_M = \left(p(p+1)F_{2p+4}(2\lambda) + p\lambda(1 - F_{2p+2}(2\lambda))\right)/\beta, \tag{2.9}$$

where λ and β are given by Eq. (1.39).

Remark 2.2.3

- First, one can note that when $q \to 1$ one has $b_M \to p(p+1)$. In that case, one has also $a_M, c_M \to p(p+1)$ which leads to $\sigma_1, \sigma_2 \to 0$. This result is expected since for $q \to 1$ all data is treated as CN-distributed and the Huber's M-estimator yields the GCWE.
- On the other hand, when $q \to 0$ which means that most data are assumed to be outliers, one has that $b_M \to p^2$ which together $a_M, c_M \to p^2$ leads to the results for the Tyler's M-estimator (see Remark 2.2.2).
- Finally, one has that for $q \to 1 \implies \sigma_1 \to 0$, while for $q \to 0 \implies \sigma_1 \to \dfrac{1}{p}$. Consequently, the value of $\sigma_1(q)$ for the Huber's M-estimator could be approximated by the linear model $\widetilde{\sigma}_1(q) = \frac{1}{p}(1-q)$. This will be illustrated in Section 2.3.

Proof. From $u(x) = \dfrac{1}{\beta}\min(1, \dfrac{\lambda}{x})$ one has that

$$\Psi(x) = \frac{\min(x, \lambda)}{\beta}. \tag{2.10}$$

Then, for CN-distributed data with $\|\mathbf{g}\|^2 \sim 1/2\chi^2(2p)$ one has

$$\begin{aligned} b_M &= \int_0^{+\infty} \Psi\left(\frac{y}{2}\right)\frac{y}{2}dy = \int_0^{2\lambda} \frac{y^2}{4\beta}\chi^2(2p)dy + \int_{2\lambda}^{+\infty} \frac{\lambda y}{2\beta}\chi^2(2p)dy \\ &= \frac{1}{\beta}\left(\int_0^{2\lambda} \frac{y^{p+1}\exp\left(-y/2\right)}{2^{p+2}\Gamma(p)}dy + \lambda\int_{2\lambda}^{+\infty} \frac{y^{p}\exp\left(-y/2\right)}{2^{p+1}\Gamma(p)}dy\right) \\ &= \frac{1}{\beta}\left(p(p+1)\int_0^{2\lambda} \frac{y^{p+1}\exp\left(-y/2\right)}{2^{p+2}\Gamma(p+2)}dy + p\lambda\left(1 - \int_0^{2\lambda} \frac{y^{p}\exp\left(-y/2\right)}{2^{p}\Gamma(p+1)}dy\right)\right) \\ &= \frac{1}{\beta}\left(p(p+1)F_{2p+4}(2\lambda) + p\lambda\left(1 - F_{2p+2}(2\lambda)\right)\right), \end{aligned}$$

which concludes the proof. □

Student's M-estimator

In this section we derive the results for the Student's M-estimator (Example 1.1.8) under the assumption that the data is t-distributed, i.e. $\dfrac{\mathcal{Q}}{\|\mathbf{g}\|^2} \overset{d}{=} \tau \sim \mathrm{IG}(\nu/2, \nu/2)$ (see Example 1.1.2 for more details).

Corollary 2.2.4: Student's M-estimator [J1]

Let $\widehat{\mathbf{\Sigma}}_t$ be defined as in Eq. (1.40). The asymptotic distribution of $\widehat{\mathbf{\Sigma}}_t - \widehat{\mathbf{\Sigma}}_{\mathrm{GCWE}}$ is defined by Theorem 2.2.1 for

$$\begin{cases} \sigma_1 = (p+\nu/2)^{-1}, \\ \sigma_2 = 2/\nu\ (p+1+\nu/2)(p+\nu/2)^{-1}. \end{cases} \tag{2.11}$$

Remark 2.2.4

- Notice that, once again, the parameter σ_1 is inversely proportional to the data dimension. The parameter σ_2 is equal to ϑ_2 defined in Eq. (1.33), since $b_M = c_M$. However, as we will see in Chapters 3 and 4, the value of σ_1 is of greater importance in most signal processing applications.
- One can also note that $\nu = 0$ yields $\sigma_1 = \dfrac{1}{p}$, as expected, since in that case the Student's M-estimators is equivalent to the Tyler's one. On the other side, when $\nu \to \infty$ the GCWE is produced and $\sigma_1, \sigma_2 \to 0$.

Proof. For the Student's t-distribution $\mathcal{Q} \sim pF_{2p,\nu}$ yields

$$f(\mathcal{Q}) = C_{p,\nu}\mathcal{Q}^{p-1}\left(1 + \frac{2}{\nu}\mathcal{Q}\right)^{-\frac{2p+\nu}{2}}$$

with $C_{p,\nu} = (2/\nu)^p\,\Gamma\left(p+\nu/2\right)/\Gamma\left(p\right)/\Gamma\left(\nu/2\right)$ where $\Gamma(\cdot)$ is the Gamma function. In this case the Student's M-estimator is the MLE. It can be easily shown that $\sigma = 1$ holds for all MLEs [130] and thus, one has

$$\Psi\left(\sigma\mathcal{Q}\right) = \Psi\left(\mathcal{Q}\right) = \frac{2p+\nu}{\nu+2\mathcal{Q}}\mathcal{Q}.$$

Now, one obtains

$$\begin{aligned}\mathbb{E}\left[\Psi^2(\mathcal{Q})\right] &= \mathbb{E}\left[\frac{(2p+\nu)^2}{(\nu+2\mathcal{Q})^2}\mathcal{Q}^2\right] = C_{p,\nu}\int_0^{+\infty}\frac{(2p+\nu)^2x^2}{\nu^2\left(1+\frac{2}{\nu}x\right)^2}x^{p-1}\left(1+\frac{2x}{\nu}\right)^{-\frac{2p+\nu}{2}}dx \\ &= C_{p,\nu}\frac{(2p+\nu)^2}{\nu^2}\frac{1}{C_{p+2,\nu}} = \frac{p(p+1)\left(p+\frac{\nu}{2}\right)}{p+1+\frac{\nu}{2}}\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\left[\mathcal{Q}\Psi'(\mathcal{Q})\right] &= \mathbb{E}\left[\frac{(2p+\nu)\nu}{(\nu+2\mathcal{Q})^2}\mathcal{Q}\right] = C_{p,\nu}\int_0^{+\infty}\frac{(2p+\nu)\nu x}{\nu^2(1+\frac{2}{\nu}x)^2}t_1^{p-1}\left(1+\frac{2x}{\nu}\right)^{-\frac{2p+\nu}{2}}dx \\ &= C_{p,\nu}\frac{\nu(2p+\nu)}{\nu^2}\int_0^{+\infty}x^{-1}x^{p+1}\left(1+\frac{2x}{\nu}\right)^{-\frac{2p+4+\nu}{2}}dx \\ &= C_{p,\nu}\frac{2p+\nu}{\nu}\frac{1}{C_{p+2,\nu}}\mathbb{E}\left[\mathcal{Q}^{-1}\right]\end{aligned}$$

where now $\mathcal{Q}/(p+2) \sim F_{2p+4,\nu}$ or equivalently $(p+2)/\mathcal{Q} \sim F_{\nu,2p+4}$ which gives

$$\mathbb{E}\left[\mathcal{Q}^{-1}\right] = \frac{1}{p+2}\frac{2p+4}{2p+4-2} = \frac{1}{p+1}$$

and finally

$$\mathbb{E}\left[\mathcal{Q}\Psi'(\mathcal{Q})\right] = \frac{\nu}{2}\frac{p}{\left(p+1+\frac{\nu}{2}\right)}.$$

To compute $\mathbb{E}\left[\Psi(\mathcal{Q})\|\mathbf{g}\|^2\right]$ let us remind that $\mathcal{Q} = \tau\|\mathbf{g}\|^2$ where τ and $\|\mathbf{g}\|$ are independent, with $\tau \sim \mathrm{IG}(\nu/2, \nu/2)$ and $\|\mathbf{g}\|^2 \sim (1/2)\chi^2_{2p}$. Thus, one can write

$$I = \mathbb{E}\left[\Psi(\mathcal{Q})\|\mathbf{g}\|^2\right] = C\iint_{\mathbb{R}_+^2}\frac{1}{1+\frac{2x}{\nu}y}x^{-\frac{\nu}{2}}\mathrm{e}^{-\frac{\nu}{2x}}y^{p+1}\mathrm{e}^{-y}dxdy$$

where $C = \left(\frac{2p}{\nu}+1\right)\frac{\nu}{2}^{\frac{\nu}{2}}/\left(\Gamma(\frac{\nu}{2})\Gamma(p)\right)$. The change of variable $u = \frac{2x}{\nu}y$ gives $du = \frac{2x}{\nu}dy$ and hence

$$I = C\int_0^{\infty}\frac{\nu}{2x}x^{-\frac{\nu}{2}}\mathrm{e}^{-\frac{\nu}{2x}}\left(\frac{\nu}{2x}\right)^{p+1}\int_0^{\infty}\frac{1}{1+u}u^{p+1}\mathrm{e}^{-\frac{\nu}{2x}u}dudx\,.$$

Then, using the equality

$$\int_0^{\infty}\frac{1}{1+u}u^{p+1}\mathrm{e}^{-\frac{\nu}{2x}u}du = \mathrm{e}^{\frac{\nu}{2x}}(p+1)!\Gamma\left(-1-p;\frac{\nu}{2x}\right),$$

where $\Gamma(\cdot;\cdot)$ stands for the upper incomplete Gamma function, one obtains

$$I = C'\int_0^{\infty}x^{-\frac{\nu}{2}-p-2}\Gamma\left(-1-p;\frac{\nu}{2x}\right)dx,$$

where $C' = C(p+1)!\,(\nu/2)^{p+2}$. Since

$$\Gamma\left(-1-p;\frac{\nu}{2x}\right) = \left(\frac{\nu}{2x}\right)^{-p-1}E_{p+2}\left(\frac{\nu}{2x}\right),$$

where E_{p+2} is the generalized exponential integral, one has

$$I = C''\int_0^{\infty}x^{-\frac{\nu}{2}-1}E_{p+2}\left(\frac{\nu}{2x}\right)dx,$$

where $C'' = C'\,(2/\nu)^{p+1}$ which leads to

$$\begin{aligned} I &= C'' \int_0^\infty x^{-\frac{\nu}{2}-1} \int_1^\infty \mathrm{e}^{-\frac{\nu}{2x}t} t^{-p-2} dt dx = C'' \int_1^\infty t^{-p-2} \int_0^\infty x^{-\frac{\nu}{2}-1} \mathrm{e}^{-\frac{\nu}{2x}t} dt dx \\ &= C''' \int_1^\infty t^{-p-2-\frac{\nu}{2}} dt = \frac{C'''}{p+1+\frac{\nu}{2}} \end{aligned}$$

with $C''' = C''\left(\frac{\nu}{2}\right)^{-\frac{\nu}{2}}\Gamma(\frac{\nu}{2})$. This, finally gives

$$\mathbb{E}\left[\Psi(\mathcal{Q})\|\mathbf{g}\|^2\right] = \frac{(p+\frac{\nu}{2})p(p+1)}{p+1+\frac{\nu}{2}},$$

which leads to the following values for a_M, b_M, c_M

$$a_M = b_M = c_M = \frac{p(p+1)(p+\frac{\nu}{2})}{p+1+\frac{\nu}{2}}.$$

Substituting previous quantities in Eq. (2.5), one obtains the final results for σ_1 and σ_2. □

K M-estimator

In this section we derive the results for the K M-estimator (Example 1.1.9) under the assumption that the data is K-distributed, i.e. $\dfrac{\mathcal{Q}}{\|\mathbf{g}\|^2} \overset{d}{=} \tau \sim \mathrm{Gam}(\nu, 1/\nu)$ (see Example 1.1.3 for more details). The results for other CES distributions can be obtained analogously.

Corollary 2.2.5: K M-estimator

Let $\widehat{\mathbf{\Sigma}}_K$ be defined as in Eq. (1.41). The asymptotic distribution of $\widehat{\mathbf{\Sigma}}_K - \widehat{\mathbf{\Sigma}}_{\mathrm{GCWE}}$ defined by Theorem 2.2.1 for

$$\begin{cases} a_M = c_M = C_1 \displaystyle\int_{\mathbb{R}_+} \frac{K^2_{\nu-p-1}(t)}{K_{\nu-p}(t)} t^{p+\nu+1} dt, \\ b_M = C_2 \displaystyle\iint_{\mathbb{R}^2_+} \frac{K_{\nu-p-1}(2\sqrt{\nu\tau t})}{K_{\nu-p}(2\sqrt{\nu\tau t})} t^{p+\frac{1}{2}} \tau^{\nu-\frac{1}{2}} e^{-\nu\tau} e^{-t} d\tau dt, \end{cases} \tag{2.12}$$

where $C_1 = (\Gamma(p)\Gamma(\nu)2^{p+\nu})^{-1}$ and $C_2 = \nu^{\nu+\frac{1}{2}}\,(\Gamma(m)/\Gamma(\nu))^{-1}$.

Remark 2.2.5

- Since the expressions for the parameters a_M, b_M and c_M are given by afore-stated integrals, it is difficult to draw concrete conclusions on the values of σ_1 and σ_2 directly from Eq. (2.12). It is probably possible to simplify the quantities using the basic properties of modified Bessel function, but we will skip this analysis in this work. However, we will see in Section 2.3 that the scale factors have the same interesting properties as other analyzed M-estimators.

Proof. The K M-estimator is defined for

$$\Psi(x) = \sqrt{\nu x}\frac{K_{\nu-p-1}(2\sqrt{\nu x})}{K_{\nu-p}(2\sqrt{\nu x})}$$

and in the MLE case $\mathcal{Q}$ has the following p.d.f.

$$f(\mathcal{Q}) = C_{p,\nu}\mathcal{Q}^{\frac{p+\nu}{2}-1}K_{\nu-p}\left(2\sqrt{\nu\mathcal{Q}}\right)$$

with $C_{p,\nu} = 2\nu^{(\nu+p)/2}/\Gamma(p)/\Gamma(\nu)$. Thus, one has

$$\mathbb{E}\left[\Psi^2(\mathcal{Q})\right] = C_{p,\nu}\int_0^{+\infty}\sqrt{\nu x}\frac{K^2_{\nu-p-1}(2\sqrt{\nu x})}{K_{\nu-p}(2\sqrt{\nu x})}x^{\frac{p+\nu}{2}-1}dx. \tag{2.13}$$

After the change of variable $t = 2\sqrt{\nu x}$ which gives $x = t^2/(4\nu)$, one can easily obtain the result Eq. (2.12). In addition, one can show that $c_M = a_M$ holds for all MLEs, which has been confirmed in the case of Student's M-estimator (see Corollary 2.2.4). Since the K-distribution belongs to the class of CCG distributions, one has that $\mathcal{Q} \sim \tau\|\mathbf{g}\|^2$ where $\tau \sim \text{Gam}(\nu, 1/\nu)$ and $\|\mathbf{g}\|^2 \sim 1/2\chi^2(2p)$ are independent. This directly leads to the result for b_M of Eq. (2.12).

□

Weibull M-estimator

In two previous examples we have analyzed the CCG distributions (Student's t-distribution and K-distribution) where the relationship between the observed CES data and their Gaussian cores can be described as a simple product model thanks to the texture parameter that is independent of the Gaussian cores. For non-CCG CES distributions, this is not the case and one has to find a non-linear relationship between the CES data and their Gaussian cores.

In this example we assume that the data is W-distributed with exponent s and scale $b = \left[p\Gamma\left(\frac{p+s-1}{s}\right)/\Gamma\left(\frac{p+s}{s}\right)\right]^s$, where $\mathcal{Q} \overset{d}{=} \mathcal{G}^{1/s}$, with $\mathcal{G} \sim \text{Gam}(\frac{p+s-1}{s}, b)$ (see Example 1.1.4).

Corollary 2.2.6: Weibull M-estimator

Let $\widehat{\mathbf{\Sigma}}_W$ be defined as in Eq. (1.42). The asymptotic distribution of $\widehat{\mathbf{\Sigma}}_W - \widehat{\mathbf{\Sigma}}_{\text{GCWE}}$ is given by Eq. (2.4) with

$$\begin{cases} \sigma_1 = (p-s)(1-s)\left(p^2+s(p+s-1)\right)^{-1}, \\ \sigma_2 = p(1-s)\left(p-s-2\right)\left(s\left(p^2+s(p+s-1)\right)(p+s-1)\right)^{-1}. \end{cases} \tag{2.14}$$

Remark 2.2.6

- Notice that both parameters σ_1 and σ_2 for the Weibull M-estimator are inversely proportional to the data dimension p.
- For $s = 0$ the parameters reduce to the ones of the Tyler's M-estimator. On the other hand, when $s = 1$, the Weibull M-estimator yields the SCM and the W-distribution reduces to the CN one. This means that the CES data reduce to their Gaussian cores and the Weibull M-estimator becomes equivalent to the GCWE providing $\sigma_1 = \sigma_2 = 0$. Consequently, similarly as for the Huber's M-estimator, one can propose an approximation $\widetilde{\sigma}_1(s) = \dfrac{1}{p}(1-s)$.

Proof. For the W-distribution with exponent s and scale b the p.d.f. of $\mathcal{Q}$ is given by

$$f(\mathcal{Q}) = C_{p,s,b}\mathcal{Q}^{p+s-2}\exp\left(-\frac{\mathcal{Q}^s}{b}\right)$$

$C_{p,s,b} = sb^{-(p+s-1)/s}/\Gamma((p+s-1)/s)$ and the Weibull M-estimator is defined for

$$\Psi(x) = \frac{s}{b}x^s - s + 1.$$

Consequently, one can write

$$\begin{aligned} a_M &= C_{p,s,b}\int_0^{+\infty}\left(\frac{s}{b}x^s - s + 1\right)^2 x^{p+s-2}\exp\left(-\frac{x^s}{b}\right)dx \\ &= C_{p,s,b}\int_0^{+\infty}\frac{s^2}{b^2}x^{p+3s-2}\exp\left(-\frac{x^s}{b}\right)dx \\ &+ 2C_{p,s,b}\int_0^{+\infty}\frac{s}{b}(1-s)x^{p+2s-2}\exp\left(-\frac{x^s}{b}\right)dx \\ &+ C_{p,s,b}\int_0^{+\infty}(s^2-2s+1)x^{p+s-2}\exp\left(-\frac{x^s}{b}\right)dx \end{aligned}$$

which thanks to the generalized gamma distribution gives $a_M = c_M = p^2 + s(p+s-1)$.

As mentioned above, in order to compute the parameter b_M we need first to express the modulate variate $\mathcal{Q}$ as a function of $\|\mathbf{g}\|^2$. For the W-distribution one has $\mathcal{Q} \stackrel{d}{=} \mathcal{G}^{1/s}$, where $\mathcal{G} \sim \text{Gam}(\frac{p+s-1}{s}, b)$. On the other hand, one has that $\|\mathbf{g}\|^2 \sim \text{Gam}(p,1)$. Using the following properties of the gamma distribution:

- $\sum_{i=1}^n x_i \sim \text{Gam}\left(\sum_{i=1}^n k_i, \theta\right)$, if $x_i \sim \text{Gam}\,(k_i,\theta)$,
- $cx \sim \text{Gam}\,(k, c\theta)$, if $x \sim \text{Gam}\,(k,\theta)$,

one can write

$$\mathcal{Q} \stackrel{d}{=} \frac{p\Gamma\left(\frac{p+s-1}{s}\right)}{\Gamma\left(\frac{p+s}{s}\right)}\left(\text{Gam}\left(\frac{(p-1)(1-s)}{s},1\right) + \|\mathbf{g}\|^2\right)^{1/s}.$$

Then, one obtains

$$\begin{aligned} b_M &= \mathbb{E}\left[\Psi(\mathcal{Q})\|\mathbf{g}\|^2\right] \\ &= s\mathbb{E}\left[\text{Gam}\left(\frac{(p-1)(1-s)}{s},1\right)\|\mathbf{g}\|^2\right] + s\mathbb{E}\left[\|\mathbf{g}\|^4\right] + (1-s)\mathbb{E}\left[\|\mathbf{g}\|^2\right] \\ &= s\frac{(p-1)(1-s)}{s}p + sp(p+1) - (s-1)p \\ &= (1-s)p^2 + sp(p+1) = p^2 + sp. \end{aligned}$$

Finally, after some simple derivations, one obtains the final results.

□

GG M-estimator

In this example we assume that the data is CGG-distributed with exponent s and scale $b = \left[p\Gamma(\frac{p}{s})/\Gamma(\frac{p+1}{s})\right]^s$, where $\mathcal{Q} \stackrel{d}{=} \mathcal{G}^{1/s}$, with $\mathcal{G} \sim \text{Gam}(\frac{p}{s}, b)$ (see Example 1.1.5).

Corollary 2.2.7: GG M-estimator

Let $\widehat{\mathbf{\Sigma}}_{\text{GG}}$ be defined by Eq. (1.43). The asymptotic distribution of $\widehat{\mathbf{\Sigma}}_{\text{GG}} - \widehat{\mathbf{\Sigma}}_{\text{GCWE}}$ is given by Eq. (2.4) with

$$\begin{cases} \sigma_1 = (1-s)\,(p+s)^{-1}, \\ \sigma_2 = (1-s)\,(s\,(p+s))^{-1}. \end{cases} \tag{2.15}$$

Remark 2.2.7

- The remark is the same as for the Weibull M-estimator, i.e. both parameters σ_1 and σ_2 are inversely proportional to the data dimension p.

- Note that for $s = 0$ the values of the parameters reduce to the ones of the Tyler's M-estimator. When $s = 1$, the GG M-estimator leads to the SCM and the underlying CGG distribution to the CN one. Consequently, the GG M-estimator is equal to its GCWE leading to $\sigma_1 = \sigma_2 = 0$.
- In this case, the proposed approximation $\widetilde{\sigma}_1(s) = \dfrac{1}{p}(1-s)$ is obviously more appropriate, since $\sigma_1(s) = \dfrac{1-s}{p+s} \approx \widetilde{\sigma}_1(s)$ is true even for relatively small values of p.

Proof. For the GG M-estimator one has

$$\Psi(x) = \frac{s}{b}x^s.$$

The parameter $\mathcal{Q}$ of the CCG distribution with exponent s and scale b has the p.d.f.

$$f(\mathcal{Q}) = C_{p,s,b}\mathcal{Q}^{p-1}\exp\left(-\frac{\mathcal{Q}^s}{b}\right),$$

where $C_{p,s,b} = sb^{-p/s}/\Gamma(p/s)$. Following the same steps as for the Weibull M-estimator, one can easily obtain $a_M = c_M = p^2 + sp$ and

$$\mathcal{Q} \overset{d}{=} \frac{p\Gamma\left(\frac{p}{s}\right)}{\Gamma\left(\frac{p+1}{s}\right)}\left(\mathrm{Gam}\left(\frac{p(1-s)}{s}, 1\right) + \|\mathbf{g}\|^2\right)^{1/s},$$

leading to $b_M = p^2 + sp$, which after some derivations leads to the final results. □

2.2.3 Discussion

Here are some general comments on the proposed results as well as their great interest in practice.

- First, to examine the values of the scale factors in Eq. (2.4), we discuss the values of $\mathbb{E}\left[\Psi^2(\sigma\mathcal{Q})\right]$, $\mathbb{E}\left[\Psi(\sigma\mathcal{Q})\|\mathbf{g}\|^2\right]$ and $\mathbb{E}\left[\Psi'(\sigma\mathcal{Q})\sigma\mathcal{Q}\right] + p^2$. Since $0 < \Psi(\sigma\mathcal{Q}) < K$ and $\mathbb{E}\left[\Psi(\sigma\mathcal{Q})\right] = p$, using Bhatia-Davis inequality [9], one has that $\mathrm{var}(\Psi(\sigma\mathcal{Q})) < (K-p)p$ and thus $\mathbb{E}\left[\Psi(\sigma\mathcal{Q})^2\right] < Kp$. Since K is of same magnitude as p and $K > p$, one obtains that $\mathbb{E}\left[\Psi(\sigma\mathcal{Q})^2\right]$ is of same magnitude as p^2 (for Tyler's estimator $\mathbb{E}\left[\Psi(\sigma\mathcal{Q})^2\right] = p^2$, for Student M-estimator $\mathbb{E}\left[\Psi(\sigma\mathcal{Q})^2\right] = \frac{p(p+1)(p+\frac{\nu}{2})}{p+1+\frac{\nu}{2}}$, for SCM $\mathbb{E}\left[\Psi(\sigma\mathcal{Q})^2\right] = p^2 + p\ldots$). From this, it follows that b_M is also of the same magnitude as p^2 since $p^2 \leq \mathbb{E}\left[\Psi(\sigma\mathcal{Q})\right]\mathbb{E}\left[\|\mathbf{g}\|^2\right] \leq \mathbb{E}\left[\Psi(\sigma\mathcal{Q})\|\mathbf{g}\|^2\right] \leq \sqrt{\mathbb{E}\left[\Psi(\sigma\mathcal{Q})^2\right]}\sqrt{\mathbb{E}\left[\|\mathbf{g}\|^4\right]} \leq \sqrt{(p^2+p)}\sqrt{\mathbb{E}\left[\Psi(\sigma\mathcal{Q})^2\right]}$. It is obvious that c_M is also of the same magnitude as p^2. Generally, for all widely used M-estimators, one obtains that $a_M, b_M, c_M = p^2 + \alpha p$, $\alpha > 0$ which leads to σ_1 inversely proportional to p. For σ_2, one can not provide precise information about its value, but it turns out that it is eather smaller (e.g., the Tyler's M-estimator, the Weibull M-estimator...) or unchanged (e.g., Student's M-estimator, the GG M-estimator...) comparing to the scale factor given in Eq. (1.32). **This ensures the strong "proximity" between M-estimators and the GCWE, justifying the approximation of M-estimators behavior with a Wishart distribution**.

- **The results derived in this chapter show that all M-estimators are asymptotically closer to the GCWE than to the true scatter matrix**. By "close", we mean that the asymptotic variance when centering about the GCWE is much smaller than the one when centering about the true scatter matrix. Also, this difference is more obvious when the data dimension p increases.

- An important consequence of the previous remarks is that any M-estimator's (including Tyler's one) behavior can be approximated by the GCWE one, namely by the Wishart distribution. This is of great interest in practice since **all the analytical performance of functionals of robust scatter estimators can be derived** basing on their equivalent built with the GCWE, while keeping the inherent robustness brought by M-estimators (contrary to the SCM). **To summarize, robust estimators are better approximated by Wishart distribution than by the asymptotic Gaussian distribution with the true scatter matrix as mean.**

- Another comment is that, roughly speaking, one has the following result for most robust scatter matrix estimators $\widehat{\boldsymbol{\Sigma}}$

$$\sqrt{pn}\left(\widehat{\boldsymbol{\Sigma}} - \widehat{\boldsymbol{\Sigma}}_{\text{GCWE}}\right) \xrightarrow[n\to\infty]{} \mathcal{GCN}\left(\mathbf{0}, \mathbf{D}, \mathbf{Q}\right),$$

  where $\mathbf{D}$ and $\mathbf{Q}$ are "fixed". **Thus, one has a gain in terms of convergence of** p. This is in agreement with the results obtained in [41] for a different convergence regime ($p, n \to \infty$ with p/n tending to a positive constant).

- Finally, it should be pointed out that the results can be applied to various signal processing problems. In Chapter 3 we will see how these results can be extended to the cases when the scatter matrix has a particular structure appearing in various signal processing applications. Then, in Chapter 4 we will explain why these results are useful for the threshold computation in the signal detection. Finally, the practical interest of the results in polarimetric SAR image processing will be demonstrated in Chapter 5.

2.3 Experimental analysis

2.3.1 Validation of the theoretical results

In this part, we present some simulations to validate the theoretical results for different M-estimators and data distributions with various values of scatter matrix and data dimension.

First, Figure 2.3 presents the relative error norm e_{T} between the empirical covariance matrix of $\sqrt{n}\text{vec}\left(\widehat{\boldsymbol{\Sigma}}_{\text{T}} - \widehat{\boldsymbol{\Sigma}}_{\text{GCWE}}\right)$, denoted as $\widehat{\boldsymbol{\Gamma}}_{\text{T}}^{(n)}$, and the matrix $\boldsymbol{\Gamma}_{\text{T}}$ from Theorem 2.2.2, i.e. $e_{\text{T}} = \dfrac{\|\widehat{\boldsymbol{\Gamma}}_{\text{T}}^{(n)} - \boldsymbol{\Gamma}_{\text{T}}\|}{\|\boldsymbol{\Gamma}_{\text{T}}\|}$ where $\|\cdot\|$ denotes the Frobenius norm. The parameter $\widehat{\boldsymbol{\Gamma}}_{\text{T}}^{(n)}$ is computed as the empirical mean of the quantities obtained from I Monte Carlo runs with $I_{\text{MC}} = 10000$. The plotted results are obtained for CN-distributed [1] data with dimension $p = 5$. The scatter matrix is (Hermitian) Toeplitz, i.e. $\boldsymbol{\Sigma}$ is defined by $[\boldsymbol{\Sigma}]_{i,j} = \rho^{(j-i)}$ for $i \leq j$ and $[\boldsymbol{\Sigma}]_{i,j} = [\boldsymbol{\Sigma}]_{j,i}^{*}$ for $i > j$, $i, j = 1, \ldots, p$. The correlation coefficient ρ is set to 0, i.e. the scatter matrix in this case is equal to the identity matrix. Figure 2.4 shows the corresponding quantity for the pseudo-covariance matrix, i.e. $e'_{\text{T}} = \dfrac{\|\widehat{\boldsymbol{\Omega}}_{\text{T}}^{(n)} - \boldsymbol{\Omega}_{\text{T}}\|}{\|\boldsymbol{\Omega}_{\text{T}}\|}$ where $\widehat{\boldsymbol{\Omega}}_{\text{T}}^{(n)}$ is empirical pseudo-covariance matrix of $\sqrt{n}\text{vec}\left(\widehat{\boldsymbol{\Sigma}}_{\text{T}} - \widehat{\boldsymbol{\Sigma}}_{\text{GCWE}}\right)$ and $\boldsymbol{\Omega}_{\text{T}}$ is given by Eq. (2.8) in Theorem 2.2.2. One can note that Figures 2.3 and 2.4 validate the results obtained in Theorem 2.2.2 since the errors tend to zero when the sample size n increases.

Figures 2.5 and 2.6 depict the corresponding results for the Student's M-estimator. The vertical axis of Figure 2.5 (resp. Figure 2.6) presents the relative error norm between the

1. The results are equal for all CES distributions, since the Tyler's M-estimator is distribution-free over the CES class.

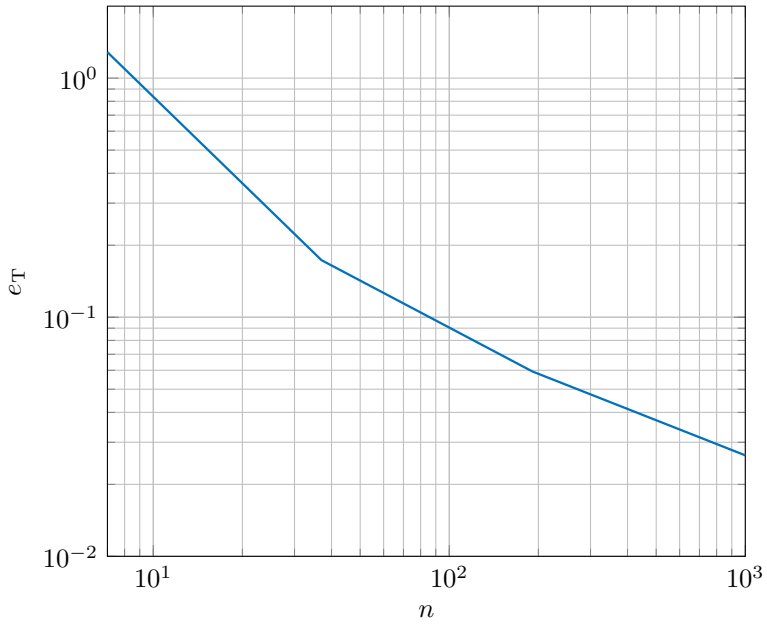


Figure 2.3 – Relative error e_{T} between the empirical covariance matrix of $\sqrt{n}\mathrm{vec}(\widehat{\mathbf{\Sigma}}_{\mathrm{T}} - \widehat{\mathbf{\Sigma}}_{\mathrm{GCWE}})$ and the corresponding theoretical results (Theorem 2.2.2) versus the sample size n; $p = 5$, $\rho = 0$.

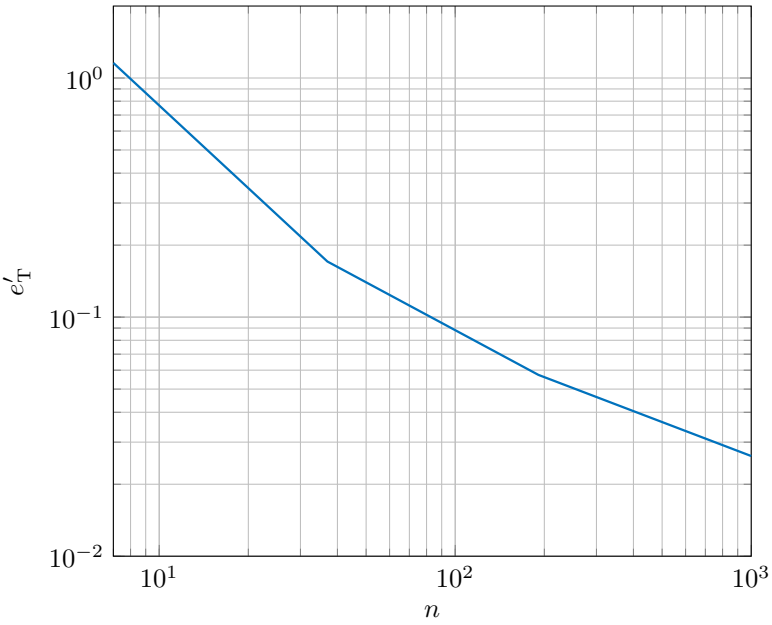


Figure 2.4 – Relative error e'_{T} between the empirical pseudo-covariance matrix of $\sqrt{n}\mathrm{vec}(\widehat{\mathbf{\Sigma}}_{\mathrm{T}} - \widehat{\mathbf{\Sigma}}_{\mathrm{GCWE}})$ and the corresponding theoretical results (Theorem 2.2.2) versus the sample size n; $p = 5$, $\rho = 0$.

empirical covariance (resp. pseudo-covariance) matrix of $\sqrt{n}(\widehat{\mathbf{\Sigma}}_t - \widehat{\mathbf{\Sigma}}_{\mathrm{GCWE}})$, denoted as $\widehat{\mathbf{\Gamma}}_{\mathrm{t}}^{(n)}$ (resp. $\widehat{\mathbf{\Omega}}_{\mathrm{t}}^{(n)}$), and the matrix $\mathbf{\Gamma}_t$ (resp. $\mathbf{\Omega}_t$) defined by Eq. (2.4) with σ_1 and σ_2 given

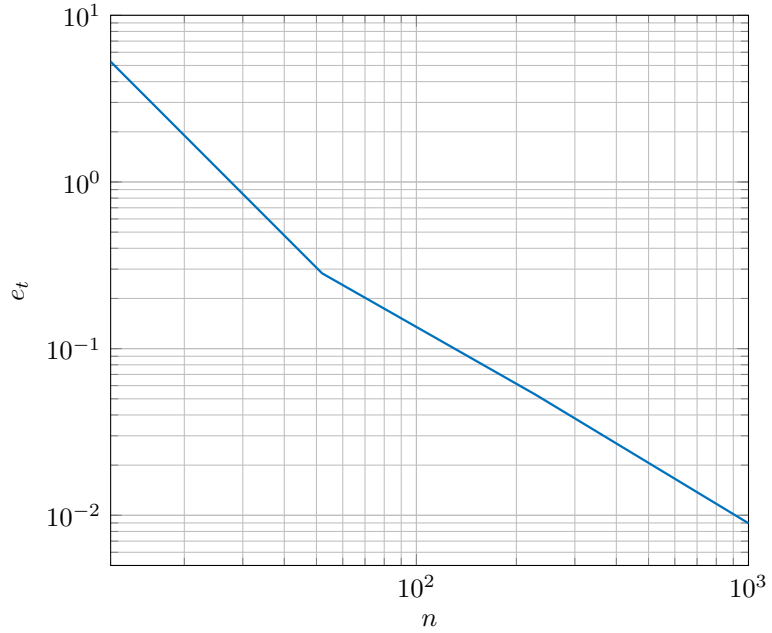


Figure 2.5 – Relative error e_t between the empirical covariance matrix of $\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_t - \widehat{\boldsymbol{\Sigma}}_{\mathrm{GCWE}})$ and the corresponding theoretical results (Corollary 2.2.4) versus the sample size n; $p = 10$, $\rho = 0.5$.

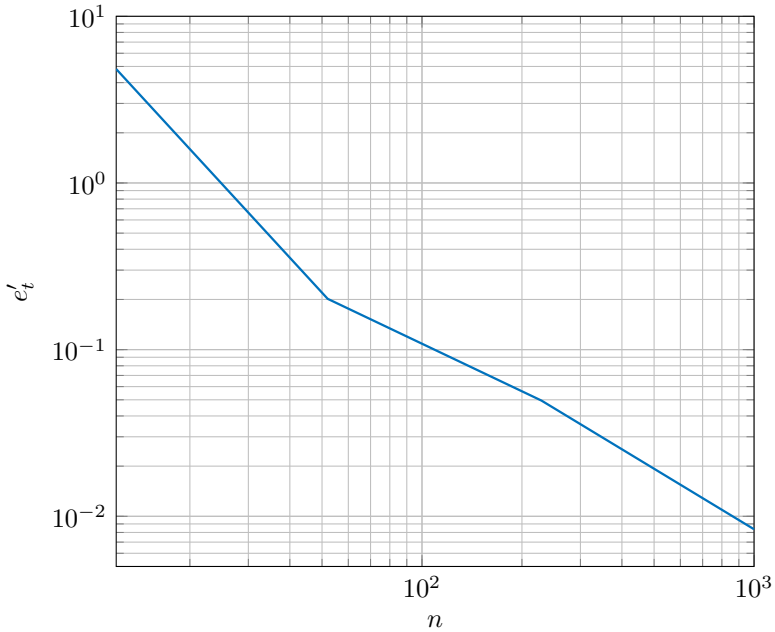


Figure 2.6 – Relative error e'_t between the empirical pseudo-covariance matrix of $\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_t - \widehat{\boldsymbol{\Sigma}}_{\mathrm{GCWE}})$ and the corresponding theoretical results (Corollary 2.2.4) versus the sample size n; $p = 10$, $\rho = 0.5$.

by Eq. (2.11) in Corollary 2.2.4, i.e. $e_t = \dfrac{\|\widehat{\boldsymbol{\Gamma}}_t^{(n)} - \boldsymbol{\Gamma}_t\|}{\|\boldsymbol{\Gamma}_t\|}$ (resp. $e'_t = \dfrac{\|\widehat{\boldsymbol{\Omega}}_t^{(n)} - \boldsymbol{\Omega}_t\|}{\|\boldsymbol{\Omega}_t\|}$). The

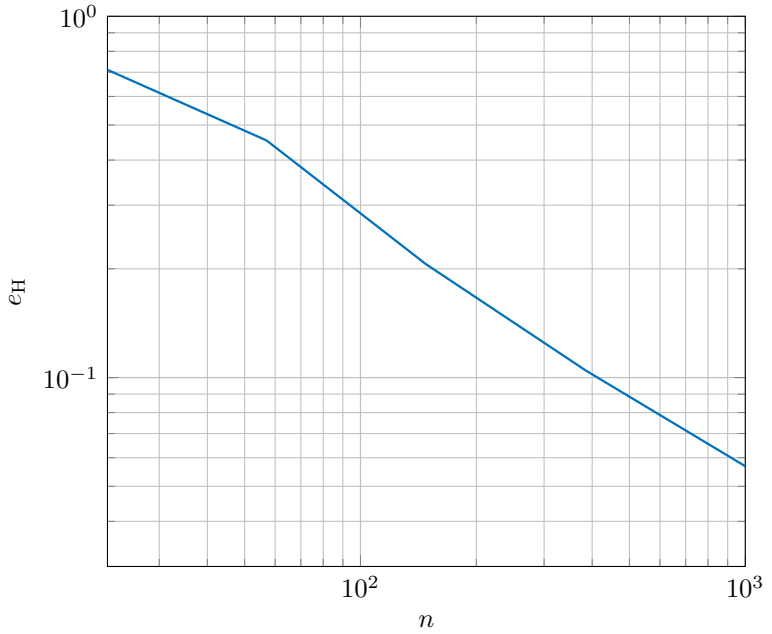


Figure 2.7 – Relative error e_{H} between the empirical covariance matrix of $\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_{\mathrm{H}} - \widehat{\boldsymbol{\Sigma}}_{\mathrm{GCWE}})$ and the corresponding theoretical results (Corollary 2.2.3) versus the sample size n; $p = 20$, $\rho = 0.5(1 + \sqrt{-1})/\sqrt{2}$.

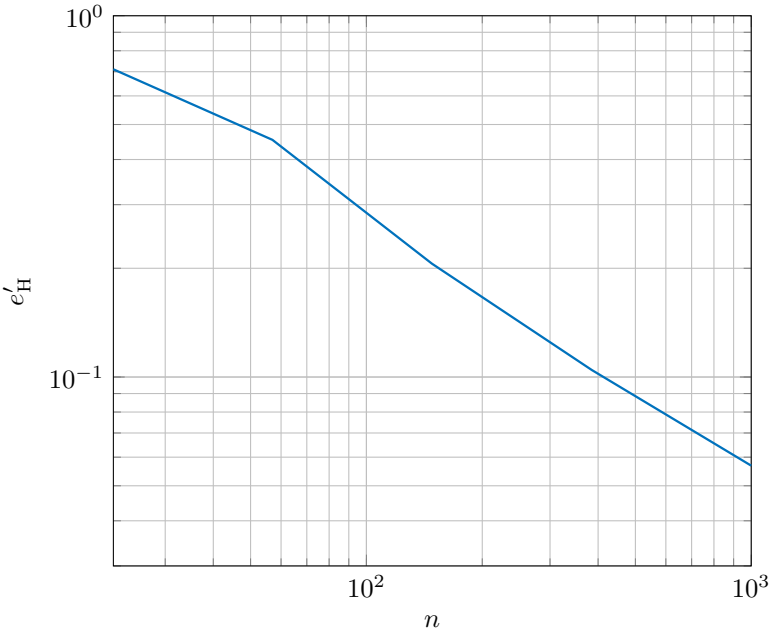


Figure 2.8 – Relative error e_{H} between the empirical pseudo-covariance matrix of $\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_{\mathrm{H}} - \widehat{\boldsymbol{\Sigma}}_{\mathrm{GCWE}})$ and the corresponding theoretical results (Corollary 2.2.3) versus the sample size n; $p = 20$, $\rho = 0.5(1 + \sqrt{-1})/\sqrt{2}$.

scatter matrix is Toeplitz with $\rho = 0.5$. The data is t-distributed with $\nu = 3$ and Student's

M-estimator is used in the MLE case, i.e., $\nu_t = \nu$. The data dimension p is set to 10. The simulations, once again, confirm the theoretical results, in this case particularly the results proposed by Corollary 2.2.4.

Finally, Figure 2.7 (resp. Figure 2.8) illustrates the results for the Huber's M-estimator. The vertical axis gives the value of the relative error norm $e_{\mathrm{H}} = \frac{\|\widehat{\mathbf{\Gamma}}_{\mathrm{H}}^{(n)} - \mathbf{\Gamma}_{\mathrm{H}}\|}{\|\mathbf{\Gamma}_{\mathrm{H}}\|}$ (resp. $e'_{\mathrm{H}} = \frac{\|\widehat{\mathbf{\Omega}}_{\mathrm{H}}^{(n)} - \mathbf{\Omega}_{\mathrm{H}}\|}{\|\mathbf{\Omega}_{\mathrm{H}}\|}$), where $\widehat{\mathbf{\Gamma}}_{\mathrm{H}}^{(n)}$ is the empirical covariance (resp. pseudo-covariance) matrix of $\sqrt{n}\mathrm{vec}\left(\widehat{\mathbf{\Sigma}}_{\mathrm{H}} - \widehat{\mathbf{\Sigma}}_{\mathrm{GCWE}}\right)$ and the matrix $\mathbf{\Gamma}_{\mathrm{H}}$ (resp. $\mathbf{\Omega}_{\mathrm{H}}$) is defined by Theorem 2.2.1 with b_M given by Eq. (2.9) in Corollary 2.2.3. The data follows the CN distribution with the Toeplitz scatter matrix with $\rho = 0.5(1 + \sqrt{-1})/\sqrt{2}$ and p is set to 20. The parameter q for Huber's M-estimator is set to 0.9 meaning that the 90% of the data is treated as CN-distributed, while the remaining 10% are assumed to be outliers.

Previous experiments validate the theoretical results of Theorem 2.2.1 for different set-ups. Let us now consider only the scale factors σ_1 and σ_2 from Eq. (2.4). Figure 2.9 plots the relative error between empirical values of the scale factors $\widehat{\sigma}_1^{(n)}$ and $\widehat{\sigma}_2^{(n)}$ and their theoretical values σ_1 and σ_2 for the K, Student's and Huber's M-estimators. The K and Student's M-estimators are analyzed in the MLE case, while the Huber's M-estimator is built with CN-distributed data. The scatter matrix is equal to identity $\mathbf{\Sigma} = \mathbf{I}$ and $p = 10$. The empirical values $\widehat{\sigma}_1^{(n)}$ and $\widehat{\sigma}_2^{(n)}$ are obtained as follows. First, the empirical covariance matrix $\widehat{\mathbf{\Gamma}}^{(n)}$ of $\sqrt{n}\mathrm{vec}\left(\widehat{\mathbf{\Sigma}} - \widehat{\mathbf{\Sigma}}_{\mathrm{GCWE}}\right)$ (where $\widehat{\mathbf{\Sigma}}$ corresponds to the used M-estimator) is computed as the empirical mean with $I_{\mathrm{MC}} = 10000$ Monte Carlo runs. Then, taking into account the structure of $\mathbf{\Gamma}$ for $\mathbf{\Sigma} = \mathbf{I}$, $\widehat{\sigma}_1^{(n)}$ and $\widehat{\sigma}_2^{(n)}$ are computed as

$$\widehat{\sigma}_1^{(n)} = \frac{1}{p^2}\mathrm{Tr}\left(\widehat{\mathbf{\Gamma}}^{(n)} - \widehat{\sigma}_2^{(n)}\mathrm{vec}\left(\mathbf{I}\right)\mathrm{vec}\left(\mathbf{I}\right)^T\right), \tag{2.16}$$

$$\widehat{\sigma}_2^{(n)} = \frac{1}{p^2 - p}\sum_{k\neq l}\left[\widehat{\mathbf{\Gamma}}^{(n)}\right]_{k,l}. \tag{2.17}$$

The shape parameter ν for K M-estimator is set to 1, while the DoF parameter for Student's M-estimator is set to 3. The parameter q for Huber's M-estimator is set to 0.95. We observe that the relative error for both scale factors goes to zero when the sample size n increases, which validates the theoretical results.

Once we have studied the scale factor for the fixed data dimension p, we can investigate the behaviour of σ_1 and σ_2 when p increases. In Figure 2.10, the scale factors for the Tyler's M-estimator in the standard asymptotic (SA) (Theorem 1.1.4) and in the GCWE regime (Theorem 2.2.2) are plotted. The empirical results for the scale factors in GCWE regime are computed using Eqs. (2.16) and (2.17). The scale factors in SA regime are computed using the same formulas where $\widehat{\mathbf{\Gamma}}^{(n)}$ is replaced by the empirical covariance matrix of $\sqrt{n}\mathrm{vec}\left(\widehat{\mathbf{\Sigma}}_{\mathrm{T}} - \mathbf{\Sigma}\right)$. The sample size n is set to 1000.

The figure is revealing in several ways. First, one can see that the empirical results match the theoretical ones. Then, one can note that the scale factors are significantly smaller in the GCWE regime than in the SA one, showing that the Tyler's M-estimator is much closer to the GCWE than to the true scatter matrix. This supports the idea that the behavior of the Tyler's M-estimator can be better described by the one of the GCWE than by its asymptotic properties. Finally, it is apparent that both scale factors in the GCWE regime tend to zero when the dimension goes to infinity, suggesting that the proposed approximation is even more accurate for high-dimensional scenarios. However, it should be noted that, even though it has a higher value in the SA regime, the second scale factor tends to zero when $p \to \infty$ both in the SA and in the GCWE regime. These conclusions are valid for all analyzed M-estimators and can be easily verified by simulation experiments.

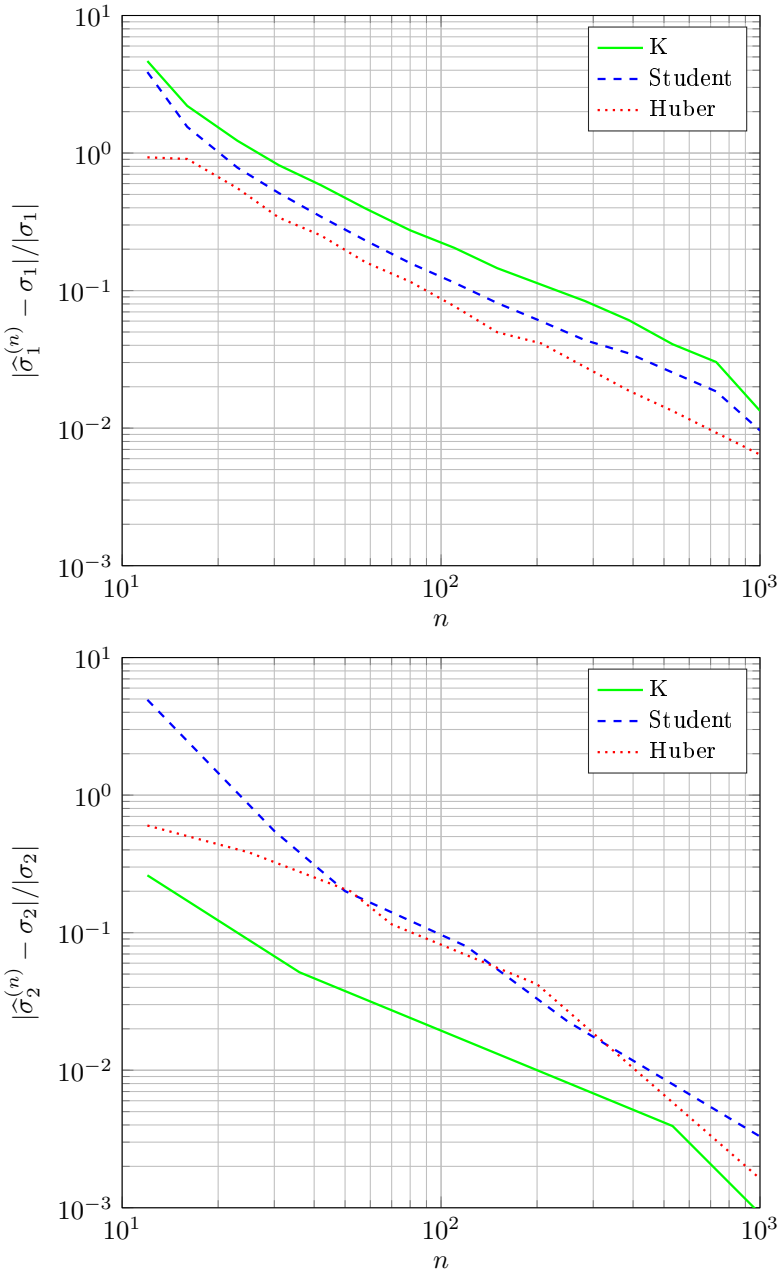


Figure 2.9 – Relative error between the estimated values of scale factors σ_1 and σ_2 and their corresponding theoretical values for the Huber's (Corollary 2.2.3), Student's (Corollary 2.2.4) and K (Corollary 2.2.5) M-estimator

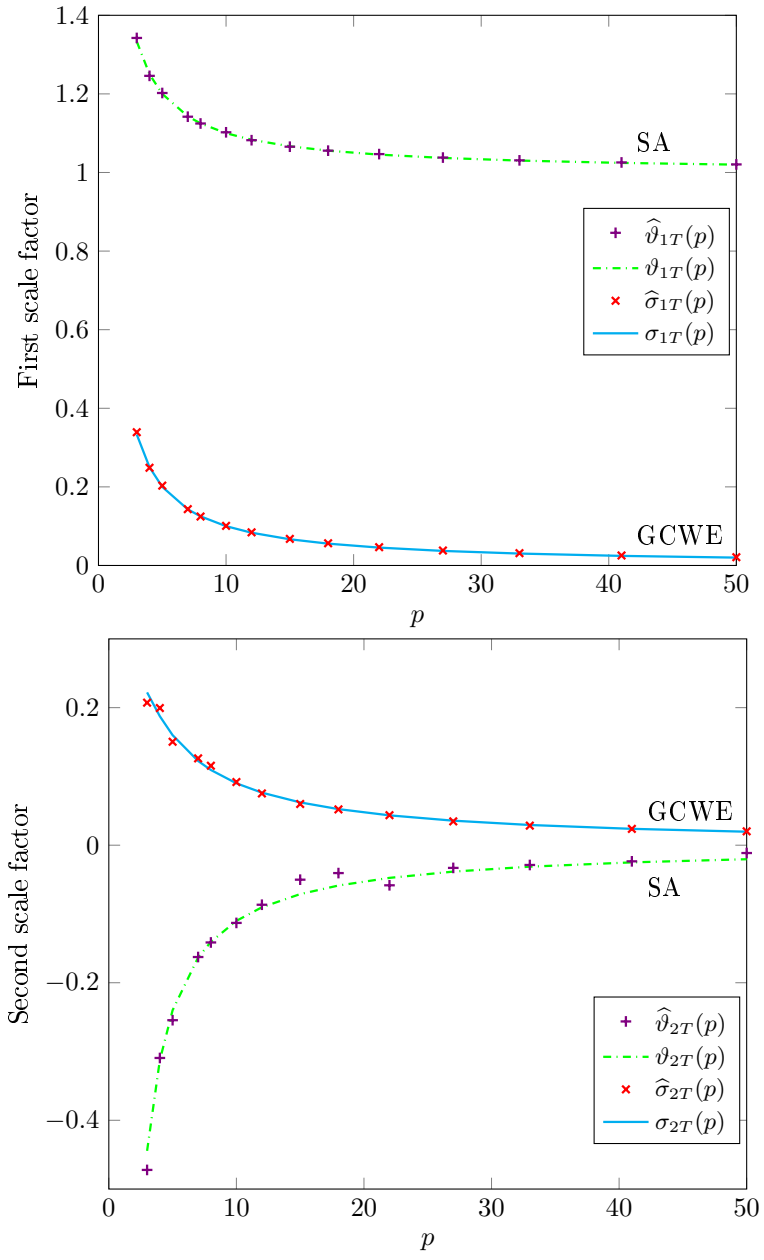


Figure 2.10 – Values of the scale factors for the Tyler's M-estimator in SA regime (Theorem 1.1.4) compared to the scale factors in GCWE regime (Theorem 2.2.2) versus data dimension p ; $n = 1000$.

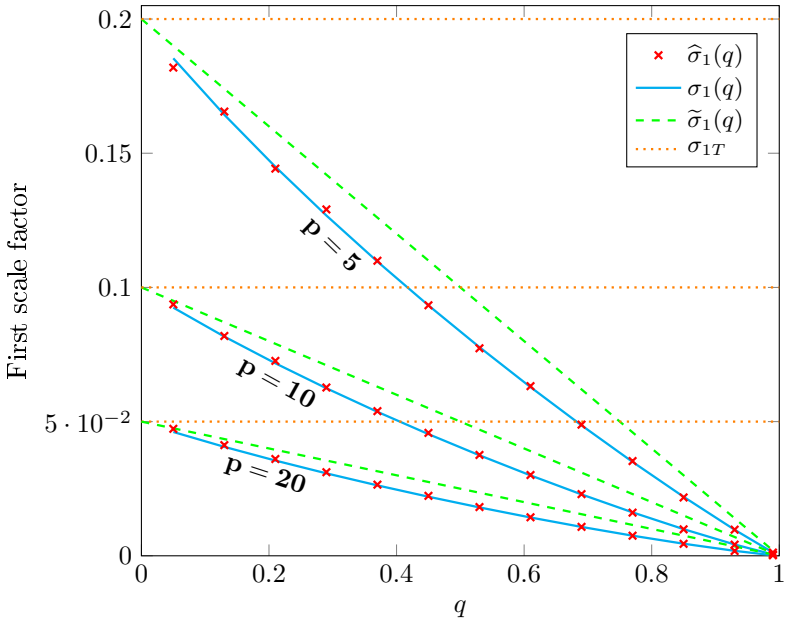


Figure 2.11 – Empirical value of the first scale factor of the Huber's M-estimator $\widehat{\sigma}_1(q)$ compared to the theoretical value $\sigma_1(q)$ (Corollary 2.2.3) and proposed approximation $\widetilde{\sigma}_1(q)$ (Remark 2.2.3) versus the parameter q; $p = 10$.

2.3.2 Study of the parameter σ_1

As it can be noted from the theoretical results for particular M-estimators, σ_1 and σ_2 do not depend only on the data dimension p, but also on the particular parameters involved in the definition of the studied M-estimator, such as ν for the Student's M-estimator, s for the Weibull M-estimator, etc. In the following, we will analyze the effect of these parameters on the values of σ_1 and σ_2.

Figure 2.11 reports the value of the first scale factor for the Huber's M-estimator for different values of the parameter q. The empirical values $\widehat{\sigma}_1(q)$ are compared to the corresponding theoretical values $\sigma_1(q)$ and to the proposed approximation $\widetilde{\sigma}_1(q) = (1-q)/p$ (Remark 2.2.3) for various values of data dimensions p. The value of the first scale factor for the Tyler's M-estimator σ_{1T} is also plotted for all values of p. The parameter $\widehat{\sigma}_1(q)$ is computed using Eq. (2.16) with $n = 1000$.

The figure highlights various interesting details on $\sigma_1(q)$. First, as well as for the Tyler's M-estimator, when the data size p increases the value of σ_1 decreases. The second important remark is that for $q \to 0 \implies \sigma_1(q) \to \sigma_{1T}$, as expected, since for $q \to 0$ the Tyler's M-estimator is produced. On the other hand, when $q \to 1$, which means that the total data is CN-distributed and the SCM is used, one has $\sigma_1(q) \to 0$. Finally, one observes that the proposed approximation $\widetilde{\sigma}_1(q)$ does not perfectly match the theoretical value of $\sigma_1(q)$, especially for small data dimension. However, for higher dimensions these quantities are very close, showing that the relationship between σ_1 and the parameter q can be well-described by $\widetilde{\sigma}_1(q)$.

Figures 2.12 and 2.13 detail the results for the first scale factor of the Student's and K M-estimators for different values of ν, respectively. The empirical values $\widehat{\sigma}_1(\nu)$ are computed using Eq. (2.16) and compared to the theoretical values $\sigma_1(\nu)$ for various values of p. The figures confirm that both analyzed M-estimators represent a kind of trade-off

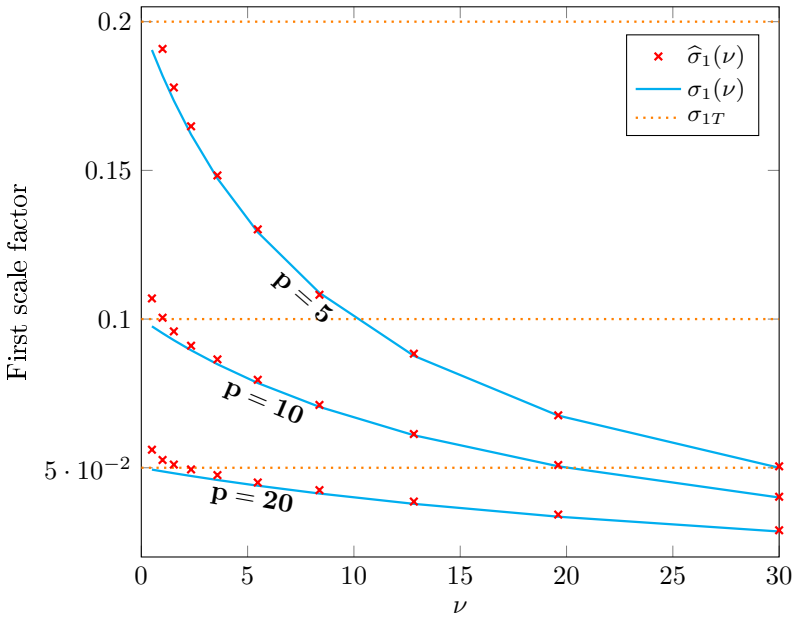


Figure 2.12 – Empirical value of the first scale factor of the Student's M-estimator $\widehat{\sigma}_1(\nu)$ compared to the theoretical value $\sigma_1(\nu)$ (Corollary 2.2.4) versus the DoF parameter ν.

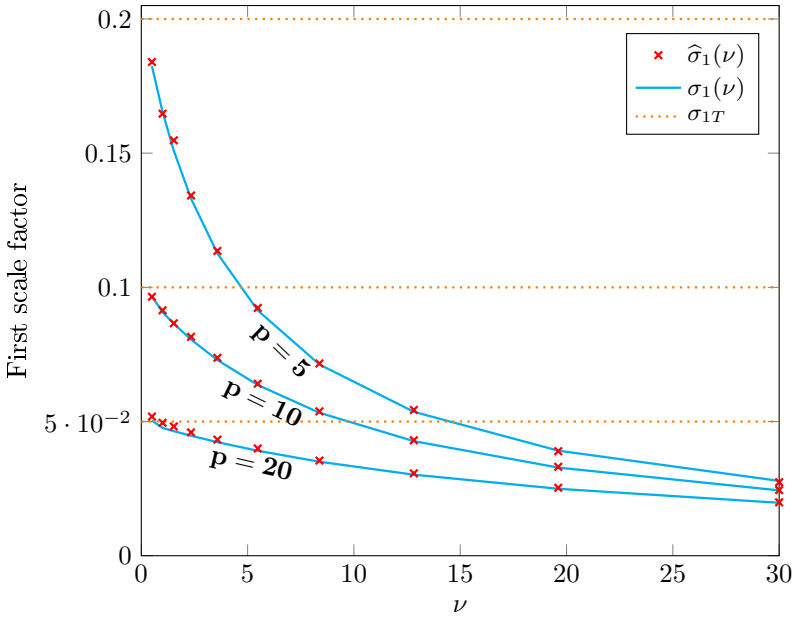


Figure 2.13 – Empirical value of the first scale factor of the K M-estimator $\widehat{\sigma}_1(\nu)$ compared to the theoretical value $\sigma_1(\nu)$ (Corollary 2.2.5) versus the shape parameter ν.

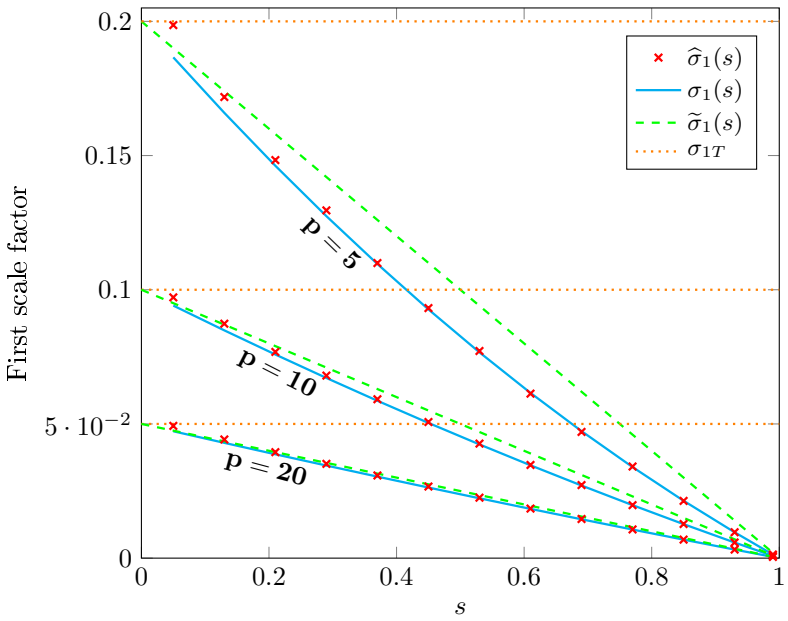


Figure 2.14 – Empirical value of the first scale factor of the Weibull M-estimator $\widehat{\sigma}_1(s)$ compared to the theoretical value $\sigma_1(s)$ (Corollary 2.2.6) and proposed approximation $\widetilde{\sigma}_1(s)$ (Remark 2.2.6) versus the parameter s; $p = 10$.

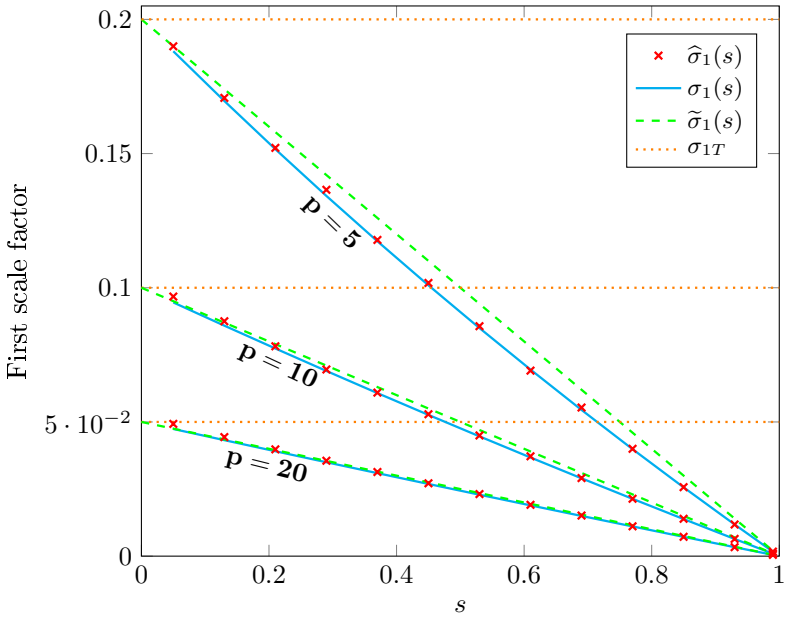


Figure 2.15 – Empirical value of the first scale factor of the GG M-estimator $\widehat{\sigma}_1(s)$ compared to the theoretical value $\sigma_1(s)$ (Corollary 2.2.7) and proposed approximation $\widetilde{\sigma}_1(s)$ (Remark 2.2.7) versus the parameter s; $p = 10$.

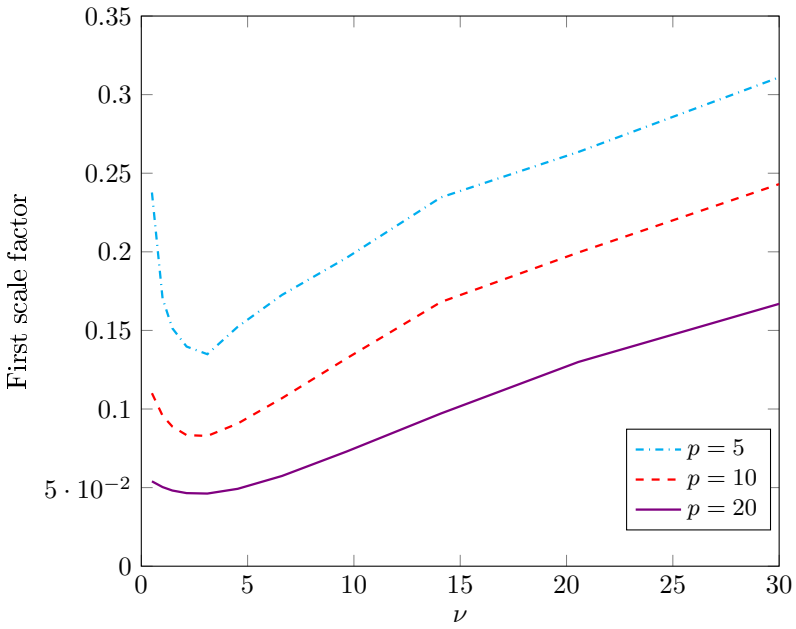


Figure 2.16 – Empirical value of the first scale factor of the K M-estimator for K-distributed data with $\nu_K = 2$ versus the shape parameter ν.

between the Tyler's M-estimator and the SCM, since $\nu \to 0 \implies \sigma_1(\nu) \to \sigma_{1T}$ and $\nu \to \infty \implies \sigma_1(\nu) \to 0$. Nevertheless, since both relationships between $\sigma_1(\nu)$ and ν are highly non-linear and $\sigma_1(\nu) \to 0$ for $\nu \to \infty$, one can not provide satisfying linear approximation for $\sigma_1(\nu) \to 0$ as was the case for the Huber's M-estimator.

In Figures 2.14 and 2.15, the results for the first scale factor of the Weibull and GG M-estimator are plotted. The horizontal scale presents the parameter s. The empirical values $\widehat{\sigma}_1(s)$ are computed using Eq. (2.16) and compared to the theoretical values $\sigma_1(s)$ (Remarks 2.2.6 and 2.2.7) for various values of p. As in the previous case, the figures confirm that both analyzed M-estimators represent a compromise between the Tyler's M-estimator and the SCM. Moreover, the values of scale factors can be well-approximated by the quantity $\widetilde{\sigma}_1(s)$, especially in the case of the GG M-estimator.

Finally, one can inspect the value of σ_1 when the M-estimator does not correspond to the MLE. Figure 2.16 demonstrates the results for $\sigma_1(\nu)$ of K M-estimator for K-distributed data with $\nu_K = 2$. The parameter ν of the K M-estimator varies between 0.5 and 30. From the figure we can see that for small values of ν, $\sigma_1(\nu)$ is close to the σ_{1T}. When $\nu > \nu_K$ we observe that $\sigma_1(\nu)$ increases which is expected, since the K M-estimator tends to the SCM and becomes less robust. The minimal value of $\sigma_1(\nu)$ is obtained when $\nu = \nu_K$, i.e. when the M-estimator corresponds to the MLE for the data distribution. An important remark here is that we have directly compared the K M-estimator to the corresponding GCWE without previous scale regulation. In fact, if we want to verify the results of Theorem 2.2.1 we need first to compute the value of the scale σ and then analyze the difference of the scaled M-estimator and the GCWE. However, as already mentioned, in some application one needs only the estimation of the shape matrix and the scale can be neglected. Nevertheless, Figure 2.16 testifies that, even when the whole information contained in the scatter matrix is important and one applies the unscaled K M-estimator, the obtained estimation in a mismatched scenario is still relatively close to the one of the GCWE for small values of ν.

This can be easily extended to Student's, Weibull and GG M-estimator.

2.4 Conclusion

In this chapter we have investigated the statistical properties of robust M-estimators. To that end, a new Gaussian-Core model has been introduced for CES distributions. Then, the GCWE is defined as the SCM built with the Gaussian cores of the observed CES data. The new approach that we have proposed consists in comparing an M-estimator to the corresponding GCWE. This allows us to derive new properties and deliver new insights into the behavior of M-estimators. The proposed approach can be summarized as follows: explaining the behavior of an "intractable" estimator $\widehat{\theta}$ by analyzing its proximity with a well-known estimator $\widehat{\theta}_1$. It has been shown that the second order statistics of M-estimators when centering around a Wishart distributed matrix are much smaller than the ones when centering around the true scatter matrix. It has also been revealed that this difference is even more meaningful for high-dimensional data. It should be stressed that these results provide a better approximation of M-estimators properties than any other analyses in the literature.

In Chapter 3, we will see how these results can be extended to the eigenvalue decomposition and principal subspace analysis of M-estimators. Several applications, depending on the scatter matrix structure, will be detailed. Then, in Chapter 4, the application of these results to the adaptive signal detection is analyzed. We will demonstrate how the results can be leveraged in the statistical analysis of adaptive robust detectors. In Chapter 5, these results will be applied to the problem of PolSAR image despeckling.

Finally, it should be noted that we have considered only complex M-estimators since they are used in signal processing applications. The results for the real case are given in Appendix C. It should be noted that the results of Theorem 2.2.1 can also be derived using the results for the real case and vector/matrix complex-to-real mapping [130]. This is briefly discussed at the end of Appendix C.

Chapter 3

EVD and PCA of the scatter matrix: new robust techniques

This chapter deals with the estimation of eigenvalue decomposition (EVD) parameters of the second order statistics. Three different cases, based on the structure of the scatter matrix, are analyzed:

- $\lambda_1 > \ldots > \lambda_p > 0$,
- $\lambda_1 = \ldots = \lambda_{j_1} > \lambda_{j_1+1} = \ldots = \lambda_{j_2} > \ldots > \lambda_{j_k} = \ldots = \lambda_p > 0$,
- $\lambda_1 > \ldots > \lambda_2 > \ldots \lambda_r > \lambda_{r+1} = \ldots = \lambda_p > 0$,

where $\lambda_1, \ldots, \lambda_p$ are eigenvalues of the scatter matrix. In the first problem, we analyze the distribution of the EVD parameters obtained from M-estimators. In the second one, we propose a new robust method in order to fuse eigenvalues and obtain a unique eigenvalue estimation for each of k blocks. Finally, in the last problem, we analyze the statistical properties of the principal subspace estimated by M-estimators. Moreover, various applications of the results are discussed.

Publications associated with this chapter:

[J2] **G. Draškovic**, A. Breloy, and F. Pascal, "On the asymptotics of Maronna's robust PCA," *IEEE Transactions on Signal Processing*, vol. 67, no. 19, pp. 4964-4975, 2019.

[C5] S. Basiri, E. Ollila, **G. Draškovi**ć, and F. Pascal, "Fusing eigenvalues," *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, United Kingdom, 2019, pp. 4968-4972.

[C9] **G. Draškovi**ć, A. Breloy et F. Pascal, "Caractérisations asymptotiques pour les composantes principales des M-estimateurs," *colloque GRETSI sur le traitement du signal et des images*, Lille, Août 2019.

3.1 Eigenvalue decomposition of M-estimators

The eigenvalue decomposition (EVD) of M-estimators is required in numerous processes. Indeed, the eigenvectors of the scatter matrix are involved in probabilistic PCA algorithms [42, 191], as well as in the derivation of robust counterparts of low rank filters or detectors [147, 69]. The eigenvalues of the scatter matrix are used in model order selection [161, 164], functions of eigenvalues are involved in various applications such as regularization parameter selection [129, 85], detection [29], and classification [16]. Hence accurately characterizing the distribution of the M-estimators EVD represents a key challenge, both from the points of view of performance analysis and optimal process design. Hereafter, we derive new asymptotic characterizations for the EVD parameters of scatter matrix M-estimators in the general context of CES-distributed samples. For the eigenvalues and eigenvectors, we derive:

- The standard Gaussian asymptotic distribution. This result is obtained by extending the analysis of [96] (for the SCM) and perturbation analysis of [101, 100] to the complex M-estimators. This asymptotic analysis provides an extension of the results obtained in [42, 14] since it gives the information about the covariance between the eigenvalues of an M-estimator. Also, contrary to the analyis done in [42, 14], all the results in this section are derived for complex data.
- The Gaussian asymptotic distribution in the GCWE regime by extending the results of Chapter 2. To do so, a central limit Theorem is established to show that the EVD parameters of M-estimators are asymptotically concentrated around their GCWE counterparts with a variance that is significantly lower than the one of the standard asymptotic regime (derived around the true expected values). These results quantify when it is acceptable to directly rely on well established results on the EVD of Wishart-distributed matrices (e.g., [123, 189]) for characterizing the EVD of M-estimators.

3.1.1 Asymptotics of M-estimators' eigenvalue decomposition

The EVD of the scatter matrix $\mathbf{\Sigma}$ is denoted as

$$\mathbf{\Sigma} \overset{\text{EVD}}{=} \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H, \quad \text{with} \quad \begin{aligned} &\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_p] \in \mathcal{U}_p^p, \\ &\mathbf{\Lambda} = \operatorname{diag}(\boldsymbol{\lambda}), \\ &\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_p]\,. \end{aligned} \tag{3.1}$$

In order to avoid ambiguity in this definition, we assume ordered eigenvalues as $\lambda_1 > \ldots > \lambda_p > 0$, and an element of each $\mathbf{u}_j$ (e.g., the first entry) for $j = 1, \ldots, n$, can be assumed to be real positive. Similarly, we define matching notations for the EVD of an M-estimator $\widehat{\mathbf{\Sigma}}$, defined as in Eq. (1.28), and its GCWE $\widehat{\mathbf{\Sigma}}_{\text{GCWE}}$, defined as in Eq. (2.2), as

$$\begin{aligned} \widehat{\mathbf{\Sigma}} &\overset{\text{EVD}}{=} \widehat{\mathbf{U}}^M \widehat{\mathbf{\Lambda}}^M \left(\widehat{\mathbf{U}}^M\right)^H, \\ \widehat{\mathbf{\Sigma}}_{\text{GCWE}} &\overset{\text{EVD}}{=} \widehat{\mathbf{U}}^{\text{GCWE}} \widehat{\mathbf{\Lambda}}^{\text{GCWE}} \left(\widehat{\mathbf{U}}^{\text{GCWE}}\right)^H. \end{aligned} \tag{3.2}$$

In the following we derive the asymptotic distributions for the quantities $\widehat{\mathbf{U}}^M$ and $\widehat{\mathbf{\Lambda}}^M$, both under the SA and the GCWE regime.

Theorem 3.1.1: Standard asymptotic [J2]

Let $\widehat{\mathbf{\Sigma}}$ be an M-estimator defined in Eq. (1.28) built from n samples drawn as $\mathbf{z} \sim \mathcal{CES}\left(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}}\right)$ and σ be the solution of Eq. (1.31). The asymptotic distribution of the EVD of $\widehat{\mathbf{\Sigma}}$, defined by Eq. (3.2), is given by

$$\begin{cases} \sqrt{n}\left(\sigma\widehat{\boldsymbol{\lambda}}^{M} - \boldsymbol{\lambda}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \vartheta_1 \mathbf{\Lambda}^2 + \vartheta_2 \boldsymbol{\lambda}\boldsymbol{\lambda}^T\right), \\ \sqrt{n}\mathbf{\Pi}_j^{\perp}\widehat{\mathbf{u}}_j^{M} \xrightarrow{d} \mathcal{CN}\left(\mathbf{0}, \mathbf{\Xi}_j\right), \end{cases} \tag{3.3}$$

where

$$\mathbf{\Xi}_j = \vartheta_1 \lambda_j \left(\mathbf{U}\mathbf{\Lambda}(\lambda_j \mathbf{I} - \mathbf{\Lambda})^{+}\right)^2 \mathbf{U}^H \tag{3.4}$$

with $\mathbf{\Pi}_j^{\perp} = \mathbf{I} - \mathbf{u}_j\mathbf{u}_j^H$, $j = 1, \ldots, p$ and ϑ_1, ϑ_2 given by Eq. (1.33).

Remark 3.1.1

- The results given in Theorem 3.1.1 are interesting since, besides the variance of each eigenvalue, they provide the correlation between them. Note that for a Wishart-distributed matrix this correlation is equal to zero, as shown in [96] for real case. Conversely, Theorem 3.1.1 shows that the eigenvalues of an M-estimator are asymptotically correlated, as stated in [42] (but not explicitly characterized). This correlation depends on the second scale parameter ϑ_2.
- Concerning the eigenvectors, note that the covariance depends only on ϑ_1 since $\mathbf{u}_j$ is scale invariant w.r.t. to the covariance matrix (see [117] for more details).

Proof. To prove the theorem we will rely on the basic results obtained in the following lemma [13][Proposition 6.2] (and e.g. [117] for the formulation in the complex case).

Lemma 3.1.1. Let $(\widehat{\mathbf{z}}_1, \ldots, \widehat{\mathbf{z}}_n)$ be a sequence of complex random vectors $\widehat{\mathbf{z}}$ and $\mathbf{z}$ a compatible fixed vector. Assume that $\sqrt{n}\left(\widehat{\mathbf{z}} - \mathbf{z}\right) \xrightarrow{d} \mathcal{GCN}\left(\mathbf{0}, \mathbf{C}, \mathbf{P}\right)$. Let $\xi\left(\mathbf{z}\right)$ be a vector function of a vector $\mathbf{z}$ with first and a second derivatives existing in $\mathbf{z}$. Then

$$\sqrt{n}\left(\xi\left(\widehat{\mathbf{z}}\right) - \xi\left(\mathbf{z}\right)\right) \xrightarrow{d} \mathcal{GCN}\left(\mathbf{0}, \xi'\left(\mathbf{z}\right)\mathbf{C}\xi'\left(\mathbf{z}\right)^H, \xi'\left(\mathbf{z}\right)\mathbf{P}\xi'\left(\mathbf{z}\right)^T\right) \tag{3.5}$$

where

$$\xi'\left(\mathbf{z}\right) = \frac{\partial \xi\left(\mathbf{z}\right)}{\partial \mathbf{z}} \tag{3.6}$$

is a matrix derivative.

Proof. Thanks to the Delta method, one has a first order approximation

$$\xi\left(\widehat{\mathbf{z}}\right) \simeq \xi\left(\mathbf{z}\right) + \xi'\left(\mathbf{z}\right)\left(\widehat{\mathbf{z}} - \mathbf{z}\right), \tag{3.7}$$

where $\xi'\left(\mathbf{z}\right) = \partial\xi\left(\mathbf{z}\right)/\partial\mathbf{z}$. Then, one has

$$\begin{aligned} \mathbb{E}\left[n\left(\xi\left(\widehat{\mathbf{z}}\right) - \xi\left(\mathbf{z}\right)\right)\left(\xi\left(\widehat{\mathbf{z}}\right) - \xi\left(\mathbf{z}\right)\right)^H\right] &= \xi'\left(\mathbf{z}\right)\mathbb{E}\left[n\left(\widehat{\mathbf{z}} - \mathbf{z}\right)\left(\widehat{\mathbf{z}} - \mathbf{z}\right)^H\right]\xi'\left(\mathbf{z}\right)^H \\ &\xrightarrow[n\to\infty]{} \xi'\left(\mathbf{z}\right)\mathbf{C}\xi'\left(\mathbf{z}\right)^H. \end{aligned} \tag{3.8}$$

Analogously, for the pseudo-covariance matrix, one has

$$\begin{aligned} \mathbb{E}\left[n\left(\xi\left(\widehat{\mathbf{z}}\right) - \xi\left(\mathbf{z}\right)\right)\left(\xi\left(\widehat{\mathbf{z}}\right) - \xi\left(\mathbf{z}\right)\right)^T\right] &= \xi'\left(\mathbf{z}\right)\mathbb{E}\left[n\left(\widehat{\mathbf{z}} - \mathbf{z}\right)\left(\widehat{\mathbf{z}} - \mathbf{z}\right)^T\right]\xi'\left(\mathbf{z}\right)^T \\ &\xrightarrow[n\to\infty]{} \xi'\left(\mathbf{z}\right)\mathbf{P}\xi'\left(\mathbf{z}\right)^T, \end{aligned} \tag{3.9}$$

which concludes the proof. □

In order to obtain the derivatives of $\boldsymbol{\lambda}$ and $\mathbf{u}_j$ w.r.t. $\mathrm{vec}\,(\boldsymbol{\Sigma})$, we differentiate $\boldsymbol{\Sigma}\mathbf{u}_j = \lambda_j \mathbf{u}_j$

$$d\boldsymbol{\Sigma}\mathbf{u}_j + \boldsymbol{\Sigma} d\mathbf{u}_j = d\lambda_j \mathbf{u}_j + \lambda_j d\mathbf{u}_j. \tag{3.10}$$

Multiplying each side of the last equation by $\mathbf{u}_j^H$, one has

$$d\lambda_j = \mathbf{u}_j^H \left(d\boldsymbol{\Sigma}\right) \mathbf{u}_j$$

since $\mathbf{u}_j^H \boldsymbol{\Sigma} = \lambda_j \mathbf{u}_j^H$ and $\mathbf{u}_j^H \mathbf{u}_j = 1$. Thus,

$$\frac{\partial \lambda_j}{\partial \mathrm{vec}\,(\boldsymbol{\Sigma})} = \mathbf{u}_j^T \otimes \mathbf{u}_j^H.$$

If $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)$, then one has

$$\frac{\partial \boldsymbol{\lambda}}{\partial \mathrm{vec}\,(\boldsymbol{\Sigma})} = \mathbf{E}^T \left(\mathbf{U}^T \otimes \mathbf{U}^H\right)$$

with $\mathbf{E} = (\mathbf{e}_1 \otimes \mathbf{e}_1 \ldots \mathbf{e}_p \otimes \mathbf{e}_p)$ where $\mathbf{e}_j$, $j = 1, \ldots, p$ are unit vectors. Further, combining the statement given in Lemma 3.1.1 with Eq. (1.32), one obtains

$$\begin{aligned} & \mathbf{E}^T \left(\mathbf{U}^T \otimes \mathbf{U}^H\right) \left(\vartheta_1 \left(\boldsymbol{\Sigma}^T \otimes \boldsymbol{\Sigma}\right)\right) \left(\mathbf{U}^* \otimes \mathbf{U}\right) \mathbf{E} \\ + \quad & \mathbf{E}^T \left(\mathbf{U}^T \otimes \mathbf{U}^H\right) \vartheta_2 \mathrm{vec}\,(\boldsymbol{\Sigma})\, \mathrm{vec}\,(\boldsymbol{\Sigma})^H \left(\mathbf{U}^* \otimes \mathbf{U}\right) \mathbf{E} \\ = \quad & \vartheta_1 \mathbf{E}^T \left(\boldsymbol{\Lambda}^T \otimes \boldsymbol{\Lambda}\right) \mathbf{E} + \vartheta_2 \mathbf{E}^T \left(\mathrm{vec}\,(\boldsymbol{\Lambda})\, \mathrm{vec}\,(\boldsymbol{\Lambda})^H\right) \mathbf{E} = \vartheta_1 \boldsymbol{\Lambda}^2 + \vartheta_2 \boldsymbol{\lambda}\boldsymbol{\lambda}^T. \end{aligned}$$

Note that in this equality $\boldsymbol{\Sigma}$ figures instead $\boldsymbol{\Sigma}_\sigma$, since we analyze the distribution of $\sigma\widehat{\lambda}^M$ instead of $\widehat{\lambda}^M$. Note also that, since the eigenvalues are real one obtains the same result using the expression for the pseudo-covariance matrix.

In order to obtain the results for eigenvectors, we will multiply Eq. (3.10) by $\mathbf{u}_k^H$, $k \neq j$. Thus, one obtains

$$\mathbf{u}_k^H \left(d\boldsymbol{\Sigma}\right) \mathbf{u}_j = \left(\lambda_j - \lambda_k\right) \mathbf{u}_k^H d\mathbf{u}_j$$

as $\mathbf{u}_k^H \mathbf{u}_j = 0$. Following the same steps as in [96] (done for the real case), it is easy to show that

$$d\mathbf{u}_j = \sum_{j \neq k} \left(\lambda_j - \lambda_k\right)^{-1} \mathbf{u}_k \mathbf{u}_k^H \left(d\boldsymbol{\Sigma}\right) \mathbf{u}_j + \mathbf{u}_j \mathbf{u}_j^H d\mathbf{u}_j.$$

In fact, the last element in the previous equality is omitted in the real case since $\mathbf{u}_j^T d\mathbf{u}_j = 0$ (from $\mathbf{u}_j^T \mathbf{u}_j = 1$). However, in the complex case $\mathbf{u}_j^H d\mathbf{u}_j \neq 0$, as from $\mathbf{u}_j^H \mathbf{u}_j = 1$ one has $\mathbf{u}_j^H d\mathbf{u}_j + d\mathbf{u}_j^H \mathbf{u}_j = 0$ and it is obvious that $\mathbf{u}_j^H d\mathbf{u}_j \neq d\mathbf{u}_j^H \mathbf{u}_j$. In some works, as in [114], the authors use different normalizations for eigenvectors that imply $\mathbf{u}_j^H d\mathbf{u}_j = 0$ and in those circumstances the results correspond to the ones in the real case. In the general (more common) case, one obtains

$$\left(\mathbf{I} - \mathbf{u}_j \mathbf{u}_j^H\right) d\mathbf{u}_j = \left(\mathbf{u}_j^T \otimes \mathbf{U} \left(\lambda_j \mathbf{I} - \boldsymbol{\Lambda}\right)^+ \mathbf{U}^H\right) d\boldsymbol{\Sigma},$$

which actually gives the projection of the derivative onto the subspace orthogonal to the one of the eigenvector. Now, employing Eq. (3.5) with previous derivatives and since

$$\begin{aligned} \left(\mathbf{u}_j^T \otimes \mathbf{U} \left(\lambda_j \mathbf{I} - \boldsymbol{\Lambda}\right)^+ \mathbf{U}^H\right) \mathbf{K} &= \mathbf{U} \left(\lambda_j \mathbf{I} - \boldsymbol{\Lambda}\right)^+ \mathbf{U}^H \otimes \mathbf{u}_j^T, \\ \left(\lambda_j \mathbf{I} - \boldsymbol{\Lambda}\right)^+ \mathbf{e}_j &= \mathbf{0}, \\ \left[\mathbf{u}_j^T \otimes \mathbf{U} (\lambda_j \mathbf{I} - \boldsymbol{\Lambda})^+ \mathbf{U}^H\right] \mathrm{vec}\,(\boldsymbol{\Sigma}) &= \mathbf{0}, \end{aligned}$$

one obtains the final results. Note that $\mathcal{GCN}$ becomes $\mathcal{CN}$ since the pseudo-covariance matrix is equal to zero. □

Theorem 3.1.2: Asymptotic GCWE [J2]

Let $\widehat{\boldsymbol{\Sigma}}$ be an M-estimator as in Eq. (1.28) built from n samples drawn as $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \boldsymbol{\Sigma}, g_{\mathbf{z}})$. Let $\widehat{\boldsymbol{\Sigma}}_{\text{GCWE}}$ be its GCWE (Definition 2.1.2) and let σ be the solution of Eq. (1.31). The asymptotic distribution of the difference between the EVD parameters of $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\Sigma}}_{\text{GCWE}}$ is given by

$$\begin{cases} \sqrt{n}\left(\sigma\widehat{\boldsymbol{\lambda}}^{M} - \widehat{\boldsymbol{\lambda}}^{\text{GCWE}}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \sigma_1\boldsymbol{\Lambda}^2 + \sigma_2\boldsymbol{\lambda}\boldsymbol{\lambda}^T\right), \\ \sqrt{n}\boldsymbol{\Pi}_j^{\perp}\left(\widehat{\mathbf{u}}_j^{M} - \widehat{\mathbf{u}}_j^{\text{GCWE}}\right) \xrightarrow{d} \mathcal{CN}\left(\mathbf{0}, \sigma_1/\vartheta_1\boldsymbol{\Xi}_j\right), \end{cases} \tag{3.11}$$

with $\boldsymbol{\Xi}_j$ and σ_1, σ_2 given by Eqs. (3.4) and (2.5), respectively.

Remark 3.1.2

- Theorem 3.1.2 characterizes the asymptotic variance of the EVD of an M-estimator compared to the one of its GCWE. It shows that their covariance structure is the same as the one in the SA regime, and differs only through the variance scales (σ_1, σ_2) (instead of $(\vartheta_1, \vartheta_2)$). As noted in Chapter 2, the total variance captured by the GCWE factors is much smaller than the standard one.
- This result supports the idea that an underlying Wishart distribution can offer a better approximation for characterizing the distribution of the M-estimator's EVD. This approximation allows us to rely on the well established results in [123, 189] and offers a thinner analysis compared to the asymptotic Gaussian results as we will se in Section 3.1.3.

Proof. Rewriting the left-hand side of Eq. (3.3)

$$\begin{aligned} \sqrt{n}\left(\sigma\widehat{\boldsymbol{\lambda}}^{M} - \widehat{\boldsymbol{\lambda}}^{\text{GCWE}}\right) &= \sqrt{n}\left(\sigma\widehat{\boldsymbol{\lambda}}^{M} - \boldsymbol{\lambda} - \widehat{\boldsymbol{\lambda}}^{\text{GCWE}} + \boldsymbol{\lambda}\right) \\ &= \sqrt{n}\left(\left(\sigma\widehat{\boldsymbol{\lambda}}^{M} - \boldsymbol{\lambda}\right) - \left(\widehat{\boldsymbol{\lambda}}^{\text{GCWE}} - \boldsymbol{\lambda}\right)\right). \end{aligned}$$

Then,

$$\begin{aligned} \text{var}_n\left(\sigma\widehat{\boldsymbol{\lambda}}^{M} - \widehat{\boldsymbol{\lambda}}^{\text{GCWE}}\right) &= \mathbb{E}\left[n\left(\sigma\widehat{\boldsymbol{\lambda}}^{M} - \widehat{\boldsymbol{\lambda}}^{\text{GCWE}}\right)\left(\sigma\widehat{\boldsymbol{\lambda}}^{M} - \widehat{\boldsymbol{\lambda}}^{\text{GCWE}}\right)^T\right] \\ &= \text{var}_n\left(\sigma\widehat{\boldsymbol{\lambda}}^{M}\right) - 2\text{cov}_n\left(\sigma\widehat{\boldsymbol{\lambda}}^{M}, \widehat{\boldsymbol{\lambda}}^{\text{GCWE}}\right) + \text{var}_n\left(\widehat{\boldsymbol{\lambda}}^{\text{GCWE}}\right). \end{aligned}$$

From Eq. (3.3), one has

$$\begin{aligned} \text{var}_n\left(\sigma\widehat{\boldsymbol{\lambda}}^{M}\right) &\xrightarrow[n\to+\infty]{} \vartheta_1\boldsymbol{\Lambda}^2 + \vartheta_2\boldsymbol{\lambda}\boldsymbol{\lambda}^T \quad \text{and} \\ \text{var}_n\left(\widehat{\boldsymbol{\lambda}}^{\text{GCWE}}\right) &\xrightarrow[n\to+\infty]{} \boldsymbol{\Lambda}^2. \end{aligned}$$

Then, it remains only to derive the expression for

$$\text{cov}_n\left(\sigma\widehat{\boldsymbol{\lambda}}^{M}, \widehat{\boldsymbol{\lambda}}^{\text{GCWE}}\right) = \mathbb{E}\left[n\left(\sigma\widehat{\boldsymbol{\lambda}}^{M} - \boldsymbol{\lambda}\right)\left(\widehat{\boldsymbol{\lambda}}^{\text{GCWE}} - \boldsymbol{\lambda}\right)^H\right].$$

Using the Delta method similarly as in Lemma 3.1.1, one can show that

$$\text{cov}_n\left(\sigma\widehat{\boldsymbol{\lambda}}^{M}, \widehat{\boldsymbol{\lambda}}^{\text{GCWE}}\right) \xrightarrow[n\to+\infty]{} \frac{\partial\boldsymbol{\lambda}}{\partial\text{vec}\left(\boldsymbol{\Sigma}\right)}\boldsymbol{\Gamma}_2\left(\frac{\partial\boldsymbol{\lambda}}{\partial\text{vec}\left(\boldsymbol{\Sigma}\right)}\right)^H,$$

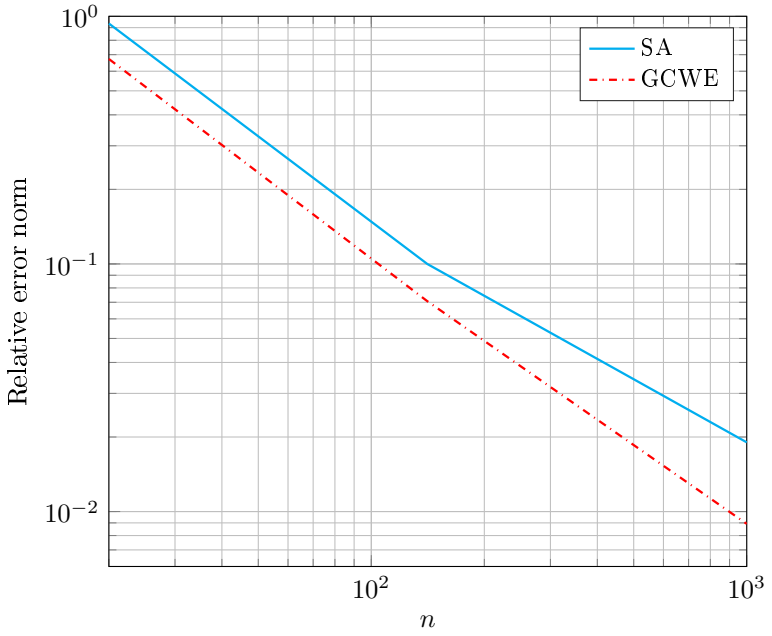


Figure 3.1 – Relative error norm between the empirical covariance matrix of $\sqrt{n}\left(\widehat{\boldsymbol{\lambda}}^t - \boldsymbol{\lambda}\right)$ and its asymptotic value (blue curve) and the corresponding quantity for $\sqrt{n}\left(\widehat{\boldsymbol{\lambda}}^t - \widehat{\boldsymbol{\lambda}}^{\mathrm{GCWE}}\right)$ (red curve); t-distributed data with $\nu = 2$, $p = 10$, $\rho = 0.9(1 + \sqrt{-1})/\sqrt{2}$.

where $\boldsymbol{\Gamma}_2$ is equal to

$$\boldsymbol{\Gamma}_2 = \gamma_1 \boldsymbol{\Sigma}^T \otimes \boldsymbol{\Sigma} + \gamma_2 \mathrm{vec}\left(\boldsymbol{\Sigma}\right) \mathrm{vec}\left(\boldsymbol{\Sigma}\right)^H \tag{3.12}$$

with γ_1 and γ_2 given by Eq. (D.7). Repeating the same steps as in Eq. (3.11), one shows that the right-hand side of the right-hand side of Eq. (3.12) becomes

$$\gamma_1 \boldsymbol{\Lambda}^2 + \gamma_2 \boldsymbol{\lambda}\boldsymbol{\lambda}^T$$

which, since $\sigma_1 = \vartheta_1 - 2\gamma_1 + 1$ and $\sigma_2 = \vartheta_2 - 2\gamma_2$, leads to the final results.

The results for the eigenvectors can be obtained following the same procedure as for the eigenvalues. □

3.1.2 Experimental validation

In this part, we present some experiments in order to validate the previous theoretical results.

Figure 3.1 plots the relative error norm [1] between the empirical covariance matrix of $\sqrt{n}\left(\widehat{\boldsymbol{\lambda}}^t - \boldsymbol{\lambda}\right)$ and its asymptotic value given by Eq. (3.3) and the relative error norm between the empirical covariance matrix of $\sqrt{n}\left(\widehat{\boldsymbol{\lambda}}^t - \widehat{\boldsymbol{\lambda}}^{\mathrm{GCWE}}\right)$ and its asymptotic value given by Eq. (3.11), where $\widehat{\boldsymbol{\lambda}}^t$ is the vector composed of the ordered eigenvalues of the Student's M-estimator. The plotted results are obtained for t-distributed data with $\nu = 2$ and dimension $p = 10$. The scatter matrix is Toeplitz with $\rho = 0.9(1 + \sqrt{-1})/\sqrt{2}$. The Student's M-estimator is built with $\nu_t = \nu$.

1. The relative error norm is defined as in Section 2.3.

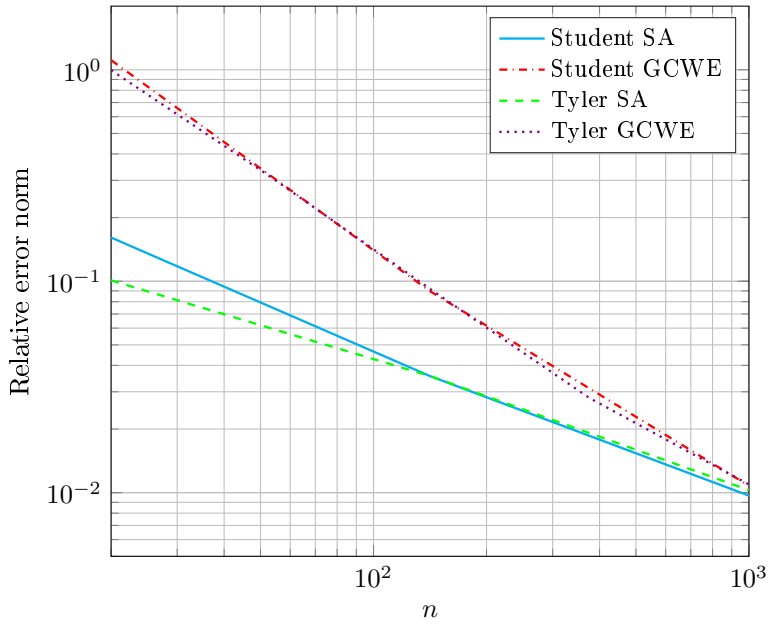


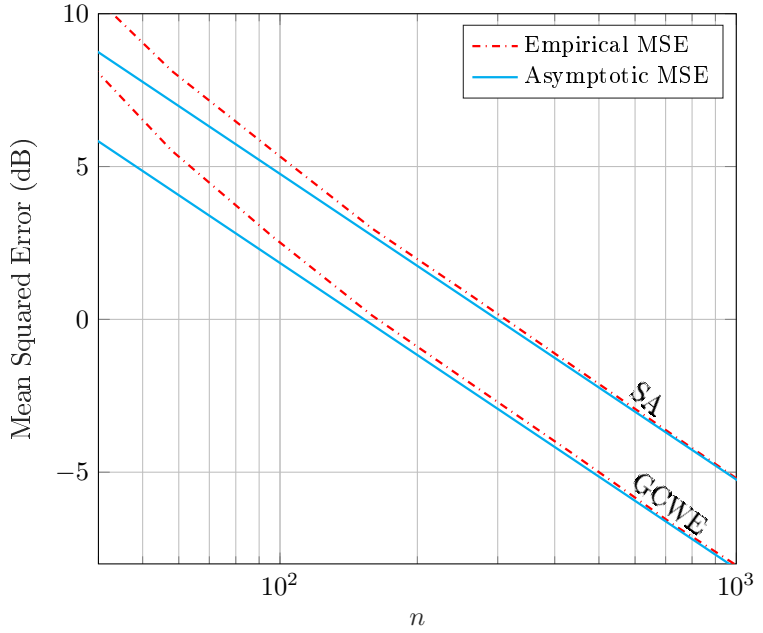
Figure 3.2 – Relative error norm between the empirical covariance matrix of $\sqrt{n}\left(\widehat{\mathbf{u}}_1 - \mathbf{u}_1\right)$ and its asymptotic value (blue curve) and the corresponding quantity for $\sqrt{n}\left(\widehat{\mathbf{u}}_1 - \mathbf{u}_1^{\text{GCWE}}\right)$ (red curve); t-distributed data with $\nu = 2$, $p = 10$, $\rho = 0.9(1 + \sqrt{-1})/\sqrt{2}$.

Figure 3.2 plots the relative error norm between the empirical covariance matrix of $\sqrt{n}\left(\widehat{\mathbf{u}}_1^M - \mathbf{u}_1\right)$ and its asymptotic value given by Eq. (3.3) and the relative error norm between the empirical covariance matrix of $\sqrt{n}\left(\widehat{\mathbf{u}}_1^M - \widehat{\mathbf{u}}_1^{\text{GCWE}}\right)$ and its asymptotic value given by Eq. (3.11), where $\widehat{\mathbf{u}}_1^M$ is the eigenvector that corresponds to the first eigenvalue (eigenvalue with highest value) of an M-estimator. The plotted results are obtained for the Student's and Tyler's M-estimator and t-distributed data with dimension $p = 10$ and $\nu = 2$.

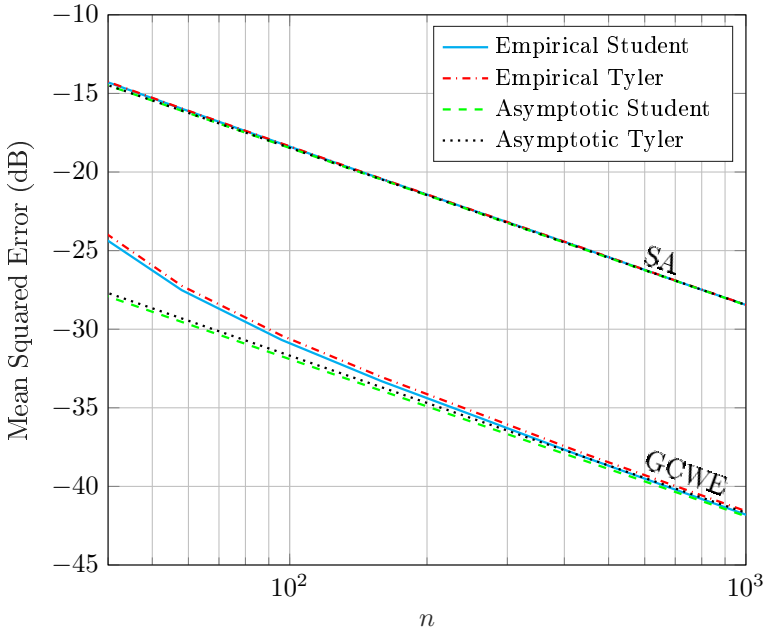
One observes that Figures 3.1 and 3.2 validates the theoretical results given in Theorems 3.1.1 and 3.1.2. Note that the eigenvalues of the Tyler's M-estimator are not analyzed due to its inherent scaling ambiguity. Consequently, the theoretical results are not valid in this case. However, one can note that the values of the eigenvectors of the Tyler's M-estimator tend to the asymptotic values defined in Eqs. (3.3) and (3.11). This is expected since, as noted in Remark 3.1.1, the eigenvectors are given as scale invariant functions of the scatter matrix and therefore, provide same results for the true scatter matrix and normalized scatter (or shape) matrix.

Figure 3.3 compares the asymptotic variances in the SA and GCWE regime. Figure 3.3a displays the empirical mean squared error (MSE) of $\widehat{\boldsymbol{\lambda}}^t - \boldsymbol{\lambda}$ and $\widehat{\boldsymbol{\lambda}}^t - \widehat{\boldsymbol{\lambda}}^{\text{GCWE}}$ as well as their corresponding asymptotic values, i.e., $\text{Tr}(\vartheta_1\boldsymbol{\Lambda}^2 + \vartheta_2\boldsymbol{\lambda}\boldsymbol{\lambda}^T)/n$ (Theorem 3.1.1) and $\text{Tr}(\sigma_1\boldsymbol{\Lambda}^2 + \sigma_2\boldsymbol{\lambda}\boldsymbol{\lambda}^T)/n$ (Theorem 3.1.2). Figure 3.3b displays the empirical MSE of $\widehat{\mathbf{u}}_1^M - \mathbf{u}$ and $\widehat{\mathbf{u}}_1^M - \widehat{\mathbf{u}}_1^{\text{GCWE}}$ where the first eigenvector for both Student's and Tyler's M-estimator and their corresponding asymptotic values, i.e., $\text{Tr}(\vartheta_1\lambda_1\mathbf{U}\boldsymbol{\Lambda}(\lambda_1\mathbf{I} - \boldsymbol{\Lambda})^{+^2}\mathbf{U}^H)/n$ (Theorem 3.1.1) and $\text{Tr}(\sigma_1\lambda_1\mathbf{U}\boldsymbol{\Lambda}(\lambda_1\mathbf{I} - \boldsymbol{\Lambda})^{+^2}\mathbf{U}^H)/n$ (Theorem 3.1.2). The data is t-distributed with $\nu = 3$ and $p = 20$.

First, one can note that the empirical MSEs tend to the proposed asymptotic values. Then, it is apparent that the MSE is significantly smaller in the GCWE regime than in the SA. The difference is even more remarkable for eigenvectors. As previously stated, the variance of eigenvectors depends only on the first scale factor (ϑ_1/σ_1), while the variance

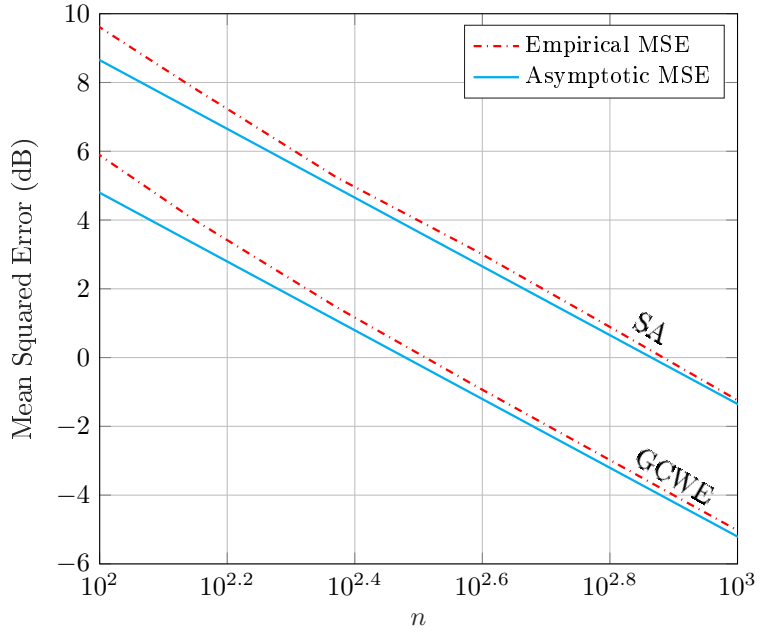


(a) Eigenvalues of the Student's M-estimator

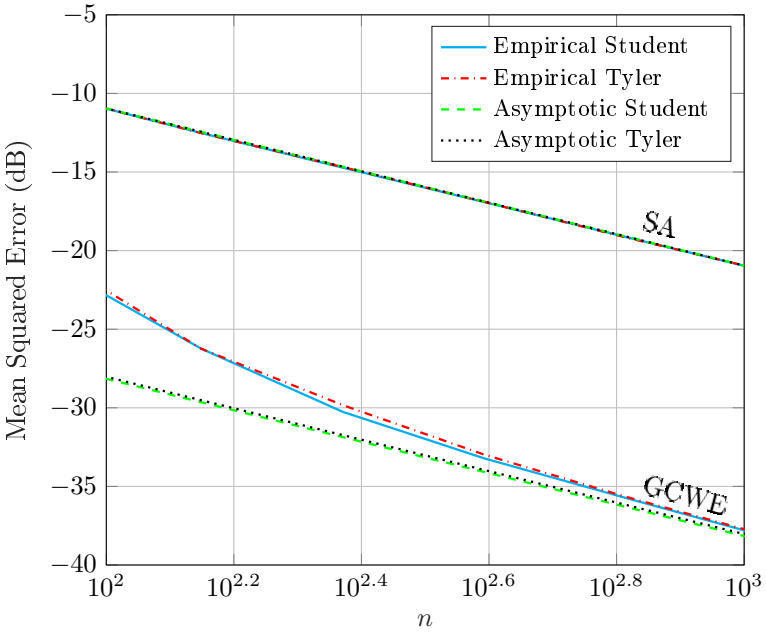


(b) Eigenvectors of the Student's and Tyler's M-estimator

Figure 3.3 – Empirical and asymptotic MSE on eigenvalues and eigenvectors for the SA and GCWE regime; t-distributed data with $\nu = 3$, $p = 20$, $\rho = 0.9(1 + \sqrt{-1})/\sqrt{2}$.



(a) Eigenvalues of the Student's M-estimator



(b) Eigenvectors of the Student's and Tyler's M-estimator

Figure 3.4 – Empirical and asymptotic MSE on eigenvalues and eigenvectors for the SA and GCWE regime; t-distributed data with $\nu = 3$, $p = 50$, $\rho = 0.9(1 + \sqrt{-1})/\sqrt{2}$.

of eigenvalues includes also the second one (ϑ_2/σ_2). We have seen in Chapter 2 that for the Student's M-estimator σ_1 is considerably smaller than ϑ_1, while the second scale factor stays unchanged, i.e. $\sigma_2 = \vartheta_2$, and does not go to zero when the dimension increases for the fixed value of ν. This explains the obtained results.

Figure 3.4 details the results obtained for the data with $p = 50$. Notice that all the conclusions drawn in the previous example are still valid in this case. Moreover, it is evident that the difference between the SA and the GCWE regime is even more significant for higher data dimension. However, one can note the empirical MSE needs higher sample size n to provide the asymptotic results.

These results support the idea that the behaviour of the EVD of an M-estimator, in this case the Student's and Tyler's M-estimator, can be better approximated with the one of their GCWE. This can be easily verified for other M-estimators. In the following section we give an example where the derived results can be leveraged.

3.1.3 Application to intrinsic bias

In [160] were derived Intrinsic (i.e. Riemannian Manifold oriented) counterparts of the Cramér-Rao inequality (ICRLB). In the context of covariance matrix estimation, these results permit notably to account for the natural Riemmanian metric on $\mathcal{H}$, and to bound the expected Riemannian distance (rather than the Euclidean one):

$$d_{nat}^2\left(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2\right) = \left\| \ln\left(\boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1/2}\right)\right\|_F^2. \tag{3.13}$$

This analysis also reveals unexpected and hidden properties of estimators, such as the bias of the SCM w.r.t. the natural metric on $\mathcal{H}^+$ (this bias phenomenon does not exist w.r.t. the Euclidean metric). In this scope, the biased ICRLB (B-ICRLB) is established for the SCM in the Gaussian context in [160, Theorem 7 and Corollary 5], and reads as follows.

Theorem 3.1.3. *B-ICRLB for SCM*
Let $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ be an n-sample distributed as $\mathbf{z}_i \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$ and $\widehat{\boldsymbol{\Sigma}}_{\text{SCM}}$ be the SCM as in Eq. (1.9). The bias w.r.t. the natural metric on $\mathcal{H}^+$ of $\widehat{\boldsymbol{\Sigma}}_{\text{SCM}}$ is

$$\mathbb{E}\left[\exp_{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{\Sigma}}_{\text{SCM}}\right] = -\eta(p,n)\boldsymbol{\Sigma} \tag{3.14}$$

with $\exp_{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{\Sigma}}_{\text{SCM}} = \boldsymbol{\Sigma}^{1/2}\log\left(\boldsymbol{\Sigma}^{-1/2}\widehat{\boldsymbol{\Sigma}}_{\text{SCM}}\boldsymbol{\Sigma}^{-1/2}\right)\boldsymbol{\Sigma}^{1/2}$, and

$$\begin{aligned}\eta(p,n) \quad &= \frac{1}{p}\left\{p\ln(n) + p - \psi(n-p+1)\right. \\ &\quad + (n-p+1)\psi(n-p+2) \\ &\quad \left. + \psi(n+1) - (n+1)\psi(n+2)\right\}\end{aligned} \tag{3.15}$$

and $\psi(x) = \Gamma'(x)/\Gamma(x)$ is the digamma function. Moreover, the natural distance Eq. (3.13) between $\widehat{\boldsymbol{\Sigma}}_{\text{SCM}}$ and $\boldsymbol{\Sigma}$ satisfies the following biased-ICRLB inequality

$$\mathbb{E}\left[d_{nat}^2\left(\hat{\boldsymbol{\Sigma}}_{\text{SCM}}, \boldsymbol{\Sigma}\right)\right] \geq \frac{p^2}{n} + p\eta\left(p,n\right)^2. \tag{3.16}$$

For CES-distributed samples, the ICRLB on d_{nat}^2 is derived in [19] as follows.

Theorem 3.1.4. *IRCLB for CES*
Let $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ be an n-sample distributed as $\mathbf{z}_i \sim \mathcal{CES}(\mathbf{0}, \boldsymbol{\Sigma}, g_{\mathbf{z}})$. Any unbiased estimator $\widehat{\boldsymbol{\Sigma}}$ of $\boldsymbol{\Sigma}$ satisfies the inequality

$$\mathbb{E}\left[d_{nat}^2\left(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}\right)\right] \geq \frac{p^2-1}{n\alpha} + \left(n(\alpha+p\beta)\right)^{-1}, \tag{3.17}$$

with $\alpha = \left(1 + \frac{\mathbb{E}\left[\mathcal{Q}^2u'(\mathcal{Q})\right]}{p(p+1)}\right)$ (where $\mathcal{Q}$ is the modular variate as in Eq. (1.14)) and $\beta = \alpha - 1$.

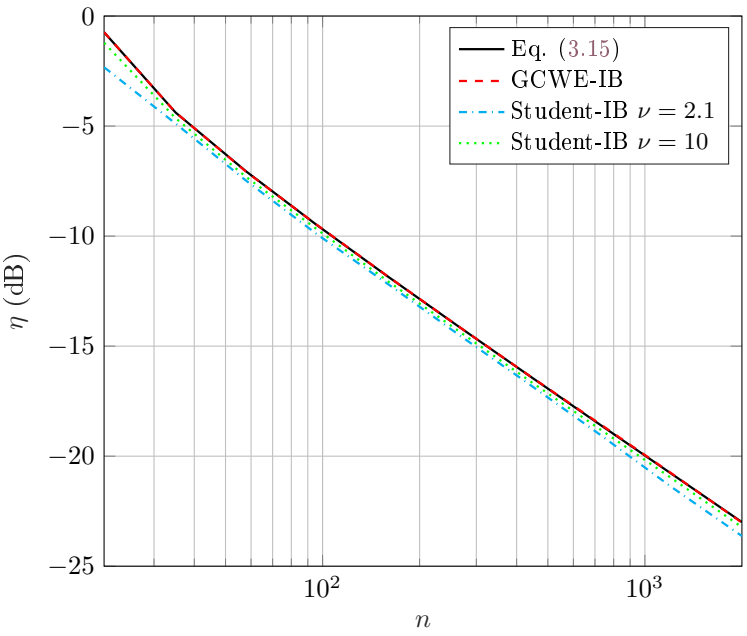


Figure 3.5 – Empirical intrinsic bias for Student's M-estimator (Student-IB) and the GCWE (GCWE-IB) compared to the theoretical result obtained for the GCWE (Eq. (3.15))

Characterizing a bias term in Theorem 3.1.4 (similarly to the one in Theorem 3.1.3) would requires to derive the intrinsic bias of an M-estimator obtained with CES-distributed samples. The problem appears intractable since this result is mainly obtained thanks to the exact distribution of the eigenvalues of a Wishart distributed matrix, and cannot be recovered through a Delta method using the standard asymptotic of Theorem 3.1.1. However the strong proximity of the eigenvalues of an M-estimator towards their GCWE described in Theorem 3.1.2 (also exhibited by the previous simulation results) gives a reasonable theoretical ground for the following approximation:

Approximation 3.1.5: Intrinsic bias of M-estimators [J2]

Let $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ be an n-sample distributed as $\mathbf{z}_i \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}})$. Let $\widehat{\mathbf{\Sigma}}$ be an M-estimator of $\mathbf{\Sigma}$ that is consistent in scale (i.e., $\sigma = 1$ in Eq. (1.31)) and $\widehat{\mathbf{\Sigma}}_{\text{GCWE}}$ its GCWE (Definition 2.1.2). The matrix $\widehat{\mathbf{\Sigma}}_{\text{GCWE}}$ is Wishart-distributed with center $\mathbf{\Sigma}$, so the derivations of Theorem 7 of [160] directly apply to its intrinsic bias. Theorem 3.1.2 then supports the approximation

$$\mathbb{E}[\exp_{\mathbf{\Sigma}}^{-1} \widehat{\mathbf{\Sigma}}] \simeq \mathbb{E}[\exp_{\mathbf{\Sigma}}^{-1} \widehat{\mathbf{\Sigma}}_{\text{GCWE}}] = -\eta(p,n)\mathbf{\Sigma}. \tag{3.18}$$

Figure 3.5 confirms the previous results and supports the proposed approximation. Indeed, it can be seen that the empirical intrinsic bias obtained with Student's M-estimator computed with t-distributed ($\nu = 2.1$ and $\nu = 10$) data coincides with the intrinsic bias based on the corresponding GCWE and the theoretical result in Eq. (3.15). As expected, for higher value of ν (green curve), this approximation is more accurate since in that case the Student's M-estimator is closer to its GCWE. For smaller values of ν (blue curve), the bias exhibits a slight deviation from the theoretical approximation, since the t-distribution has heavier tails and the Student's M-estimator is closer to the Tyler's one.

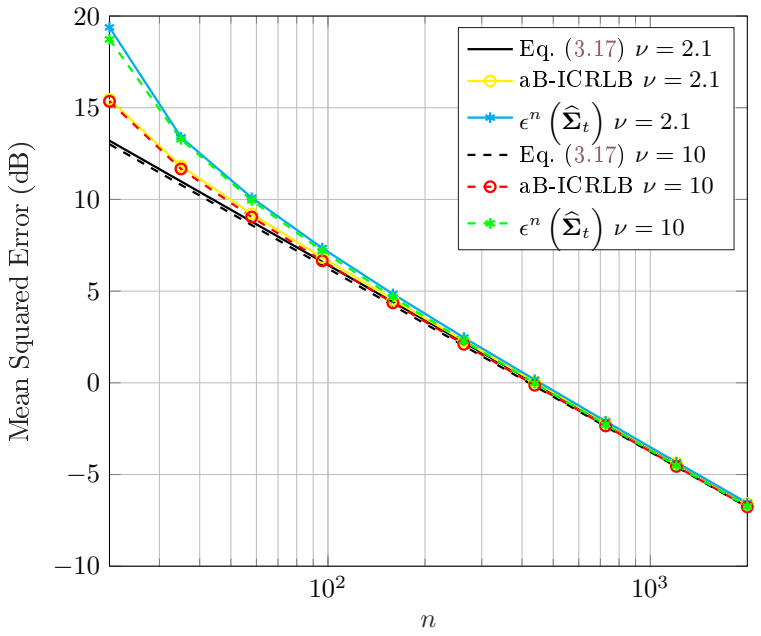


Figure 3.6 – Empirical mean of $d^2_{\text{nat}}\left(\widehat{\mathbf{\Sigma}}_t, \mathbf{\Sigma}\right)$ denoted as $\epsilon^n\left(\widehat{\mathbf{\Sigma}}_t\right)$ versus theoretical CRLB for an unbiased estimator in the CES framework (Eq. (3.17)) and approximated biased instrinsic CRLB (aB-ICRLB)

Finally, we propose to incorporate an equivalent bias term in Eq. (3.17) to obtain an accurate approximation of the B-ICRLB of M-estimators built form CES-distributed samples.

Approximation 3.1.6: B-ICRLB for CES [J2]

Let $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ be an n-sample distributed as $\mathbf{z}_i \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}})$. Let $\widehat{\mathbf{\Sigma}}$ be an M-estimator of $\mathbf{\Sigma}$ that is consistent in scale (i.e., $\sigma = 1$ in Eq. (1.31)). We have the following approached B-ICRLB

$$\mathbb{E}\left[d^2_{nat}\left(\widehat{\mathbf{\Sigma}}, \mathbf{\Sigma}\right)\right] \geq \frac{p^2-1}{n\alpha} + \left(n(\alpha + p\beta)\right)^{-1} + p\eta\left(p, n\right)^2, \tag{3.19}$$

with α and β defined in Theorem 3.1.4 and $\eta\left(p, n\right)$ from Eq. (3.15).

Figure 3.6 illustrates this approximation. The empirical mean of the natural Riemannian distance of $\widehat{\mathbf{\Sigma}}_t$ (denoted as $\epsilon^n(\widehat{\mathbf{\Sigma}}_t)$) is compared to the theoretical ICRLB in Eq. (3.17)) and the approached B-ICRLB in Eq. (3.19). As expected, one can see that the approached bias term in B-ICRLB offers a more accurate theoretical approximation for bounding the expected natural distance.

3.2 eFusion

Given an n-sample $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ of i.i.d. p-dimensional vectors, the SCM defined in Eq. (C.1), uniquely minimizes the loss function

$$l(\mathbf{\Sigma}; \widehat{\mathbf{\Sigma}}_{\text{SCM}}) = \text{Tr}(\mathbf{\Sigma}^{-1}\widehat{\mathbf{\Sigma}}_{\text{SCM}}) + \log\{\det(\mathbf{\Sigma})\} \tag{3.20}$$

over $\mathbf{\Sigma} \in \mathcal{S}^{p\times p}$ given that $p < n$, where $\mathcal{S}^{p\times p}$ denotes the set of $p \times p$ positive definite symmetric matrices. The loss function Eq. (3.20) corresponds to two times the negative log-likelihood function when sampling from p-variate normal distribution. Insufficient number of samples causes significant estimation error in the SCM. When the sample size n is not of orders of magnitude larger than the dimensionality, p, it has long been recognized that larger eigenvalues of the SCM tend to overestimate, whereas the smaller eigenvalues tend to underestimate the true eigenvalues. To fill this gap, regularized or penalized estimators of covariance matrix have been introduced in a series of papers [102, 104, 26, 7, 1, 136, 163, 129, 11, 127].

A regularized estimator of covariance matrix may be an optimally weighted average of the SCM and a well-structured target estimator, which determines what type of structure is imposed on the estimator. The weight parameter controls how much structure is required [103, 104, 11, 127]. Another approach in regularizing the SCM is to shrink the eigenvalues towards each other, and not towards a predefined target value. Such an approach, called *elasso*, was developed in [171], where the authors developed a family of non-smooth penalty functions that not only shrink the eigenvalues towards each other, but they may result in partitioning the eigenvalues into subgroups.

A regularized SCM (RSCM) $\widehat{\mathbf{\Sigma}}_{\text{RSCM}}$ is then defined as the minimizer of

$$L(\mathbf{\Sigma}; \widehat{\mathbf{\Sigma}}_{\text{RSCM}}; \eta) = l(\mathbf{\Sigma}; \widehat{\mathbf{\Sigma}}_{\text{RSCM}}) + \eta\Pi(\mathbf{\Sigma}), \tag{3.21}$$

where $\Pi(\mathbf{\Sigma})$ denotes a nonnegative penalty function, with $\eta \geq 0$ being the regularization parameter.

In this section, we exploit the assumption that $\mathbf{\Sigma}$ has a structure with only a few distinct eigenvalues, i.e., there are k groups of identical eigenvalues. We propose an RSCM estimator that groups the eigenvalues by penalizing large differences between successive eigenvalues.

3.2.1 Proposed method

Let $d_1 \geq \cdots \geq d_p > 0$ and $\lambda_1 \geq \cdots \geq \lambda_p > 0$ denote the ordered eigenvalues of $\widehat{\mathbf{\Sigma}}_{\text{SCM}}$ and $\mathbf{\Sigma}$, respectively. Furthermore, let $r_j = \log(\lambda_j) - \log(\lambda_{j+1})$ denote the differences, referred to as gaps, between successive log-eigenvalues of $\mathbf{\Sigma}$ and let $r_j^{[0]} = \log(d_j) - \log(d_{j+1})$ denote the gaps between log-eigenvalues of $\widehat{\mathbf{\Sigma}}_{\text{SCM}}$ for $j = 1, \ldots, p-1$. We propose to find $\widehat{\mathbf{\Sigma}}$ as the minimizer of Eq. (3.21) based on the following non-convex penalty

$$\Pi(\mathbf{\Sigma}) = \sum_{j=1}^{p-1} \rho_c\left(\frac{r_j}{s}\right), \tag{3.22}$$

where $\rho_c(\cdot) : \mathbb{R} \to \mathbb{R}^+$ denotes Tukey' s biweight function [152]:

$$\rho_c(r) = \frac{1}{6} \cdot \min\left\{1, 1 - \left(1 - \frac{r^2}{c^2}\right)^3\right\}, \quad r \in \mathbb{R}, \tag{3.23}$$

where c is a user-defined tuning parameter and s is the sample standard deviation (SD) of $r_j^{[0]}$ for $j = 1, \ldots, p-1$.

In the minimization of Eq. (3.21), our penalty function in Eq. (3.22) assigns relatively large weights to smaller gaps r_j's, whereas very large gaps attain relatively smaller weights due to the boundedness of Tukey's loss function. The latter property is required in order to achieve the grouping effect of eigenvalues. We refer to Eq. (3.22) as *eFusion penalty* and the corresponding estimator $\widehat{\mathbf{\Sigma}}$ as the eFusion RSCM estimator. We devise an iteratively reweighting (IR) algorithm for computing the proposed RSCM estimator. The IR algorithms are commonly used in finding approximate solutions to such non-convex optimization problems [75, 183], as in the case of M-estimators.

According to [171, Lemma 2.2], for an orthogonally invariant penalty $\Pi(\mathbf{\Sigma})$, the RSCM estimator $\widehat{\mathbf{\Sigma}}$ and the SCM $\widehat{\mathbf{\Sigma}}_{\text{SCM}}$ possess the same set of eigenvectors, with the associated eigenvalues following the same ordering. Note that the eFusion penalty Eq. (3.22) is orthogonally invariant and hence we only need to solve the eigenvalues $\widehat{\lambda}_1, \ldots, \widehat{\lambda}_p$. Thus, due to [171], the eigenvalues of the eFusion estimator $\widehat{\mathbf{\Sigma}}$ can be found as minimizers of

$$\begin{aligned}\mathcal{L}(\boldsymbol{\lambda}; \mathbf{d}, \eta) &= \sum_{j=1}^{p} \left(\frac{d_j}{\lambda_j} + \log(\lambda_j) \right) + \eta \sum_{j=1}^{p-1} \rho_c \left(\frac{r_j}{s} \right) \\ &= \mathbf{d}^T \boldsymbol{\lambda}^{-1} + \log(\boldsymbol{\lambda})^T \mathbf{1} + \eta \sum_{j=1}^{p-1} \rho_c \left(\frac{r_j}{s} \right), \end{aligned} \tag{3.24}$$

over $\lambda_1 \geq \cdots \geq \lambda_p > 0$, i.e., over the ordered eigenvalues.

Above $\mathbf{d} = (d_1, \ldots, d_p)^T$, $\boldsymbol{\lambda}^{-1} = (1/\lambda_1, \ldots, 1/\lambda_p)^T$ and $\mathbf{1}$ is a vector of size $p \times 1$ with all elements equal to one. By setting the gradient of Eq. (3.24) w.r.t. $\boldsymbol{\lambda}$ to zero $\nabla_{\boldsymbol{\lambda}} \mathcal{L} = \mathbf{0}$ we get

$$-\text{diag}(\boldsymbol{\lambda})^{-2} \mathbf{d} + \text{diag}(\boldsymbol{\lambda})^{-1} \left(\mathbf{1} + \frac{\eta}{s} \sum_{j=1}^{p-1} \rho_c' \left(\frac{r_j}{s} \right) \kappa_j \right) = \mathbf{0},$$

where

$$\kappa_j = \Bigg[\underbrace{0 \quad \cdots \quad 0}_{j-1} \quad 1 \quad -1 \quad 0 \quad \cdots \quad 0 \Bigg]^T_{1 \times p}.$$

After some straightforward mathematics, we obtain the following estimating equation.

$$\mathbf{f}(\boldsymbol{\lambda}) = \text{diag} \left(\mathbf{1} + \frac{\eta}{s} \mathbf{v} \right) \boldsymbol{\lambda} - \mathbf{d} = \mathbf{0}, \tag{3.25}$$

where $\mathbf{v} = (v_1, \ldots, v_p)^T$ with $v_j = \rho_c'(r_j/s) - \rho_c'(r_{j-1}/s)$ for $j \in \{1, \ldots, p\}$. Note that, $v_1 = \rho_c'(r_1/s)$ and $v_p = -\rho_c'(r_{p-1}/s)$.

The solution to Eq. (3.25) can be obtained by solving the following system of equations

$$1 + \frac{\eta}{s} \left(\rho_c' \left(\frac{r_j}{s} \right) - \rho_c' \left(\frac{r_{j-1}}{s} \right) \right) - \frac{d_j}{\lambda_j} = 0, \tag{3.26}$$

for $j = 1, \ldots, p$.

This can be reformulated as

$$1 + \eta \left(r_j w_j - r_{j-1} w_{j-1} \right) / s^2 - d_j / \lambda_j = 0, \tag{3.27}$$

where $w_j = \rho_c'(r_j/s)/(r_j/s)$ are referred to as weights. By substituting $r_j = \log(\lambda_j) - \log(\lambda_{j+1})$ and $r_{j-1} = \log(\lambda_{j-1}) - \log(\lambda_j)$ to Eq. (3.27), we obtain the following system of fixed-point equations

$$\log(\lambda_j) = \frac{\frac{s^2}{\eta}(d_j/\lambda_j - 1) + w_j \log \lambda_{j+1} + w_{j-1} \log \lambda_{j-1}}{w_j + w_{j-1}}, \tag{3.28}$$

for $j = 1, \ldots, p$. Not that for $j = 1$ and $j = p$, Eq. (3.28) reduces to

$$\begin{aligned} \log(\lambda_1) &= \frac{s^2}{w_1 \eta} \left(\frac{d_1}{\lambda_1} - 1 \right) + \log \lambda_2, \\ \log(\lambda_p) &= \frac{s^2}{w_{p-1} \eta} \left(\frac{d_p}{\lambda_p} - 1 \right) + \log \lambda_{p-1}. \end{aligned}$$

Algorithm 1: Iteratively reweighted eFusion algorithm [C5]

Input : $\mathbf{d}$: Eigenvalues of the SCM $\widehat{\mathbf{\Sigma}}_{\text{SCM}}$;
η: Penalty parameter;
c: Tukey tuning constant.
Output : $\widehat{\boldsymbol{\lambda}}$: Penalized eigenvalues verifying Eq. (3.28)

Initialize: $k \leftarrow 0$; $\boldsymbol{\lambda}^{[0]} \leftarrow \mathbf{d}$

1 Compute $s = \text{SD}(\mathbf{r}^{[0]})$,

Repeat

2 Update the gaps:
$r_j^{[k]} \leftarrow \log(\lambda_j^{[k]}) - \log(\lambda_{j+1}^{[k]}),\ \ j = 1, \ldots, p-1,$

3 Update the weights:
$w_j^{[k]} \leftarrow \rho_c'(r_j^{[k]}/s)/(r_j^{[k]}/s),\ \ j = 1, \ldots, p-1,$

4 Update the eigenvalue estimates:
$\log \lambda_j^{[k+1]} \leftarrow \frac{1}{w_j^{[k]} + w_{j-1}^{[k]}} \Big(\frac{s^2}{\eta}(d_j/\lambda_j^{[k]} - 1) + w_j^{[k]} \log \lambda_{j+1}^{[k]} + w_{j-1}^{[k]} \log \lambda_{j-1}^{[k+1]} \Big),$
for $j = 1, \ldots, p$.

5 $k \leftarrow k + 1$

until *convergence*

6 $\hat{\boldsymbol{\lambda}} \leftarrow \big(\exp(\log \lambda_1^{[k+1]}), \ldots, \exp(\log \lambda_p^{[k+1]}) \big)^T$

In the spirit of Iteratively Reweighted Least Squares (IRLS), we devise an IR-eFusion algorithm to find the solution $\widehat{\lambda}_1, \ldots, \widehat{\lambda}_p$ that verify Eq. (3.28).

Using $\mathbf{d}$ as the initial value for $\boldsymbol{\lambda}$, our approach, detailed in Algorithm 1, iterates the following steps until convergence. First, the gaps r_j and the weights w_j are computed for all the eigenvalues. Then, Eq. (3.28) is used to update each eigenvalue estimate in a coordinate-wise fashion, i.e., in updating $\log(\lambda_j)$ we use the already updated $\log(\lambda_{j-1})$. Note that, in Step 4, we avoid updating $\log \lambda_j^{[k+1]}$ if $w_j^{[k]} + w_{j-1}^{[k]} = 0$, i.e., $\log \lambda_j^{[k+1]} \leftarrow \log \lambda_j^{[k]}$.

Notice that the full covariance matrix estimator can be reconstructed by combining the proposed eigenvalues estimators together with the corresponding eigenvectors obtained from the SCM eigendecompostion ([171, Lemma 2.2]).

3.2.2 On choosing the tuning parameter

One of the key challenges in such an IR algorithm with a robust penalty loss function is to find a procedure that provides "optimal" values of tuning parameters. In general, optimizing parameters in an analytical way is a difficult problem.

In order to find an optimal value of the tuning parameter c for Tukey's biweight function we will analyze the distribution of $r_j^{[0]} = \log(d_j) - \log(d_{j+1})$. To that end, let us consider the following binary hypothesis test

$$\begin{cases} H_0 : \lambda_j = \lambda_{j+1}, \\ H_1 : \lambda_j > \lambda_{j+1}, \end{cases} \qquad j = 1, \ldots, p-1.$$

Our goal is to detect when two consecutive eigenvalues are equal and thus, we want to derive the distribution of $r_j^{[0]} = \log(d_j) - \log(d_{j+1})$ under the null hypothesis H_0. Once the distribution is derived, the tuning parameter can be obtained as a threshold that assures a given probability of false alarm. The distribution of $r_j^{[0]}$ can be derived using the result for the joint distribution of d_j and d_{j+1} derived in [190]. More precisely, under the assumption that the data is uncorrelated ($\mathbf{\Sigma} = \mathbf{I}$), one has the following joint distribution of two ordered

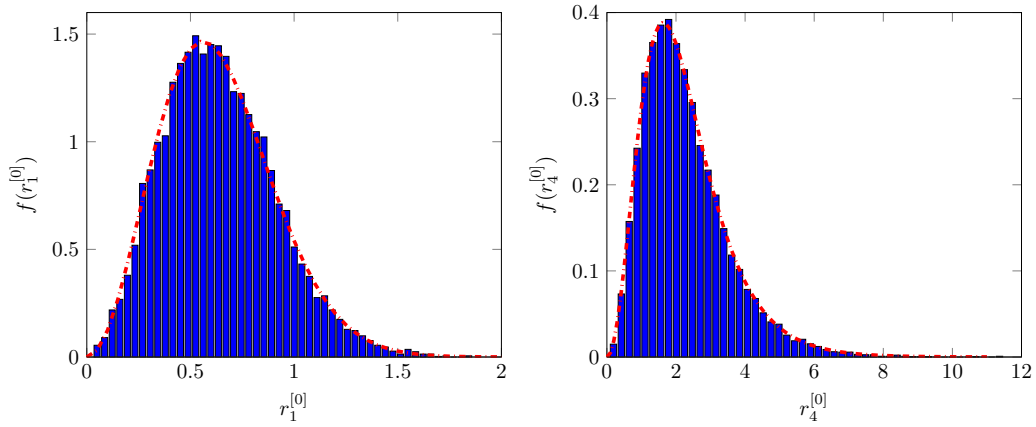


Figure 3.7 – Empirical distribution of $r_1^{[0]}$ (left panel) and resp. $r_4^{[0]}$(right panel) compared to the corresponding theoretical distribution for $p = n = 5$.

consecutive eigenvalues of a Wishart-distributed matrix

$$f_{d_j,d_{j+1}}(x_j, x_{j+1}) = \frac{K}{(j-1)!} \sum_{\mathbf{n}} \sum_{\mathbf{m}} s(\mathbf{n},\mathbf{m}) \left|\mathbf{D}(x_{j+1})\right| \prod_{k=j}^{j+1} \varphi(n_k, m_k, x_k) \prod_{k=1}^{j-1} g(k)$$

where $g(k) = \Gamma(n-p+n_k+m_k-1, x_j)$, the $(s,t)^{\underline{\text{th}}}$ element of $\mathbf{D}(y)$ is given by $\gamma(n-p+l_{s,\mathbf{n}}+l_{t,\mathbf{m}}-1, y)$, Γ and γ are respectively upper and lower incomplete Gamma functions, K is a normalizing constant, $\varphi(n_k, m_k, x_k) = x_k^{n-p+n_k+m_k-2}\exp(-x_k)$ and

$$\sum_{\mathbf{n}} = \sum_{n_1=1}^{p} \sum_{n_2=1, n_2\neq n_1}^{p} \cdots \sum_{n_{j+1}=1, n_{j+1}\neq\{n_1,\ldots,n_j\}}^{p}$$

(analogous for $\sum_{\mathbf{m}}$). The definitions for $s(\mathbf{n},\mathbf{m})$, $l_{s,\mathbf{n}}$ and $l_{t,\mathbf{m}}$ can be found in [190]. In order to derive the distribution of r_j, we perform a change of variables by introducing $x = \log(x_j/x_{j+1})$ and $y = x_{j+1}$. Then, computing the Jacobian $J(x,y) = \frac{1}{x_j} = \exp(-x)/y$ and using $f_{r_j}(x) = \int_0^{+\infty} f_{d_j,d_{j+1}}(x,y)\left|J(x,y)\right|^{-1} dy$ one obtains the result

$$f_{r_j^{[0]}}(x) = \int_0^{+\infty} \left(\frac{K}{(j-1)!} \sum_{\mathbf{n}} \sum_{\mathbf{m}} s(\mathbf{n},\mathbf{m}) \left|\mathbf{D}(y)\right| \right.$$
$$\varphi(n_j, m_j, \exp(x)y)\varphi(n_{j+1}, m_{j+1}, y)$$
$$\left. \prod_{k=1}^{j-1} \Gamma(n-p+n_j+m_j-1, \exp(x)y)\exp(x)y \right) dy. \tag{3.29}$$

One can note that the distribution of $r_j^{[0]}$ depends on the position j, on the data dimension p and on the number of observations n. Indeed, for a small sample size n the SCM's eigenvalues are very distinct, $r_j^{[0]}$'s have high fluctuations, especially for large j (small eigenvalues). On the other hand, for a sufficiently large n, the eigenvalues of SCM are closer to their true values and the gaps between them are smaller. This implies that for larger n the interval of acceptable parameter c is wider. For instance, with our settings and $n = 700$ (instead of 3000) the "optimal" parameter c is approximately between 1.13 and 1.5. When n increases the gaps that correspond to true identical eigenvalues decrease, while the gaps corresponding to distinct eigenvalues rapidly increase. Therefore the choice of c is much

more flexible, e.g. for $n = 3000$ all values between 0.42 and 2.96 give good results. Small values of c can result in more groups than expected, while large values have tendency to fuse even very different eigenvalues.

For high dimension p and for small eigenvalues (so large j), Eq. (3.29) is computationally demanding, but we can empirically show that some cases can be well approximated with distributions obtained with significantly smaller parameters. Also, as it can be seen from Figure 3.7, the variations of $r_j^{[0]}$ are much higher for smaller eigenvalues. The range becomes narrower when j decreases. Here, we have plotted the limited case when $n = p$. Obviously, in order to better estimate the covariance matrix, one needs c that depends on the position j. In this analysis, we assume that n is big enough that the variation of $r_j^{[0]}$ are smaller under H_0 and thus a unique c can be used for all j. More general case is still under study. In order to compute that value, we propose then to look, for instance, at the distribution of $r_j^{[0]}$ for $p = 4$, $n = 16$ and $P_{fa} = 0.01$. For j equal to $1, 2, 3$, one obtains c equal to $0.794, 0.86, 1.159$, respectively. These are used in Section 3.2.3 as candidate values.

3.2.3 Experiments and discussion

We compare the performance of the proposed estimator with the elasso [171]. In elasso, $\Pi(\mathbf{\Sigma}) = \sum_{j=1}^{p} a_j \log(\lambda_j)$, is used as the penalty function, where the weights a_j are obtained by centering decreasing quantiles from the Marčenko-Pastur law.

We generate a random sample of size $n = 3000$ from a $p = 100$ dimensional multivariate normal distribution. Similar to [171], the covariance matrix $\mathbf{\Sigma}$ has 40 eigenvalues equal to 20, 30 equal to 10 and 30 equal to 2.

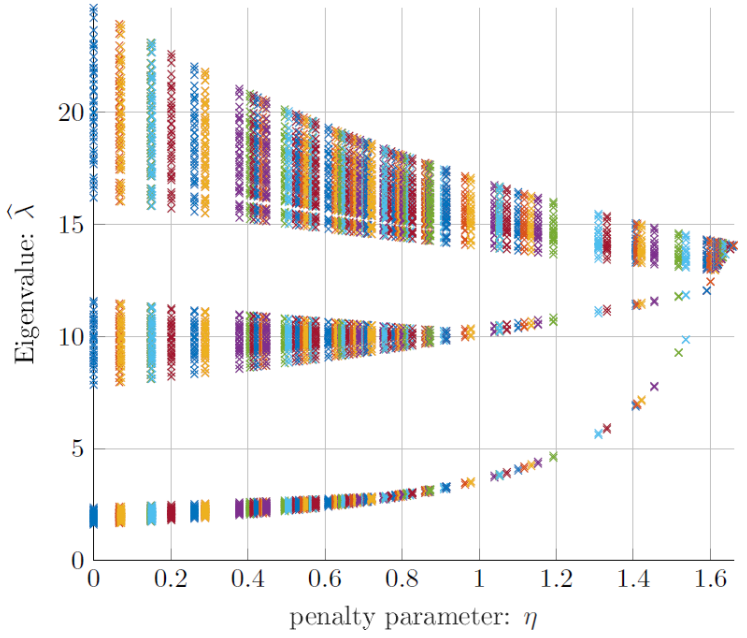
Figure 3.8 displays the process of grouping eigenvalues with elasso (top panel) and eFusion (bottom panel) versus the value of the penalty parameter. The results show a significant improvement that our estimator can offer.

First, it gives an unbiased estimation as the three groups of eigenvalues are well separated and close to their true values. Second, one does not need to search for optimal penalty parameter, contrary to elasso, where it is necessary to use the cross validation to choose a value of η. On the other hand, the eFusion method provides good results for all large values of η. Moreover, from Figure 3.8a, one can note that independently of the chosen value of η, the elasso method provides biased estimators of eigenvalues. For instance, if we choose $\eta \approx 1$, only the second group of eigenvalues ($\lambda_2 = 10$) is well-estimated, while the estimators from other two groups are visibly biased, i.e., the eigenvalues from the first group ($\lambda_1 = 20$) are underestimated, while the eigenvalues from the third group ($\lambda_3 = 2$) are overestimated.

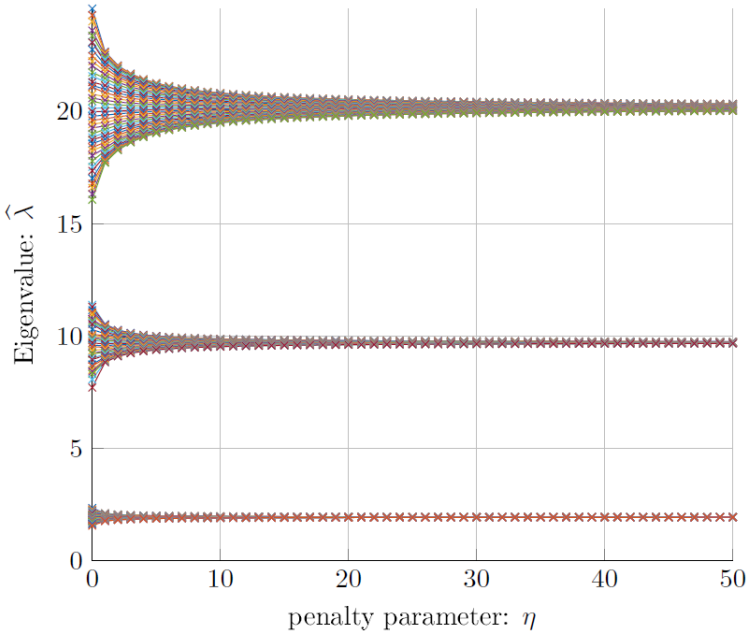
However, a poor choice of c can significantly degrade the estimation obtained with eFusion. In Figure 3.9 the possible scenarios for poor choices of c are illustrated. First, in Figure 3.10a the results obtained for small value of c are displayed. One can note that in that case, for all displayed values of η, the algorithm fails to fuse eigenvalues from the same group and provides biased estimation. In general, the higher eigenvalues are better estimated since, as stated in previous section, the gaps between them are relatively small. This is not the case for small eigenvalues, which finally results in an overestimated number of blocks.

In Figure 3.10b one can see the results obtained for high value of c. As illustrated, a high value of c can produce the fusion of eigenvalues from different blocks. Consequently, the number of blocks in this case is underestimated.

However, even when the parameter c is chosen in the manner that the method provides the exact number of different groups of eigenvalues, the estimations of these values can be slightly biased. This is the case when the sample size is not high enough and the initial estimation obtained with the SCM is not accurate. Figure 3.10 illustrates this point. The sample size is set to 800 and the parameter c is equal to 1.2. From the figure, we can see that the value of the first eigenvalue is slightly overestimated, while the value of the second one is underestimated, confirming our claim. Nevertheless, one can note that even in this



(a) elasso path

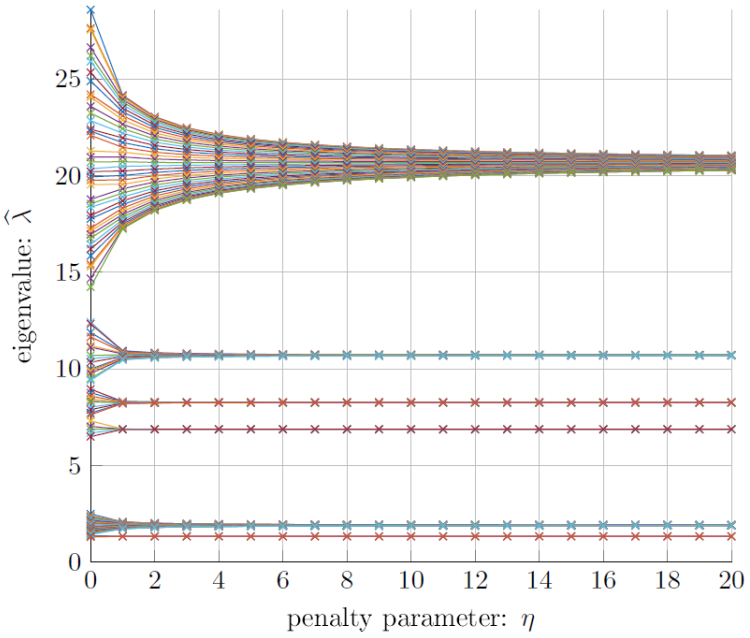


(b) eFusion path for $c = 0.794$

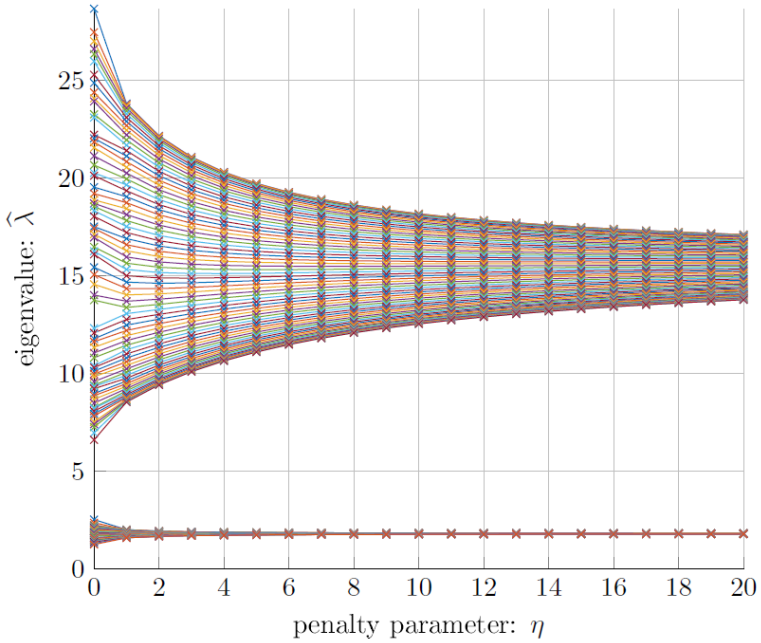
Figure 3.8 – An example of grouping of eigenvalues of elasso (top panel) and eFusion (bottom panel) for $c = 0.794$.

case the eFusion method provides significantly better estimation than the elasso one.

Finally, if the sample size is high enough, as in the first example (Figure 3.8), the eFusion estimator can offer similar results for different choices of c. In order to compare these results we can measure an affine-invariant (Riemannian) distance between positive definite matrices $\boldsymbol{\Sigma}$ and $\widehat{\boldsymbol{\Sigma}}$, defined as $d(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}) = \| \log(\boldsymbol{\Sigma}^{-1/2}\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1/2})\|_F$. Table 3.1 lists the values of the



(a) $c = 0.7$



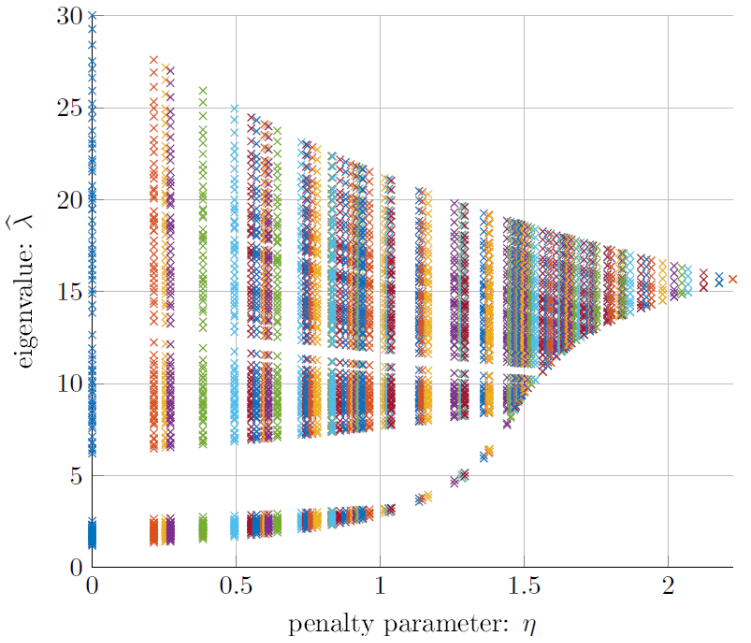
(b) $c = 1.7$

Figure 3.9 – Effects of poor choices of the parameter c, $n = 1000$.

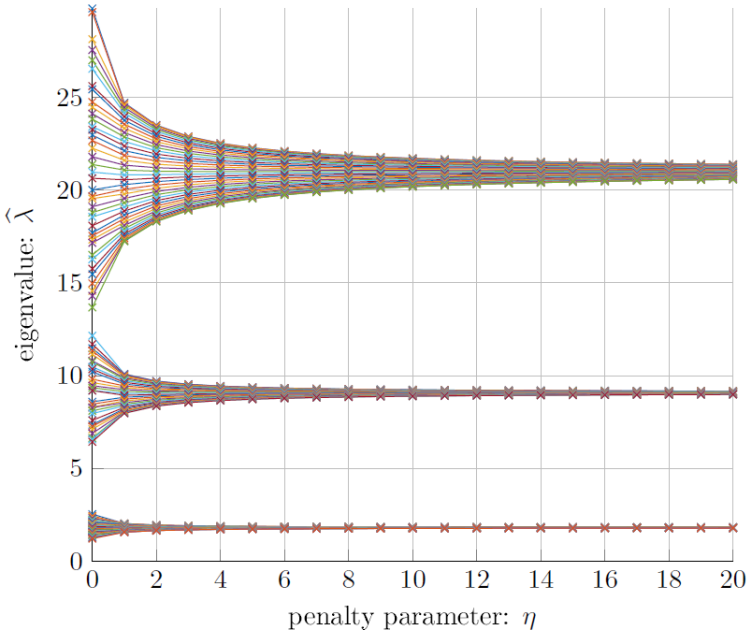
distance for different values of c for the eFusion estimator. The results are obtained for the same initialization d_j, $j = 1, \ldots, p$, and $\eta = 50$.

One can note that the distance has the same value for different choices of c. In fact, to be precise, the values differs from the eighth decimal, which is insignificant.

Finding the best way to choose an optimal value of c, or a range of acceptable values, is still under study. In general, to find a satisfactory value of c one could use the method



(a) elasso path



(b) eFusion path

Figure 3.10 – Grouping of eigenvalues of elasso (top panel) and eFusion (bottom panel) with $c = 1.2$, $n = 800$.

proposed in Section 3.2.2 with relatively small values of p and n, taking a ratio p/n that is larger than the one of the given data. This prevents us from having too small parameter c and consequently biased estimation of eigenvalues.

In the spirit of this thesis, we can propose the extension of the eFusion method adapted

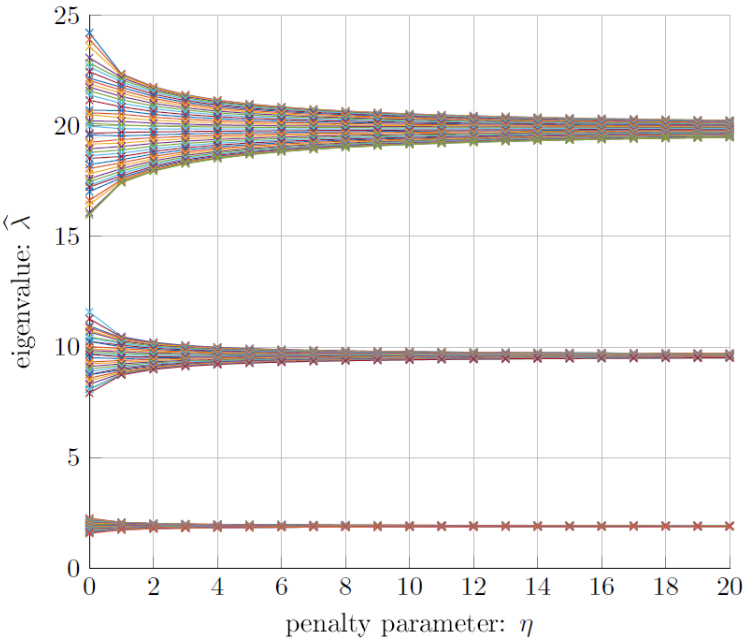


Figure 3.11 – eFusion grouping of eigenvalues with Student's M-estimator $\nu_t = 2$ for $c = 0.794$; t-distributed data, $\nu = 2$, $n = 3000$.

c	0.7	0.8	0.9	1	1.1	1.2	1.3
$d(\mathbf{\Sigma}, \widehat{\mathbf{\Sigma}})$	1.5168	1.5168	1.5168	1.5168	1.5168	1.5168	1.5168
c	1.4	1.5	1.6	1.7	1.8	1.9	2
$d(\mathbf{\Sigma}, \widehat{\mathbf{\Sigma}})$	1.5168	1.5168	1.5168	1.5168	1.5168	1.5168	1.5168

Table 3.1 – Performance analysis of eFusion for different c

to the non-Gaussian data. Thanks to the analysis done in Section 3.1, we can suggest to use an M-estimator instead of the SCM when the data are CES-distributed. Consequently, the initial eigenvalue estimates are close to the ones obtained with SCM in the Gaussian context. Figure 3.11 demonstrates this point. The data follow t distribution with $\nu = 2$. The Student's M-estimator is used as the MLE, i.e., $\nu_t = \nu$. One can note that the algorithm provides the results comparable to the ones of Figure 3.8, i.e., when the SCM is used with Gaussian data and under same set-up ($c = 0.794$ and $n = 3000$). Moreover, Figure 3.12 details the results in a mismatch scenario, i.e. when $\nu_t \neq \nu$. One can note that, once again, eFusion provides an accurate estimation of the eigenvalues.

Furthermore, Figure 3.13a displays the grouping of eigenvalues initialized with the SCM built with t-distributed data under the previous set-up, i.e., $\nu = 2$, $n = 3000$ and $c = 0.794$. One can see that the SCM provides highly biased initial estimates, which cannot be correctly grouped for any value of the penalty parameter η. We also observe that for $c = 0.794$, the method overestimates the number of groups. We can then choose a higher value of c and analyze the results. Figure 3.13b illustrates the results obtained for $c = 3$. We note that the number of classes is closer to the true value, as expected. However, not only the number of classes is not well-estimated, but the values of final estimators stay highly biased. This, once again, confirms the interest of using M-estimators in a non-Gaussian environment.

Finally, the convergence of this method is under analysis. Note that this version of the method has been done for real data and can be easily extended to the complex one. Future

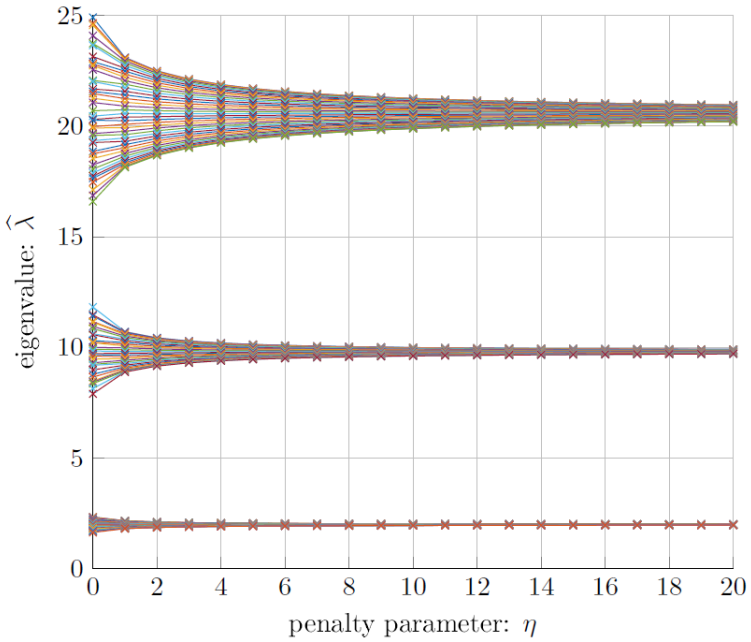


Figure 3.12 – eFusion grouping of eigenvalues with Student's M-estimator $\nu_t = 4$ for $c = 0.794$; t-distributed data, $\nu = 2$, $n = 3000$.

work will also concentrate on its extension to the high dimensional case when $p > n$.

3.3 Principal subspace estimation from M-estimators

In this section, the analysis done in Section 3.1 is extended to the principal subspace (i.e. the subspace spanned by the r strongest eigenvectors) of M-estimators. As for the EVD of M-estimators, for principal subspaces we derive:

- The standard Gaussian asymptotic distribution. This result is obtained by extending the perturbation analysis of [101, 100] to the complex M-estimators. This asymptotic analysis provides an extension of [168] since it provides the exact structure of the asymptotic covariance and pseudo-covariance matrix of principal subspace. Also, contrary to [168], the results are derived for complex data.
- The convergence rate towards the principal subspace of a GCWE by extending the results of Chapter 2.

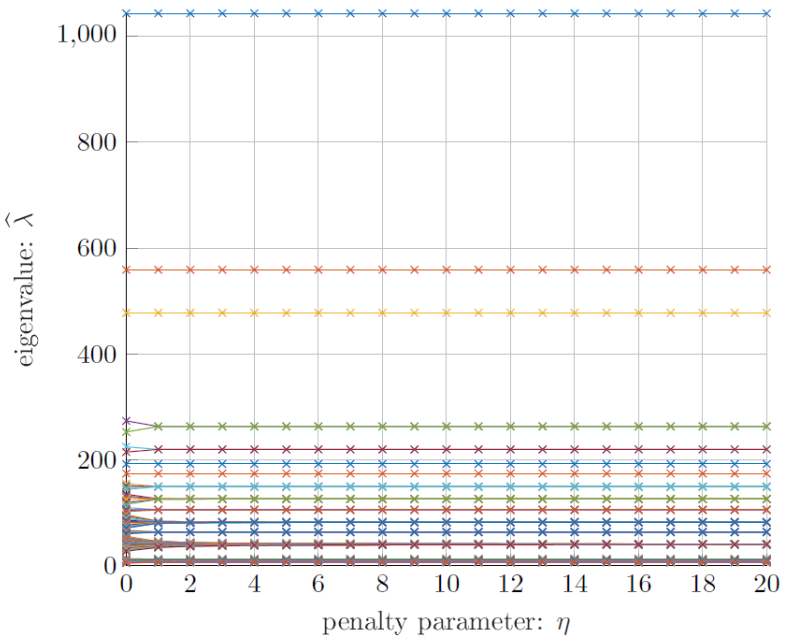
3.3.1 Asymptotics of M-estimators' principal subspace

Consider the case of a low-rank plus identity scatter matrix (also referred to as factor model), that is commonly used in signal processing to account for low dimensional signals embedded in white noise:

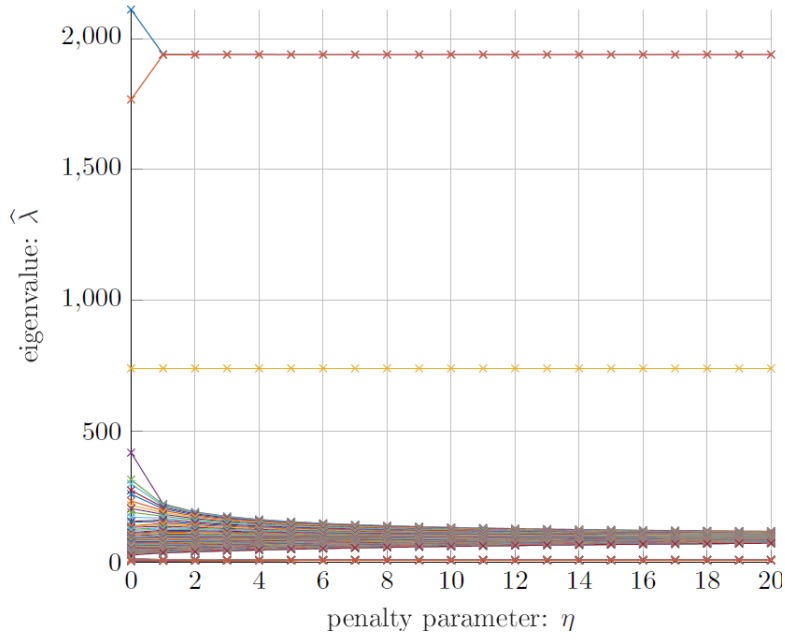
$$\mathbf{\Sigma} = \mathbf{\Sigma}_r + \gamma^2 \mathbf{I}_p \quad \overset{\text{EVD}}{=} \quad [\mathbf{U}_r | \mathbf{U}_r^{\perp}] \mathbf{\Lambda} [\mathbf{U}_r | \mathbf{U}_r^{\perp}]^H \tag{3.30}$$

with the rank r matrix $\mathbf{\Sigma}_r = \mathbf{U}_r \mathbf{\Lambda}_r \mathbf{U}_r^H$, with $\mathbf{U}_r \in \mathcal{U}_r^p$ and $\mathbf{\Lambda}_r \in \mathbb{R}^{r \times r}$.

We focus on the estimation of the orthogonal projector onto the range space spanned by $\mathbf{U}_r$, the r strongest eigenvectors of $\mathbf{\Sigma}$ (referred to as "principal subspace").



(a) $c = 0.794$



(b) $c = 3$

Figure 3.13 – eFusion grouping of eigenvalues for the SCM in a non-Gaussian context; t-distributed data with $\nu = 2$, $n = 3000$.

We define $\mathcal{R}_r\{.\}$ the operator that extracts this principal subspace from a given matrix, i.e.:

$$\begin{array}{rccc} \mathcal{R}_r: & \mathcal{H}^+ & \longrightarrow & \mathcal{G}_r^p \\ & \mathbf{\Sigma} \overset{\text{EVD}}{=} [\mathbf{U}_r|\mathbf{U}_r^\perp]\mathbf{\Lambda}[\mathbf{U}_r|\mathbf{U}_r^\perp]^H & \longmapsto & \mathbf{U}_r\mathbf{U}_r^H \end{array} \tag{3.31}$$

where $\mathcal{G}_r^p$ is the set of rank r orthogonal projectors of $\mathbb{C}^{p\times p}$.

Let us consider an M-estimator $\widehat{\mathbf{\Sigma}}$ built with an n-sample drawn as $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma})$ where

$\mathbf{\Sigma}$ is low-rank structured as in Eq. (3.30), and let $\widehat{\mathbf{\Sigma}}_{\text{GCWE}}$ be its GCWE from Definition 2.1.2. We have the corresponding principal subspaces as:

$$\begin{cases} \mathbf{\Pi}_r = \mathcal{R}_r\{\mathbf{\Sigma}\}, \\ \widehat{\mathbf{\Pi}}_r^M = \mathcal{R}_r\{\widehat{\mathbf{\Sigma}}\}, \\ \widehat{\mathbf{\Pi}}_r^{\text{GCWE}} = \mathcal{R}_r\{\widehat{\mathbf{\Sigma}}_{\text{GCWE}}\}. \end{cases} \tag{3.32}$$

In the following, we derive the asymptotic distributions for the quantities $\widehat{\mathbf{\Pi}}^M$, both under the SA and the GCWE regime.

Theorem 3.3.1: Standard asymptotic [J2]

Let $\widehat{\mathbf{\Pi}}_r^M$ be the estimator of the projector $\mathbf{\Pi}_r$ obtained from an M-estimator as in Eq. (3.32). The asymptotic distribution of $\widehat{\mathbf{\Pi}}_r^M$ is given by

$$\sqrt{n}\text{vec}\left(\widehat{\mathbf{\Pi}}_r^M - \mathbf{\Pi}_r\right) \xrightarrow{d} \mathcal{GCN}\left(\mathbf{0}, \vartheta_1 \mathbf{\Gamma}_{\mathbf{\Pi}}, \vartheta_1 \mathbf{\Gamma}_{\mathbf{\Pi}} \mathbf{K}\right), \tag{3.33}$$

where

$$\mathbf{\Gamma}_{\mathbf{\Pi}} = \mathbf{A}^T \otimes \mathbf{B} + \mathbf{B}^T \otimes \mathbf{A} \tag{3.34}$$

with $\mathbf{A} = \mathbf{U}_r \left(\gamma^2 \mathbf{\Lambda}_r^{-2} + \mathbf{\Lambda}_r^{-1}\right) \mathbf{U}_r^H$, $\mathbf{B} = \gamma^2 \mathbf{\Pi}_r^\perp$ and ϑ_1, ϑ_2 given by Eq. (1.33).

Proof. If we define the pseudo-inverse of $\mathbf{\Sigma}_r$ as

$$\mathbf{\Phi} = \mathbf{U}_r \mathbf{\Lambda}_r^{-1} \mathbf{U}_r^H, \tag{3.35}$$

one has the following Taylor series expansion from [101]

$$\widehat{\mathbf{\Pi}}_r = \mathbf{\Pi}_r + \delta\mathbf{\Pi}_r + \ldots + \delta^i \mathbf{\Pi}_r + \ldots$$

where

$$\begin{aligned} \delta\mathbf{\Pi}_r &= \mathbf{\Pi}_r^\perp \Delta\mathbf{\Sigma}\mathbf{\Phi} + \mathbf{\Phi}\Delta\mathbf{\Sigma}\mathbf{\Pi}_r^\perp, \\ \delta^i\mathbf{\Pi}_r &= -\mathbf{\Pi}_r^\perp \left(\delta^{i-1}\mathbf{\Pi}\right) \Delta\mathbf{\Sigma}\mathbf{\Phi} + \mathbf{\Pi}_r^\perp \left(\delta^{i-1}\mathbf{\Pi}\right) \Delta\mathbf{\Sigma}\mathbf{\Phi}, \end{aligned}$$

with $\Delta\mathbf{\Sigma} = \widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}$.

In the asymptotic regime, when $n \to \infty$, we can write the following first order approximation of Taylor expansion

$$\widehat{\mathbf{\Pi}}_r = \mathbf{\Pi}_r + \delta\mathbf{\Pi}_r$$

since $\Delta\mathbf{\Sigma}$ is close to zero. Hence, taking the vec of $\widehat{\mathbf{\Pi}}_r - \mathbf{\Pi}_r = \delta\mathbf{\Pi}_r$, one gets

$$\text{vec}\left(\widehat{\mathbf{\Pi}}_r - \mathbf{\Pi}_r\right) = \mathbf{F}\text{vec}\left(\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\right)$$

with

$$\mathbf{F} = \left(\mathbf{\Phi}^T \otimes \mathbf{\Pi}_r^\perp + \left(\mathbf{\Pi}_r^\perp\right)^T \otimes \mathbf{\Phi}\right).$$

It is now obvious that the covariance (resp. pseudo-covariance) matrix of $\sqrt{n}\left(\mathbf{\Pi}_r^M - \mathbf{\Pi}_r\right)$ is equal to $\mathbf{F}\mathbf{C}_M\mathbf{F}^H$ (resp. $\mathbf{F}\mathbf{P}_M\mathbf{F}^T$) where $\mathbf{C}_M$ and $\mathbf{P}_M$ are given in Eq. (1.32). Further

$$\begin{aligned} \mathbf{F}\mathbf{C}_M &= \vartheta_1 \left(\mathbf{\Phi}^T \otimes \mathbf{\Pi}_r^\perp + \left(\mathbf{\Pi}_r^\perp\right)^T \otimes \mathbf{\Phi}\right) \left(\mathbf{\Sigma}^T \otimes \mathbf{\Sigma}\right) \\ &+ \vartheta_2 \left(\mathbf{\Phi}^T \otimes \mathbf{\Pi}_r^\perp + \left(\mathbf{\Pi}_r^\perp\right)^T \otimes \mathbf{\Phi}\right) \text{vec}\left(\mathbf{\Sigma}\right) \text{vec}\left(\mathbf{\Sigma}\right)^H \\ &= \vartheta_1 \left(\mathbf{\Phi}^T\mathbf{\Sigma}^T \otimes \mathbf{\Pi}_r^\perp\mathbf{\Sigma} + \left(\mathbf{\Pi}_r^\perp\right)^T \mathbf{\Sigma}^T \otimes \mathbf{\Phi}\mathbf{\Sigma}\right) \end{aligned}$$

as $\left(\mathbf{\Phi}^T \otimes \mathbf{\Pi}_r^\perp + \left(\mathbf{\Pi}_r^\perp\right)^T \otimes \mathbf{\Phi}\right) \text{vec}\left(\mathbf{\Sigma}\right) = \mathbf{0}$ using $\left(\mathbf{T}^T \otimes \mathbf{R}\right) \text{vec}\left(\mathbf{S}\right) = \text{vec}\left(\mathbf{RST}\right)$ and $\mathbf{\Pi}_r^\perp \mathbf{\Sigma} \mathbf{\Phi} = \mathbf{\Phi} \mathbf{\Sigma} \mathbf{\Pi}_r^\perp = \mathbf{0}$. Finally, after the postmultiplication by $\mathbf{F}^H$ and since

$$\begin{aligned}\mathbf{\Sigma} &= \mathbf{\Sigma}^H \neq \mathbf{\Sigma}^T \\ \mathbf{\Phi} &= \mathbf{\Phi}^H \neq \mathbf{\Phi}^T \\ \mathbf{\Pi}_r^\perp &= \left(\mathbf{\Pi}_r^\perp\right)^H \neq \left(\mathbf{\Pi}_r^\perp\right)^T\end{aligned}$$

one obtains

$$\mathbf{F}\mathbf{C}\mathbf{F}^H = \left(\left(\mathbf{\Phi}\mathbf{\Sigma}\mathbf{\Phi}\right)^T \otimes \mathbf{\Pi}_r^\perp \mathbf{\Sigma} \mathbf{\Pi}_r^\perp + \left(\mathbf{\Pi}_r^\perp \mathbf{\Sigma} \mathbf{\Pi}_r^\perp\right)^T \otimes \mathbf{\Phi}\mathbf{\Sigma}\mathbf{\Phi}\right)$$

which with $\mathbf{\Phi}$ given by Eq. (3.35) and $\mathbf{\Sigma} = \mathbf{U}_r \mathbf{\Lambda}_r \mathbf{U}_r^H + \gamma^2 \mathbf{I}_p$ yields the final result.

Analogously, one can derive the results for the pseudo-covariance using the equality Eq. (A.4). □

Theorem 3.3.2: Asymptotic GCWE [J2]

Let $\widehat{\mathbf{\Pi}}_r^M$ and $\widehat{\mathbf{\Pi}}_r^{\text{GCWE}}$ be the estimators of the projector $\mathbf{\Pi}_r$ defined in Eq. (3.32). The asymptotic distribution of $\widehat{\mathbf{\Pi}}_r^M$ is given by

$$\sqrt{n}\text{vec}\left(\widehat{\mathbf{\Pi}}_r^M - \widehat{\mathbf{\Pi}}_r^{\text{GCWE}}\right) \xrightarrow{d} \mathcal{GCN}\left(\mathbf{0}, \sigma_1 \mathbf{\Gamma}_{\mathbf{\Pi}}, \sigma_1 \mathbf{\Gamma}_{\mathbf{\Pi}} \mathbf{K}\right) \tag{3.36}$$

with $\mathbf{\Gamma}_{\mathbf{\Pi}}$ and σ_1, σ_2 given by Eqs. (3.34) and (2.5), respectively.

Proof. Following the same steps as in the proof of Theorem 3.1.2 and using the results of Theorem 3.3.1, one can easily prove the theorem. □

Remark 3.3.2

- Theorem 3.3.1 (resp. 3.3.2) extends the results of Theorem 3.1.1 (resp. 3.1.2) to the principal subspace of M-estimators. We can draw the same conclusions as in Remark 3.1.2, notably, that an underlying Wishart equivalent offers a more accurate asymptotic equivalent than the standard Gaussian one.

3.3.2 Experiments

In this part we present some experiments in order to validate the theoretical results.

The data is t-distributed with the scatter matrix constructed as $\mathbf{\Sigma} = \mathbf{\Sigma}_r + \mathbf{I}$ with $\mathbf{\Sigma}_r = \mathbf{U}_r \mathbf{\Lambda}_r \mathbf{U}^H$ where only the 5 first eigenvalues and eigenvectors of $\mathbf{\Sigma}$ are kept, and scaled so that $\text{Tr}(\mathbf{\Sigma}_r) = 100$. The results are illustrated for the Student's and Tyler's M-estimator.

Figure 3.14 plots the relative error norm between the empirical covariance matrix of $\sqrt{n}\text{vec}\left(\widehat{\mathbf{\Pi}}_r^t - \mathbf{\Pi}_r\right)$ and its asymptotic value given by Eq. (3.33) and the relative error norm between the empirical covariance matrix of $\sqrt{n}\text{vec}\left(\widehat{\mathbf{\Pi}}_r^t - \widehat{\mathbf{\Pi}}_r^{\text{GCWE}}\right)$ and its asymptotic value given by Eq. (3.36), where $\widehat{\mathbf{\Pi}}_r^t$ is the projector obtained with the Student's M-estimator.

Analogously, Figure 3.15 plots the relative error norm between the empirical covariance matrix of $\sqrt{n}\text{vec}\left(\widehat{\mathbf{\Pi}}_r^{\text{T}} - \mathbf{\Pi}_r\right)$ and its asymptotic value given by Eq. (3.33) and the relative error norm between the empirical covariance matrix of $\sqrt{n}\text{vec}\left(\widehat{\mathbf{\Pi}}_r^{\text{T}} - \widehat{\mathbf{\Pi}}_r^{\text{GCWE}}\right)$ and its

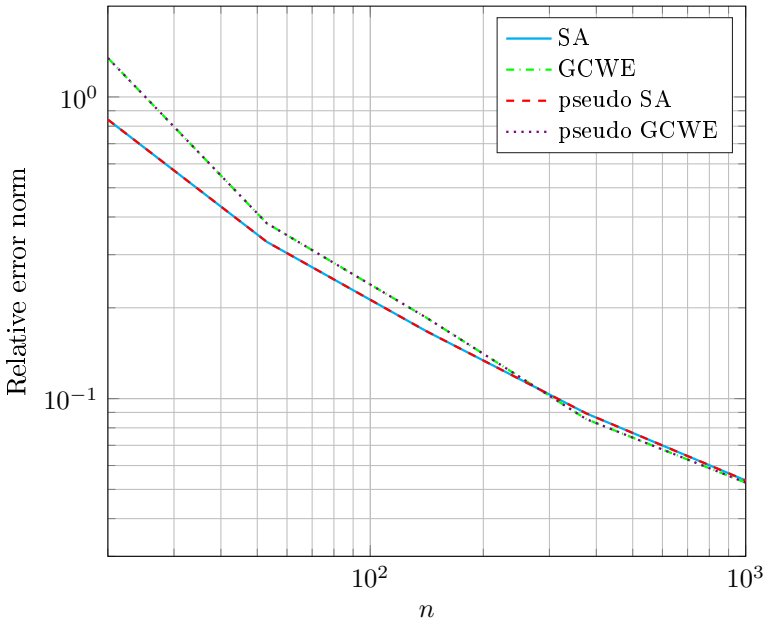


Figure 3.14 – Relative error norm between the empirical covariance (resp. pseudo-covariance) matrix of $\sqrt{n}\text{vec}\left(\widehat{\mathbf{\Pi}}_r^t - \mathbf{\Pi}_r\right)$ and its asymptotic value and the corresponding result for $\sqrt{n}\text{vec}\left(\widehat{\mathbf{\Pi}}_r^t - \widehat{\mathbf{\Pi}}_r^{\text{GCWE}}\right)$; t-distributed data with $\nu = 2$, $p = 10$.

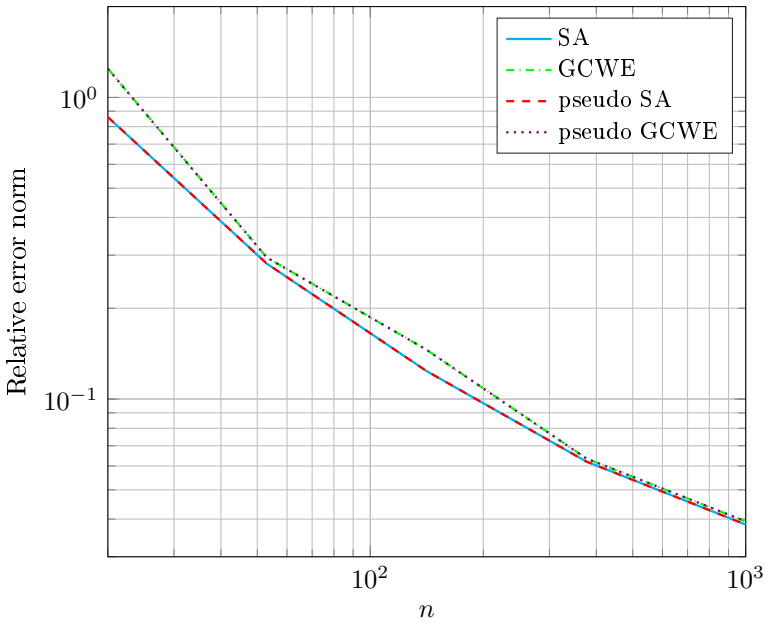
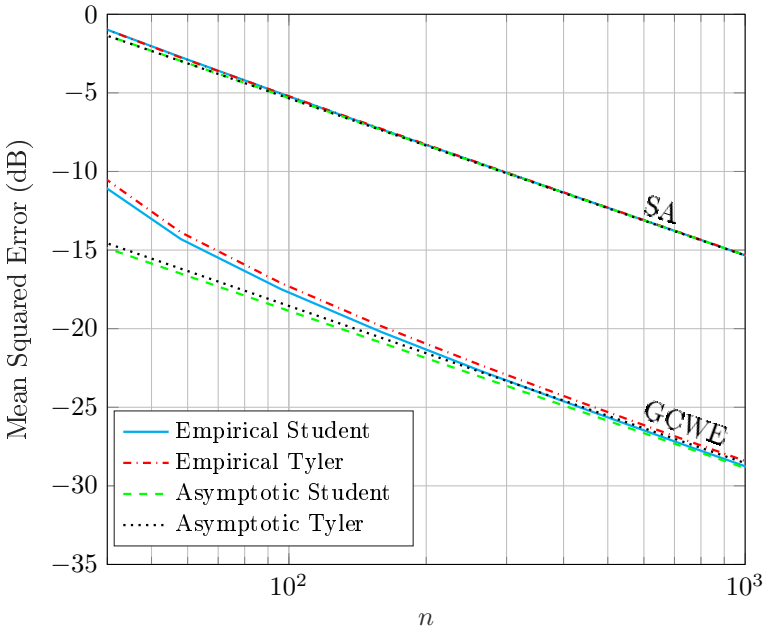
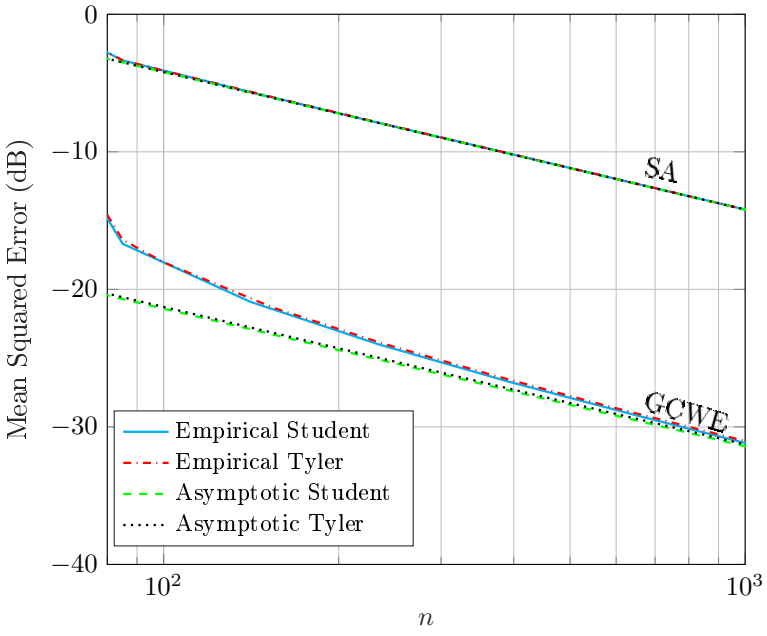


Figure 3.15 – Relative error norm between the empirical covariance (resp. pseudo-covariance) matrix of $\sqrt{n}\text{vec}\left(\widehat{\mathbf{\Pi}}_r^{\text{T}} - \mathbf{\Pi}_r\right)$ and its asymptotic value and the corresponding result for $\sqrt{n}\text{vec}\left(\widehat{\mathbf{\Pi}}_r^{\text{T}} - \widehat{\mathbf{\Pi}}_r^{\text{GCWE}}\right)$; t-distributed data with $\nu = 2$, $p = 10$.



(a) $p = 20$



(b) $p = 50$

Figure 3.16 – Empirical and asymptotic MSE on the projectors of Student's and Tyler's M-estimator for the SA and GCWE regime; t-distributed data with $\nu = 3$.

asymptotic value given by Eq. (3.36), where $\widehat{\mathbf{\Pi}}_r^{\mathrm{T}}$ is the projector obtained with the Tyler's M-estimator.

One can see that both figures validate the theoretical results proposed by Theorems 3.3.1 and 3.3.2, since all error norms goes to zero when the sample size increases. The projector, as a function of eigenvectors, is also a scale-invariant function of the scatter matrix and thus, its asymptotic distribution depend only on the first scale factor. Moreover, one notes that the Tyler's M-estimator (estimator of the shape matrix) provides also an adequate estimation of the principal subspace, as demonstrated in Figure 3.15.

Figure 3.16 displays the empirical MSE of $\sqrt{n}\mathrm{vec}\left(\widehat{\mathbf{\Pi}}_r^M - \mathbf{\Pi}_r\right)$ and $\sqrt{n}\mathrm{vec}\left(\widehat{\mathbf{\Pi}}_r^M - \widehat{\mathbf{\Pi}}_r^{\mathrm{GCWE}}\right)$ and their corresponding theoretical values, i.e., $\mathrm{Tr}(\vartheta_1 \mathbf{\Gamma}_{\mathbf{\Pi}})/n$ (Theorem 3.3.1) and $\mathrm{Tr}(\sigma_1 \mathbf{\Gamma}_{\mathbf{\Pi}})/n$ (Theorem 3.3.2), for both Student's and Tyler's M-estimator. The data is t-distributed with $\nu = 3$.

One can note that the empirical MSEs tend to the proposed asymptotic values. Then, once again, we observe that the MSE is significantly smaller in the GCWE regime than in the SA. The conclusion are valid for both $p = 20$ and $p = 50$, with slightly higher difference for higher value of p.

Thanks to theses results, one can approximate the behaviour of the M-estimatoes' principal subspace with the ones of the GCWE. In the following part, we will see where these results, together with ones from Section 3.1, can be applied.

3.3.3 Number of sources detection / Rank estimation

The rank estimation, or more generally the model order selection, is an important problem in data analysis. It consists in determining r when the covariance of the data is low-rank structured as in Eq. (3.30). In the context of Gaussian distributed samples, several rank estimators have been proposed as functions of the eigenvalues of the SCM $\widehat{\boldsymbol{\lambda}}^{\mathrm{SCM}} = [\widehat{\lambda}_1, \ldots, \widehat{\lambda}_p]$ [161]. Notably, we can cite two of the most commonly used:

- The Akaike Information Criterion (AIC) [4], that minimizes the following criterion

$$\widehat{r}_{\mathrm{AIC}} = \underset{k \in [0, p-1]}{\operatorname{argmin}} \left[n(p-k) \times ic(\widehat{\boldsymbol{\lambda}}^{\mathrm{SCM}}) + k(2p-k) \right] \tag{3.37}$$

  with

$$ic(\widehat{\boldsymbol{\lambda}}^{\mathrm{SCM}}) = \ln \left(\frac{\prod_{i=k}^{p-1} \widehat{\lambda}_i^{\frac{1}{l-1}}}{\frac{1}{l}\sum_{i=k}^{p-1} \widehat{\lambda}_i} \right). \tag{3.38}$$

- The Minimum Description Length (MDL) [180] (also referred to as Bayesian Information Criterion (BIC)), that minimizes the following criterion

$$\widehat{r}_{\mathrm{MDL}} = \underset{k \in [0, p-1]}{\operatorname{argmin}} \left[2n(p-k) \times ic(\widehat{\boldsymbol{\lambda}}^{\mathrm{SCM}}) + k(2p-k)\ln(n) \right]. \tag{3.39}$$

In the context of CES-distributed samples, a "plug-in" approach using $ic(\widehat{\boldsymbol{\lambda}}^{\mathrm{M}})$ (computed from the EVD of an M-estimator) in Eqs. (3.37) or (3.39) can be envisioned rather that re-deriving information criterions assuming non-Gaussian samples. This approach is motivated by the fact that $\widehat{\boldsymbol{\lambda}}^{\mathrm{M}}$ quickly converges to eigenvalues of an equivalent Wishart model (cf. Theorem 3.3.2 and Remark 3.1.2), i.e. the problem can be processed as if the

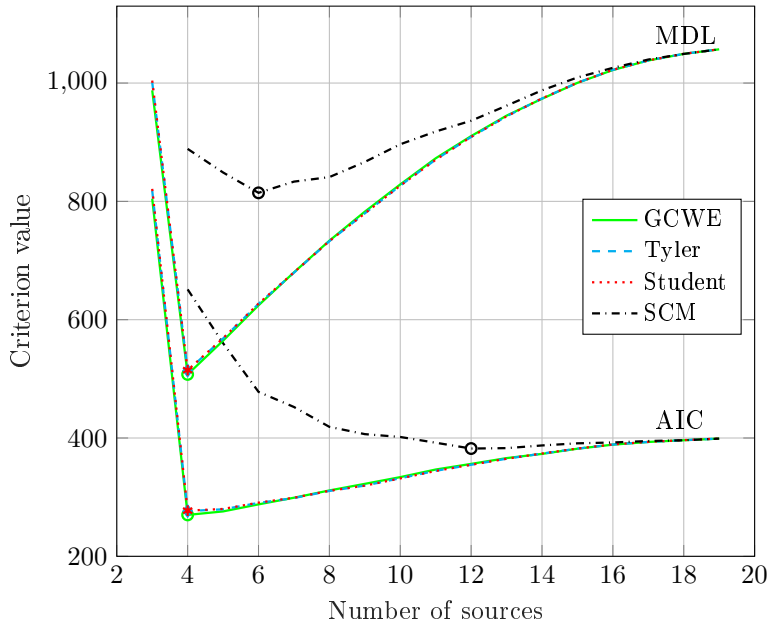


Figure 3.17 – Number of sources estimation with AIC and MDL: Results obtained with Eqs. (3.37) and (3.39) for the SCM, Student's M-estimator and Tyler's M-estimator compared to the theoretical GCWE; Student t-distributed data with $\nu = 2.1$; $p = 20$, $r = 4$.

data were initially Gaussian.

To illustrate this point, we consider the problem of determining how many sources are observed by an array of sensors. We assume that a planar array of p sensors observes signals produced by r sources that are centered around a known frequency ω, affecting the array from locations $\theta_1, \ldots, \theta_{n_s}$. The sources-plus-white-noise signal received by the array of sensors can be expressed as

$$\mathbf{z} = \mathbf{A}\mathbf{s} + \mathbf{n} \tag{3.40}$$

with $\mathbf{s}$ the signal vector, $\mathbf{n}$ the additive white noise and $\mathbf{A} = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_{n_s})]$ and

$$\mathbf{a}(\theta) = \left[a_1(\theta)e^{-j\omega\tau_1(\theta)}, \ldots, a_p(\theta)e^{-j\omega\tau_p(\theta)}\right]^T, \tag{3.41}$$

where a_j the amplitude response of the j^{th} sensor towards direction θ and $\tau_j(\theta)$ propagation delay between the reference point and the j^{th} sensor. The total covariance matrix has thus a low rank structure as in Eq. (3.30), and is given by:

$$\mathbf{\Sigma} = \mathbf{A}\mathbf{S}\mathbf{A}^H + \gamma^2\mathbf{I}. \tag{3.42}$$

where $\mathbb{E}\left[\mathbf{s}\mathbf{s}^H\right] = \mathbf{S}$ and $\mathbb{E}\left[\mathbf{n}\mathbf{n}^H\right] = \gamma^2\mathbf{I}$.

In the considered problem, the received data $\mathbf{z}$ is CES-distributed with a scatter matrix $\mathbf{\Sigma}$ given by Eq. (3.42). We resort to the proposed plug-in approach to estimate r. Figure 3.17 shows the values of the AIC and MLE criteria computed with different M-estimators. The data is t-distributed with $\nu = 2.1$. The number of sensors is set to 20, while the number of sources to estimate is equal to $r = 4$. The number of samples is $n = 200$. A circle indicates the minimum value of each criterion in order to highlight the estimated number r of sources. We observe that Student's and Tyler's M-estimators give an accurate estimation of r. Conversely, the result for the SCM is, as expected, not accurate due to the

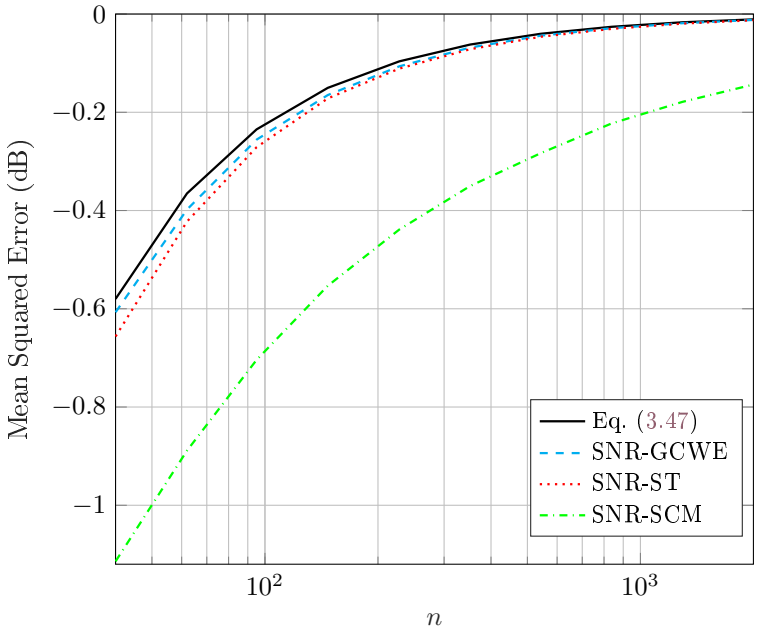


Figure 3.18 – Empirical SNR Loss obtained with the Student's M-estimator (SNR-ST), GCWE (SNR-GCWE) and SCM (SNR-SCM) versus the theoretical result given by Eq. (3.47); t-distributed data with $p = 20$, $r = 5$, $\nu = 3$.

non Gaussianity of the observations. Interestingly, the values of the criteria for Student's and Tyler's M-estimators are almost identical to the ones computed with the (theoretical) GCWE, which validates the proposed approach.

3.3.4 SNR Loss

In the context of STAP, the covariance of the clutter plus noise $\mathbf{\Sigma}$ is low rank structured as in Eq. (3.30), where the rank r can be evaluated thanks to the Brennan rule [20]. The optimal filter $\mathbf{w}_{opt}$ [177] given by

$$\mathbf{w}_{opt} = \mathbf{\Sigma}^{-1}\mathbf{p}, \tag{3.43}$$

where $\mathbf{p}$ is a known steering vector. In the low-rank clutter case an alternative is to use the low-rank STAP filter $\mathbf{w}_R$ [95, 79] defined as

$$\mathbf{w}_r = \mathbf{\Pi}_r^{\perp}\mathbf{p}. \tag{3.44}$$

with $\mathbf{\Pi}_r = \mathcal{R}_r\{\mathbf{\Sigma}\}$ (cf. Eq. (3.31)). In practice, adaptive STAP filters are built with an estimate of the matrix $\mathbf{\Sigma}$ or the projector $\mathbf{\Pi}_r^{\perp}$ computed with secondary data $\mathbf{z}_i \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, \mathbf{g_z})$.

The SNR Loss ρ of an adaptive filter $\widehat{\mathbf{w}}$ is given by

$$\rho = \frac{SNR_{out}}{SNR_{max}} = \frac{|\widehat{\mathbf{w}}^H\mathbf{p}|^2}{(\widehat{\mathbf{w}}^H\mathbf{\Sigma}\widehat{\mathbf{w}})\,(\mathbf{p}^H\mathbf{\Sigma}\mathbf{p})}. \tag{3.45}$$

For an adaptive low rank filter $\widehat{\mathbf{w}}_r$, this expression becomes

$$\rho = \gamma^2\frac{\left(\mathbf{p}^H\widehat{\mathbf{\Pi}}_r^{\perp}\mathbf{p}\right)^2}{\mathbf{p}^H\widehat{\mathbf{\Pi}}_r^{\perp}\mathbf{\Sigma}\widehat{\mathbf{\Pi}}_r^{\perp}\mathbf{p}}. \tag{3.46}$$

with a given estimator $\widehat{\mathbf{\Pi}}_r$. In [79], it has been shown that when the data are Gaussian-distributed and $\widehat{\mathbf{\Pi}}_r = \mathcal{R}_r\{\widehat{\mathbf{\Sigma}}_{\text{SCM}}\}$ the expected SNR Loss is given by

$$\mathbb{E}\left[\rho\right] = 1 - r/n. \tag{3.47}$$

This result can directly provide a good approximation of the expected SNR Loss of adaptive low rank filters build from M-estimators in the general context of CES-distributed samples. Indeed, this approach is again motivated by the fact that $\widehat{\mathbf{\Pi}}_r^M$ quickly converges to principal subspace of an equivalent Wishart model (cf. Theorem 3.3.2). To illustrate this point, Figure 3.18 draws a comparison between the SNR Losses of various low rank filters built from t-distributed data. The low rank covariance matrix is built as in Section 3.3.2. One can notice that the value of SNR-ST is, as expected, very close to the one of SNR-GCWE, which supports the idea to approximate the behavior of SNR-ST with the one of SNR-GCWE [79].

3.4 Conclusion

In this chapter, the estimation of the EVD parameters of the scatter matrix has been addressed. it can be divided in three parts depending on the structure of the scatter matrix.

In Section 3.1, we have analyzed the asymptotic distribution of eigenvalues and eigenvectors of the scatter matrix, when the scatter matrix is assumed to have different eigenvalues.

In Section 3.3, the factor model, i.e., a low-rank scatter matrix, is assumed. The asymptotic distribution of principal subspace has been investigated.

In both cases, the asymptotics for both standard and GCWE regimes have been derived. Interestingly, we have shown that the behavior of the EVD parameters is more accurately characterized by an equivalent Wishart model than by their standard asymptotic Gaussian distribution, thanks to the results of Chapter 2.

Moreover, the possible applications of the theoretical results have been discussed. We have addressed the complex issue of characterizing the intrinsic bias of M-estimators in the CES context. So far, this quantity has been studied for the SCM in the Gaussian context thanks to the distribution of the eigenvalues of a Wishart matrix. Extending this analysis to M-estimators in the general CES context represents, at first sight, an intractable problem because of their unknown exact distribution. However, the established convergence of the eigenvalues of an M-estimator toward their GCWE counterpart allows to derive an accurate approximation of this intrinsic bias. In the context of model order selection (i.e., rank estimation) from non-Gaussian samples, we show that the use of M-estimators (rather than the SCM) in theoretic criteria derived for Gaussian models yields the same results as the one obtained with the theoretical GCWE. Again, this justifies a plug-in approach (using M-estimators in processes derived under the Wishart assumption), instead of a complete re-derivation that would require to assume an exact CES distribution. The performance of low rank filters built from M-estimators are derived in the same way (i.e., approached by the one of their GCWE) to illustrate that the approach also holds for adaptive processes based on the eigenvectors.

Finally, Section 3.2 is somewhat different. In this section, we assumed that the covariance matrix of Gaussian-distributed data possesses few different eigenvalues. We have introduced a new regularized covariance matrix estimator based on a novel eFusion penalty that promotes similarity and grouping of eigenvalues. The grouping effect is achieved as large gaps between successive eigenvalues are not penalized excessively. This feature is obtained by utilizing Tukey's function as the penalty function for the gaps. The important topic of how to choose the tuning parameter c of Tukey's function has been addressed. The extension to the CES-data has been introduced. In this case, the method could be presented as *"two-steps robust"*, since we first use a robust M-estimator to pre-estimate the eigenvalues and then apply the robust Tukey's function in order to group them and get the final estimates. The

main benefits of the eFusion are unbiasedness (accurate grouping) and robustness to the choice of the penalty parameter. Future works will focus on a generalization to $p > n$ case.

The conclusions drawn in this chapter are in perfect line with the previous claims of this thesis: *Always prefer robust methods because they offer an important gain in terms of robustness to outliers, heterogeneous data, missing data, etc. Their performances can now be well-characterized, with a minor loss, using the Gaussian-based optimal processing methods.*

Chapter 4

Robust detection

This chapter is dedicated to robust signal detection. We analyze the asymptotic performances of the robust Mahalanobis distance and of various robust adaptive detectors in the context of non-Gaussian observations. We focus on the single steering case in homogeneous environment and analyze the properties of different detectors such as Adaptive (Normalized) Matched Filter (AMF/ANMF), Kelly's GLRT, and Rao test. Furthermore, we analyze the asymptotic performances of the low-rank ANMF which is based on the M-estimators' principal subspaces.

In addition to the standard asymptotic distribution, we extend the results of Chapter 2 and derive the asymptotic distribution in the GCWE regime. In this context, we show that, the distribution of a statistic built with M-estimators can be accurately approximated by the one of the same statistic built with the GCWE. The loss due to this approximation is theoretically derived and shown to be negligible in most cases. This explicit equivalent statistic is especially interesting since it permits to tune robust Mahalanobis distances and robust detectors with well-established results from the Gaussian framework.

Publications associated with this chapter:

[C2] **G. Draškovic**, F. Pascal, A. Breloy and J.-Y. Tourneret, "New asymptotic properties for the robust ANMF," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-17*, New Orleans, USA, March 2017, pp. 3429-3433.

[C7] **G. Draškovic**, F. Pascal, A. Breloy et J.-Y. Tourneret, "Nouvelles propriétés asymptotiques de détecteurs robustes," *colloque GRETSI sur le traitement du signal et des images*, Juan-les-Pins, France, Septembre 2017.

[J3] **G. Draškovic**, A. Breloy and F. Pascal, "On the performance of robust plug-in detectors using M-estimators," *Signal Processing*, vol. 167, pp. 107282, 2020.

4.1 Non-Gaussian detection

In Section 1.2, we have presented the signal detection problem in the Gaussian framework and reviewed the well-known detection statistics. As a core component, these statistics involve the SCM [13] in their construction. Since this estimator of the noise covariance matrix is sensitive to heavy-tailed distributed samples, this family of Gaussian detectors can exhibit poor performance in non-Gaussian environments.

When the data turn to be non-Gaussian, the CES distributions (Definition 1.1.10) are employed as alternative models. Particularly, for radar clutter modeling, the SIRV model (Definition 1.1.12) has been widely used [12, 179, 142]. Detection procedures assuming CES/CG distributed samples have been proposed and widely studied in the literature. First, detection procedures assuming known parameters for the noise can found in [62, 156, 155] among others. Then, adaptive detectors have been derived, generally based on a 2-step GLRT, for different covariance matrix estimators [63, 32, 33]. More recently, various robust covariance matrix estimators have been used for detection purposes [6, 2, 131, 139, 138]. An overview of recent advances in radar detection, including robust detection approaches, can be found in [45].

However, in practice, the density generator of the true underlying distribution is unknown. In this case, a *robust detector* can be built as a classical Gaussian detector [132] where an M-estimator of the scatter [119] is plugged-in instead of the SCM. The study of these robust detection processes is not trivial since M-estimators are expressed as solutions of fixed point equations. So far, only the properties in standard asymptotic regime of M-estimators have been proposed in [169, 117]. These works have permitted to analyze the asymptotic properties of robust detectors in standard regime [139, 138, 117].

In this chapter, we provide an accurate performance analysis of the asymptotic distribution of robust detectors, detectors built with an M-estimator in the CES context, by comparing to a *Gaussian-Core Equivalent Detector* (GCED).

In the following, we first analyze the robust Mahalanobis distance. Besides the signal detection, this measure is widely used in various signal processing and machine learning applications. Then, the statistical properties of full and low-rank robust detectors are investigated.

4.2 Mahalanobis distance

The Mahalanobis distance [115, 113] is one of the most common measures in multivariate statistics and signal processing. It is based on the correlation between variables thanks to which different models can be identified and analyzed. The Mahalanobis distance of $\mathbf{z}$ from $\boldsymbol{\mu}$ is given by $\Delta(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where

$$\Delta^2(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (\mathbf{z} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \boldsymbol{\mu}), \tag{4.1}$$

where $\boldsymbol{\mu}$ is the population mean and $\boldsymbol{\Sigma}$ is the common scatter matrix. Since we work with vectors with known mean, we will w.l.o.g. analyze $\Delta^2(\boldsymbol{\Sigma}) = \mathbf{z}^H \boldsymbol{\Sigma}^{-1} \mathbf{z}$. If the data are normal distributed, $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$, and the distance is based on the true scatter matrix $\boldsymbol{\Sigma}$, then it follows a scaled chi-squared distribution

$$\Delta^2(\boldsymbol{\Sigma}) \sim (1/2)\, \chi^2_{2p}. \tag{4.2}$$

Since the scatter matrix is usually unknown, the distance is computed with an estimator. If the SCM is plugged in instead of the true scatter matrix and under the Gaussian assumption,

the distance becomes β'-distributed [1] with an asymptotic chi-squared distribution

$$\Delta^2\left(\widehat{\mathbf{\Sigma}}_{\text{SCM}}\right) \sim n\beta'\left(p, n-p+1\right), \tag{4.3}$$

where $\beta'(a,b)$ denotes a Beta prime distribution with real shape parameters a and b.

Beside testing if an observed random sample is from a multivariate normal distribution (detecting outliers) [153, 78], the Mahalanobis distance is also a useful way to determine similarities between sets of known and unknown data. Thus, it is widely used in classification problems [184, 181], feature selection problems [145], anomaly detection in hyperspectral imaging [24, 60], etc.

4.2.1 Asymptotic of the robust Mahalanobis distance

The object of our study is to analyze robust Mahalanobis distances, i.e. distances computed with M-estimators, comparing it to the one based on the GCWE (Definition 2.1.2). To that end, let us first state two definitions.

Definition 4.2.1: Robust Mahalanobis distance

Consider a set of $n+1$ samples $(\mathbf{z}, \mathbf{z}_1, \ldots, \mathbf{z}_n)$ drawn as $\mathbf{z}_i \sim \mathcal{CES}\left(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}}\right)$. Let $\widehat{\mathbf{\Sigma}}$ be an M-estimator as in Eq. (1.28) built from $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ and $\Delta = \Delta(\mathbf{\Sigma})$ be the Mahalanobis distance defined by Eq. (4.1). The distance $\Delta_M = \Delta(\sigma\widehat{\mathbf{\Sigma}})$, that uses the M-estimator $\widehat{\mathbf{\Sigma}}$ instead of the traditional SCM, is referred to as a robust Mahalanobis distance.

Definition 4.2.2: Gaussian-Core Equivalent Mahalanobis (GCEM)

Consider a set of $n+1$ samples $(\mathbf{z}, \mathbf{z}_1, \ldots, \mathbf{z}_n)$ drawn as $\mathbf{z}_i \sim \mathcal{CES}\left(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}}\right)$. Let Δ_M be the robust Mahalanobis distance as in Definition 4.2.1. Let $\widehat{\mathbf{\Sigma}}_{\text{GCWE}}$ be the GCWE of $\widehat{\mathbf{\Sigma}}$ (cf. Definition 2.1.2). The quantity $\Delta_{\text{GCEM}} = \Delta(\widehat{\mathbf{\Sigma}}_{\text{GCWE}})$ is referred to as Gaussian-Core Equivalent Mahalanobis distance (GCEM) of Δ_M.

Once we have defined the robust Mahalanobis distance and its GCEM, we can derive the asymptotic distribution of the robust Mahalanobis distance Δ_M defined in Definition 4.2.1. However, before stating the theorem, let us recall that the distribution of Δ_M depends both on the distribution of $\mathbf{z}$ and of $\widehat{\mathbf{\Sigma}}$. Hereafter, we derive the asymptotic distribution of the robust Mahalanobis distance conditionally to $\mathbf{z}$, i.e. taking into account only the asymptotic distribution of $\widehat{\mathbf{\Sigma}}$. However, when analyzing the exact distribution of Δ_M, one must consider both the distribution of $\mathbf{z}$ and of $\widehat{\mathbf{\Sigma}}$.

Theorem 4.2.3: Robust Mahalanobis distance [J1]

Let Δ_M be the robust Mahalanobis distance defined in Definition 4.2.1 and Δ_{GCEM} its GCEM defined in Definition 4.2.2. Then, the asymptotic conditional distribution of Δ_M^2 is given by

$$\textbf{(SA)} \qquad \sqrt{n}\left(\Delta_M^2 - \Delta^2\right)_{\mathbf{z}} \xrightarrow{d} \mathcal{N}\left(0, (\vartheta_1 + \vartheta_2)\Delta^4\right), \tag{4.4}$$

$$\textbf{(GCEM)} \qquad \sqrt{n}\left(\Delta_M^2 - \Delta_{\text{GCEM}}^2\right)_{\mathbf{z}} \xrightarrow{d} \mathcal{N}\left(0, (\sigma_1 + \sigma_2)\Delta^4\right), \tag{4.5}$$

where the notation $(.)_{\mathbf{z}}$ stresses the conditional distribution to $\mathbf{z}$, ϑ_1 and ϑ_2 are given by Eq. (1.33) and σ_1 and σ_2 are given by Eq. (2.5).

1. Beta prime distribution corresponds to a scaled F-distribution.

Remark 4.2.3

- The asymptotic variance of the robust Mahalanobis distance when centering around its GCEM is smaller than the one in the SA regime, i.e. when centering around the distance based on the true scatter matrix, since $\sigma_1 + \sigma_2 < \vartheta_1 + \vartheta_2$. The results are valid even for small values of n, which will be demonstrated in the simulation part.
- As many results in this thesis, these findings reveal that the distribution of the robust (squared) Mahalanobis distance is better approximated with a scaled Beta prime distribution (Eq. (4.3)) than with a scaled chi-squared distribution (Eq. (4.2)).

Proof. Following the standard idea of this thesis, we can write

$$\mathbb{E}\left[n\left(\Delta_M^2 - \Delta_{\text{GCEM}}^2\right)^2\right] = (\phi_M^{(n)} - 2\phi_{M-\text{GCWE}}^{(n)} + \phi_{\text{GCWE}}^{(n)}), \tag{4.6}$$

where

$$\begin{aligned}
\phi_M^{(n)} &= \mathbb{E}\left[n\left(\Delta_M^2 - \Delta^2\right)^2\right], \\
\phi_{M-\text{GCWE}}^{(n)} &= \mathbb{E}\left[n\left(\Delta_M^2 - \Delta^2\right)\left(\Delta_{\text{GCEM}}^2 - \Delta^2\right)\right], \\
\phi_{\text{GCWE}}^{(n)} &= \mathbb{E}\left[n\left(\Delta_{\text{GCEM}}^2 - \Delta^2\right)^2\right].
\end{aligned}$$

Using the Delta method and Theorem 1.1.3, one can obtain

$$\phi_M^{(n)} \xrightarrow[n\to+\infty]{} \phi_M = \left(\Delta^2\right)' \mathbf{\Gamma}_1 \left(\left(\Delta^2\right)'\right)^H, \tag{4.7}$$

where $\left(\Delta^2\right)' = \partial\Delta^2/\partial\left(\text{vec}(\mathbf{\Sigma})\right)$ and $\mathbf{\Gamma}_1$ is defined by Eq. (1.32) (with $\mathbf{\Sigma} = \sigma\mathbf{\Sigma}_\sigma$).

Moreover, one has a more general result

$$\phi_{M-\text{GCWE}}^{(n)} \xrightarrow[n\to+\infty]{} \left(\Delta^2\right)' \mathbf{\Gamma}_2 \left(\left(\Delta^2\right)'\right)^H, \tag{4.8}$$

where $\mathbf{\Gamma}_2 = \gamma_1\left(\mathbf{\Sigma}^T \otimes \mathbf{\Sigma}\right) + \gamma_2 \text{vec}\left(\mathbf{\Sigma}\right)\text{vec}\left(\mathbf{\Sigma}\right)^H$ with γ_1 and γ_2 are given in Eq. (D.7). In [140] it has been shown that $\left(\Delta^2\right)' = \text{vec}^H\left(\mathbf{z}\mathbf{z}^H\right)\left(\mathbf{\Sigma}^T \otimes \mathbf{\Sigma}\right)^{-1}$. From Eqs. (A.9) and Eq. (A.1), one has $\left(\mathbf{\Sigma}^T \otimes \mathbf{\Sigma}\right)^{-1}\text{vec}\left(\mathbf{\Sigma}\right) = \text{vec}\left(\mathbf{\Sigma}^{-1}\right)$. Consequently, one obtains

$$\begin{aligned}
\phi_M &= \left(\Delta^2\right)' \mathbf{\Gamma}_1 \left(\left(\Delta^2\right)'\right)^H \\
&= \vartheta_1 \text{vec}^H\left(\mathbf{z}\mathbf{z}^H\right)\left(\mathbf{\Sigma}^T \otimes \mathbf{\Sigma}\right)^{-1}\text{vec}\left(\mathbf{z}\mathbf{z}^H\right) \\
&+ \vartheta_2 \text{vec}^H\left(\mathbf{z}\mathbf{z}^H\right)\text{vec}\left(\mathbf{\Sigma}^{-1}\right)\text{vec}\left(\mathbf{\Sigma}^{-1}\right)^H\text{vec}\left(\mathbf{z}\mathbf{z}^H\right) \\
&= \vartheta_1 \text{vec}^H\left(\mathbf{z}\mathbf{z}^H\right)\text{vec}\left(\mathbf{\Sigma}^{-1}\mathbf{z}\mathbf{z}^H\mathbf{\Sigma}^{-1}\right) \\
&+ \vartheta_2 \text{Tr}(\mathbf{z}\mathbf{z}^H\mathbf{\Sigma}^{-1})\text{Tr}(\mathbf{\Sigma}^{-1}\mathbf{z}\mathbf{z}^H) \\
&= \vartheta_1 \text{Tr}(\mathbf{z}^H\mathbf{\Sigma}^{-1}\mathbf{z}\mathbf{z}^H\mathbf{\Sigma}^{-1}\mathbf{z}) + \vartheta_2 \text{Tr}(\mathbf{z}^H\mathbf{\Sigma}^{-1}\mathbf{z})^2 \\
&= (\vartheta_1 + \vartheta_2)(\mathbf{z}^H\mathbf{\Sigma}^{-1}\mathbf{z})^2 = (\vartheta_1 + \vartheta_2)\Delta^4,
\end{aligned}$$

proving the result of Eq. (4.4). It is now clear that $\phi_{\text{GCWE}} = \Delta^4$ and $\phi_{M-\text{GCWE}} = \left(\Delta^2\right)'\mathbf{\Gamma}_2\left(\left(\Delta^2\right)'\right)^H = (\gamma_1 + \gamma_2)\Delta^4$ which leads to the final result in Eq. (4.5). □

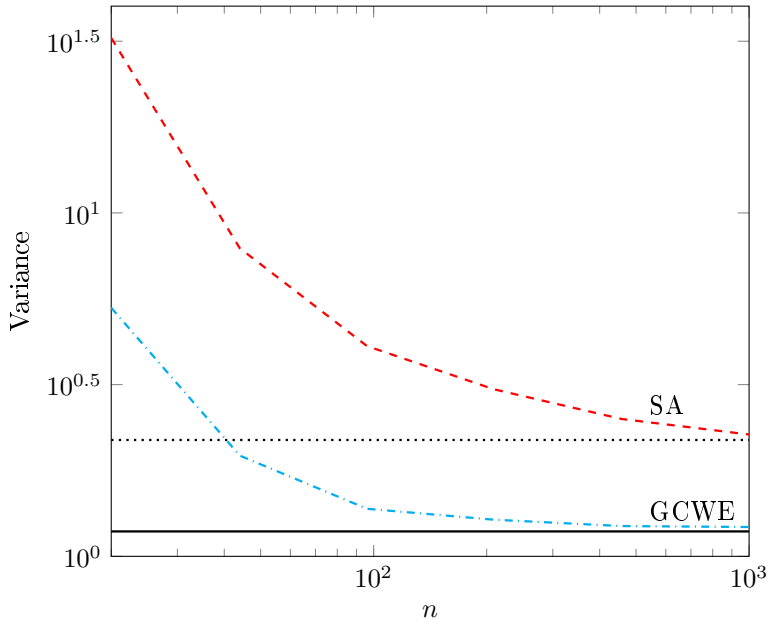


Figure 4.1 – Scaled empirical variance of the robust Mahalanobis distance in the SA regime (red curve) and when centered around its GCEM (blue curve), compared to the theoretical results (Theorem 4.2.3).

4.2.2 Experiments

Hereafter, we provide some simulations in order to validate the theoretical result.

Figure 4.1 displays the results for Student's M-estimator are presented. The data follow the complex t-distribution with $\nu = 2$. The empirical variance of the robust distance and the one of the difference between the robust distance and the GCEM (compared to the theoretical result of Theorem 4.2.3) are plotted.

One can notice that the value of the robust distance is closer to its GCEM, than to the distance computed with the true scatter matrix. This implies that the distribution of robust distances can be better approximated with the theoretical distribution of the GCEM than with the asymptotic distribution based on the true scatter matrix. Figure 4.2 illustrates this point.

Figure 4.2 plots the empirical distribution of the robust Mahalanobis distance built with the Student's M-estimator and two corresponding distributions proposed in Eqs. (4.2) and (4.3) for t-distributed data with $\nu = 2$, $p = 10$ and $n = 100$. We observe that the empirical distribution matches significantly better the scaled Beta prime than the scaled chi-squared distribution. The essential advantage of these findings in, for instance, outlier detection is that they support the idea to use the robust M-estimators to estimate the scatter matrix and to rely on the theoretical distribution of the Wishart-based distance when computing the detection threshold.

In the following part, we extend these results to the signal detection problem.

4.3 Robust full-rank signal detection

We consider the problem of signal detection defined in Section 1.2 with the assumption that the secondary data is non-Gaussian distributed. The objective is again to detect the

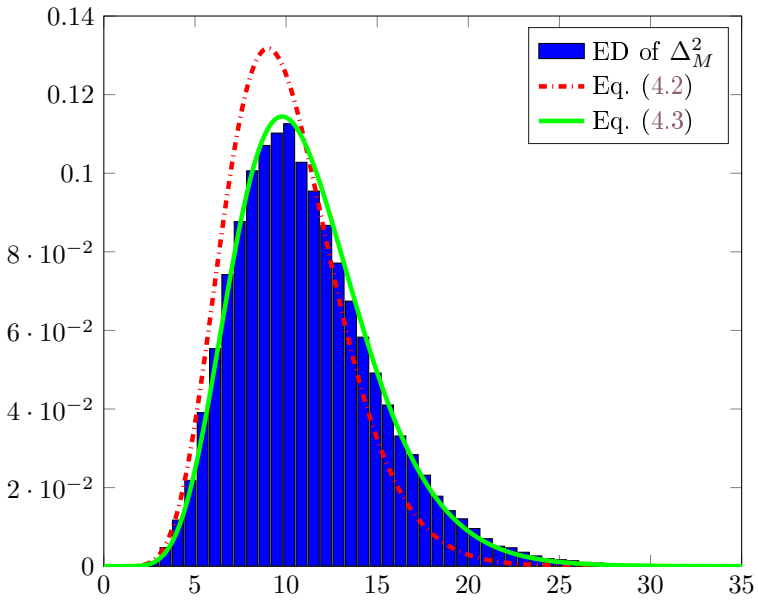


Figure 4.2 – Empirical distribution of the t-robust Mahalanobis distance versus the asymptotic distribution (Eq. (4.2)) and theoretical approximative distribution (Eq. (4.3)); t-distributed data with $\nu = 2$, $n = 100$, $p = 10$.

complex signal $\mathbf{p}$, from the received data $\mathbf{z} = \alpha\mathbf{p} + \mathbf{c}$, where α is an unknown deterministic parameter. The binary hypothesis test is given by

$$\begin{cases} H_0 : \mathbf{z} = \mathbf{c} & \mathbf{z}_i = \mathbf{c}_i, \quad i = 1, \ldots, n, \\ H_1 : \mathbf{z} = \alpha\mathbf{p} + \mathbf{c} & \mathbf{z}_i = \mathbf{c}_i, \quad i = 1, \ldots, n, \end{cases} \tag{4.9}$$

where now $\mathbf{c}_i$ are i.i.d. with $\mathbf{c}_i \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}})$. These vectors are used to estimate the unknown scatter matrix $\mathbf{\Sigma}$.

Then, we can define a robust detector.

Definition 4.3.1: Robust detector

Consider a set of $n+1$ samples $(\mathbf{z}, \mathbf{z}_1, \ldots, \mathbf{z}_n)$ drawn as $\mathbf{z}_i \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}})$ with the detection problem expressed in Eq. (4.9). Let $\widehat{\mathbf{\Sigma}}$ be an M-estimator as in Eq. (1.28) built from $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ and $\Lambda(\cdot)$ be a decision statistic (either Kelly in Eq. (1.45), AMF in Eq. (1.49), ANMF in Eq. (1.52), or Rao in Eq. (1.59)). The statistic $\widehat{\Lambda}^M = \Lambda(\sigma\widehat{\mathbf{\Sigma}})$, that uses the M-estimator $\widehat{\mathbf{\Sigma}}$ as a "plug-in" instead of the traditional SCM, is referred to as a robust detector.

Note that the ANMF can be used only with $\widehat{\mathbf{\Sigma}}$ since it is a scale-invariant function of the scatter matrix.

Thanks to the GCWE (Definition 2.1.2) we can now define Gaussian-Core Equivalent Detector (GCED) of a robust detector.

Definition 4.3.2: Gaussian-Core Equivalent Detector (GCED)

Consider a set of $n+1$ samples $(\mathbf{z}, \mathbf{z}_1, \ldots, \mathbf{z}_n)$ drawn as $\mathbf{z}_i \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}})$ with the detection problem expressed in Eq. (4.9). Let $\Lambda(\sigma\widehat{\mathbf{\Sigma}})$ be a robust detector as in Definition 4.3.1. Let $\widehat{\mathbf{\Sigma}}_{\text{GCWE}}$ be the GCWE of $\widehat{\mathbf{\Sigma}}$ (cf. Definition 2.1.2). The quantity $\widehat{\Lambda}^{\text{GCED}} = \Lambda(\widehat{\mathbf{\Sigma}}_{\text{GCWE}})$ is referred to as Gaussian-Core Equivalent Detector (GCED) of $\widehat{\Lambda}^M$.

In the following, we propose several theorems to characterize the distribution of robust detectors (as in Definition 4.3.1) conditionally to the tested sample $\mathbf{z}$. First, the results are obtained for the standard asymptotic, then the convergence towards the GCED are derived.

4.3.1 Asymptotics of robust detectors

Before deriving the asymptotic distribution of robust detectors, please note that $\widehat{\Lambda}_{\text{Kelly}}$, $\widehat{\Lambda}_{\text{Rao}}$ and $\widehat{\Lambda}_{\text{AMF}}$ defined in Eqs. (1.45), (1.59) and (1.49), respectively, go to zero when $n \to \infty$, as well as their variances. Thus, in order to compare these detectors to their GCEDs, we will analyze the statistics multiplied by n, i.e., $n\widehat{\Lambda}_{\text{Kelly}}$, $n\widehat{\Lambda}_{\text{Rao}}$ and $n\widehat{\Lambda}_{\text{AMF}}$. For the sake of simplicity, from now on, we refer to these quantities as $\widehat{\Lambda}_{\text{Kelly}}$, $\widehat{\Lambda}_{\text{Rao}}$ and $\widehat{\Lambda}_{\text{AMF}}$. The corresponding GCEDs are defined w.r.t. these quantities.

In the SA regime, the robust adaptive detector is compared to its non-adaptive form, i.e. detector built using the true SM. For $\widehat{\Lambda}_{\text{Kelly}}$, $\widehat{\Lambda}_{\text{Rao}}$ and $\widehat{\Lambda}_{\text{AMF}}$, one obtains the following results.

Theorem 4.3.3: SA of the robust AMF, Kelly and Rao [J3]

Let $\widehat{\Lambda}^M$ be a scaled robust detector as in Definition 4.3.1 with $\widehat{\Lambda}^M \in \{\widehat{\Lambda}_{\text{Kelly}}, \widehat{\Lambda}_{\text{Rao}}, \widehat{\Lambda}_{\text{AMF}}\}$ and $\Lambda = \Lambda(\mathbf{\Sigma})$. Conditionally to the distribution of $\mathbf{z}$, the asymptotic distribution of $\widehat{\Lambda}^M$ is given by

$$\sqrt{n}\left(\widehat{\Lambda}^M - \Lambda\right)_{\mathbf{z}} \xrightarrow{d} \mathcal{N}\left(0, \vartheta_1\sigma_{\mathcal{X}} + \vartheta_2\Lambda^2\right) \tag{4.10}$$

with $\sigma_{\mathcal{X}} = \Lambda\left(2\mathbf{z}^H\mathbf{\Sigma}^{-1}\mathbf{z} - \Lambda\right)$ and ϑ_1,ϑ_2 defined in Eq. (1.33).

Proves of all theorems are given together after Theorem 4.3.6 (see Proof 4.3.1).

Note that, similarly to Mahalanobis distance, the distribution of robust detectors depends both on the distribution of $\mathbf{z}$ and of $\widehat{\mathbf{\Sigma}}$. As previously, we analyze the asymptotic distribution of robust detectors conditionally to $\mathbf{z}$.

Due to the specific scale invariance property, the variance of the robust ANMF has a particular form (depending only on the first scale factor) derived in [139]. Here, we recall this result in order to compare it with the result in the GCED regime that will be derived later on.

Theorem 4.3.4: SA of the robust ANMF

Let $\widehat{\Lambda}^{M}_{\text{ANMF}}$ be the robust ANMF detector as in Definition 4.3.1 and Λ_{NMF} the statistics computed with $\mathbf{\Sigma}$. Conditionally to the distributions of $\mathbf{z}$, the asymptotic distribution of $\widehat{\Lambda}^{M}_{\text{ANMF}}$ is given by [139]

$$\sqrt{n}\left(\widehat{\Lambda}^{M}_{\text{ANMF}} - \Lambda_{\text{NMF}}\right)_{\mathbf{z}} \xrightarrow{d} \mathcal{N}\left(0, \vartheta_1 \sigma_{\mathcal{H}}\right), \tag{4.11}$$

where $\sigma_{\mathcal{H}} = 2\Lambda_{\text{NMF}}(\Lambda_{\text{NMF}} - 1)^2$ and ϑ_1 is defined in Eq. (1.33).

Thanks to the results for the SA regime, we can obtain the corresponding ones for the convergence towards the GCED.

Theorem 4.3.5: GCED of the robust AMF, Kelly and Rao [J3]

Let $\widehat{\Lambda}^{M}$ be a scaled robust detector as in Definition 4.3.1 with $\widehat{\Lambda}^{M} \in \{\widehat{\Lambda}_{\text{Kelly}}, \widehat{\Lambda}_{\text{Rao}}, \widehat{\Lambda}_{\text{AMF}}\}$. Let $\widehat{\Lambda}^{\text{GCED}}$ be the GCED of $\widehat{\Lambda}^{M}$ defined in Definition 4.3.2. Then, conditionally to the distribution of $\mathbf{z}$, the asymptotic of $\widehat{\Lambda}^{M} - \widehat{\Lambda}^{\text{GCED}}$ is given by

$$\sqrt{n}\left(\widehat{\Lambda}^{M} - \widehat{\Lambda}^{\text{GCED}}\right)_{\mathbf{z}} \xrightarrow{d} \mathcal{N}\left(0, \sigma_1 \sigma_{\mathcal{X}} + \sigma_2 \Lambda^2\right) \tag{4.12}$$

with $\sigma_{\mathcal{X}} = \Lambda\left(2\mathbf{z}^H \mathbf{\Sigma}^{-1}\mathbf{z} - \Lambda\right)$ and σ_1, σ_2 defined in Eq. (2.5).

Theorem 4.3.6: GCED of the robust ANMF [J3]

Let $\widehat{\Lambda}^{M}_{\text{ANMF}}$ be the robust ANMF detector as in Definition 4.3.1. Conditionally to the distributions of $\mathbf{z}$, the asymptotic distribution of $\widehat{\Lambda}^{M}_{\text{ANMF}} - \widehat{\Lambda}^{\text{GCED}}_{\text{ANMF}}$ is given by

$$\sqrt{n}\left(\widehat{\Lambda}^{M}_{\text{ANMF}} - \widehat{\Lambda}^{\text{GCED}}_{\text{ANMF}}\right)_{\mathbf{z}} \xrightarrow{d} \mathcal{N}\left(0, \sigma_1 \sigma_{\mathcal{H}}\right) \tag{4.13}$$

where $\sigma_{\mathcal{H}} = 2\Lambda_{\text{NMF}}(\Lambda_{\text{NMF}} - 1)^2$ and σ_1 defined in Eq. (2.5).

Proof. In order to prove all theorems, we first provide common steps for all detectors, where the detection test is noted as $\Lambda(\cdot)$. Using the standard steps, one can show that one obtains

$$\begin{aligned}
\sigma^{(n)}_{\Lambda} &= \mathbb{E}\left[n\left(\widehat{\Lambda}^{M} - \widehat{\Lambda}^{\text{GCED}}\right)\left(\widehat{\Lambda}^{M} - \widehat{\Lambda}^{\text{GCED}}\right)^H\right] \\
&= \sigma^{(n)}_{\Lambda 1} - 2\sigma^{(n)}_{\Lambda 2} + \sigma^{(n)}_{\Lambda 3}
\end{aligned}$$

with

$$\begin{aligned}
\sigma^{(n)}_{\Lambda 1} &= \mathbb{E}\left[n\left(\widehat{\Lambda}^{M} - \Lambda\right)\left(\widehat{\Lambda}^{M} - \Lambda\right)^H\right], \\
\sigma^{(n)}_{\Lambda 2} &= \mathbb{E}\left[n\left(\widehat{\Lambda}^{M} - \Lambda\right)\left(\widehat{\Lambda}^{\text{GCED}} - \Lambda\right)^H\right], \\
\sigma^{(n)}_{\Lambda 3} &= \mathbb{E}\left[n\left(\widehat{\Lambda}^{\text{GCED}} - \Lambda\right)\left(\widehat{\Lambda}^{\text{GCED}} - \Lambda\right)^H\right].
\end{aligned}$$

Thanks to the Lemma 3.1.1, one has

$$\begin{aligned}\sigma_{\Lambda 1}^{(n)} &\xrightarrow[n\to+\infty]{} \Lambda' \mathbf{\Gamma}_1 \left(\Lambda'\right)^H, \\ \sigma_{\Lambda 2}^{(n)} &\xrightarrow[n\to+\infty]{} \Lambda' \mathbf{\Gamma}_2 \left(\Lambda'\right)^H, \\ \sigma_{\Lambda 3}^{(n)} &\xrightarrow[n\to+\infty]{} \Lambda' \mathbf{\Gamma}_3 \left(\Lambda'\right)^H.\end{aligned}$$

where $\Lambda' = \dfrac{\partial \Lambda(\mathbf{\Sigma})}{\partial \mathrm{vec}(\mathbf{\Sigma})}$ and $\mathbf{\Gamma}_1$, $\mathbf{\Gamma}_2$ and $\mathbf{\Gamma}_3$ are given by Eqs. (1.32), (D.8) and (1.11), respectively. This leads to

$$\sigma_{\Lambda}^{(n)} \xrightarrow[n\to+\infty]{} \Lambda' \mathbf{\Gamma} \left(\Lambda'\right)^H \tag{4.14}$$

where $\mathbf{\Gamma}$ is defined in Eq. (2.4).

Using the previous equations and deriving the derivatives Λ', one can obtain the final results for all detectors.

In order to obtain derivatives, we will start with Rao statistic Λ'_{Rao} since the derivatives for other detectors can be easily obtained from this one.

Let us first rewrite Λ_{Rao} as

$$\Lambda_{\mathrm{Rao}} = \frac{a}{(1+\frac{1}{n}b)(1+\frac{1}{n}b+\frac{1}{n}a)}$$

with $a = \left|\mathbf{p}^H \mathbf{\Sigma}^{-1} \mathbf{z}\right|^2 / \left(\mathbf{p}^H \mathbf{\Sigma}^{-1} \mathbf{p}\right)$ and $b = \mathbf{z}^H \mathbf{\Sigma}^{-1} \mathbf{z}$. Then, one has

$$\begin{aligned}\bullet \text{ Rao:} \quad \partial\Lambda_{\mathrm{Rao}} &= \frac{\partial a(1+\frac{1}{n}b)(1+\frac{1}{n}b+\frac{1}{n}a) - a(\frac{1}{n}\partial b)(1+\frac{1}{n}b+\frac{1}{n}a) - a(1+\frac{1}{n}b)(\frac{1}{n}\partial b+\frac{1}{n}\partial a)}{(1+\frac{1}{n}b)^2(1+\frac{1}{n}b+\frac{1}{n}a)^2} \\ &= \Lambda_{\mathrm{Rao}}\left(\frac{\partial a}{a} - \frac{\frac{1}{n}\partial b}{1+\frac{1}{n}b} - \frac{\frac{1}{n}\partial b+\frac{1}{n}\partial a}{1+\frac{1}{n}b+\frac{1}{n}a}\right).\end{aligned}$$

Now

$$\begin{aligned}\partial a &= \frac{\partial\left(\mathbf{p}^H \mathbf{\Sigma}^{-1} \mathbf{z}\right)\left(\mathbf{z}^H \mathbf{\Sigma}^{-1} \mathbf{p}\right)\left(\mathbf{p}^H \mathbf{\Sigma}^{-1} \mathbf{p}\right)}{\left(\mathbf{p}^H \mathbf{\Sigma}^{-1} \mathbf{p}\right)^2} \\ &- \frac{\left(\mathbf{p}^H \mathbf{\Sigma}^{-1} \mathbf{z}\right)\partial\left(\mathbf{z}^H \mathbf{\Sigma}^{-1} \mathbf{p}\right)\left(\mathbf{p}^H \mathbf{\Sigma}^{-1} \mathbf{p}\right)}{\left(\mathbf{p}^H \mathbf{\Sigma}^{-1} \mathbf{p}\right)^2} \\ &- \frac{\left(\mathbf{p}^H \mathbf{\Sigma}^{-1} \mathbf{z}\right)\left(\mathbf{z}^H \mathbf{\Sigma}^{-1} \mathbf{p}\right)\partial\left(\mathbf{p}^H \mathbf{\Sigma}^{-1} \mathbf{p}\right)}{\left(\mathbf{p}^H \mathbf{\Sigma}^{-1} \mathbf{p}\right)^2}.\end{aligned}$$

By using Eqs. (A.5), (A.8), (A.9) and (A.1), one can show that

$$\partial\left(\mathbf{p}^H \mathbf{\Sigma}^{-1} \mathbf{z}\right) = -\partial\left(\mathrm{vec}\left(\mathbf{\Sigma}\right)^H\right)\left(\mathbf{\Sigma}^T \otimes \mathbf{\Sigma}\right)^{-1} \mathrm{vec}\left(\mathbf{z}\mathbf{p}^H\right)$$

and thus

$$\partial a = -\partial\left(\mathrm{vec}\left(\mathbf{\Sigma}\right)^H\right)\left(\mathbf{\Sigma}^T \otimes \mathbf{\Sigma}\right)^{-1} a\left(\frac{\mathbf{z}\mathbf{p}^H}{\mathbf{p}^H \mathbf{\Sigma}^{-1} \mathbf{z}} + \frac{\mathbf{p}\mathbf{z}^H}{\mathbf{z}^H \mathbf{\Sigma}^{-1} \mathbf{p}} - \frac{\mathbf{p}\mathbf{p}^H}{\mathbf{p}^H \mathbf{\Sigma}^{-1} \mathbf{p}}\right)$$

and

$$\partial b = -\partial\left(\mathrm{vec}\left(\mathbf{\Sigma}\right)^H\right)\left(\mathbf{\Sigma}^T \otimes \mathbf{\Sigma}\right)^{-1} \mathrm{vec}\left(\mathbf{z}\mathbf{z}^H\right).$$

Analogously,

- Kelly: $\partial\Lambda_{\mathrm{Kelly}} = \Lambda_{\mathrm{Kelly}}\left(\dfrac{\partial a}{a} + \dfrac{\frac{1}{n}\partial b}{1+\frac{1}{n}b}\right)$,
- AMF: $\partial\Lambda_{\mathrm{AMF}} = \Lambda_{\mathrm{AMF}}\dfrac{\partial a}{a}$.

One can note that

$$\frac{\frac{1}{n}\partial b}{1+\frac{1}{n}b} \xrightarrow[n\to+\infty]{} 0,$$
$$\frac{\frac{1}{n}\partial b+\frac{1}{n}\partial a}{1+\frac{1}{n}b+\frac{1}{n}a} \xrightarrow[n\to+\infty]{} 0$$

and

$$\Lambda_{\text{Rao}} \xrightarrow[n\to+\infty]{} \Lambda_{\text{AMF}},$$
$$\Lambda_{\text{Kelly}} \xrightarrow[n\to+\infty]{} \Lambda_{\text{AMF}}.$$

Therefore

$$\Lambda'_{\text{Rao}} \xrightarrow[n\to+\infty]{} \Lambda'_{\text{AMF}},$$
$$\Lambda'_{\text{Kelly}} \xrightarrow[n\to+\infty]{} \Lambda'_{\text{AMF}}.$$

and

$$\Lambda'_{\text{AMF,Kelly,Rao}} = -\Lambda_{\text{AMF}}\left(\frac{\text{vec}\left(\mathbf{z}\mathbf{p}^H\right)^H}{\mathbf{z}^H\boldsymbol{\Sigma}^{-1}\mathbf{p}}+\frac{\text{vec}\left(\mathbf{p}\mathbf{z}^H\right)^H}{\mathbf{p}^H\boldsymbol{\Sigma}^{-1}\mathbf{z}}-\frac{\text{vec}\left(\mathbf{p}\mathbf{p}^H\right)^H}{\mathbf{p}^H\boldsymbol{\Sigma}^{-1}\mathbf{p}}\right)\left(\boldsymbol{\Sigma}^T\otimes\boldsymbol{\Sigma}\right)^{-1} \tag{4.15}$$

Finally, the variance in Eq. (3.5) becomes

$$\begin{aligned}\Lambda'_{\text{AMF}}\boldsymbol{\Gamma}_1\left(\Lambda'_{\text{AMF}}\right)^H &= \Lambda^2_{\text{AMF}}\left(\frac{\text{vec}\left(\mathbf{z}\mathbf{p}^H\right)^H}{\mathbf{z}^H\boldsymbol{\Sigma}^{-1}\mathbf{p}}+\frac{\text{vec}\left(\mathbf{p}\mathbf{z}^H\right)^H}{\mathbf{p}^H\boldsymbol{\Sigma}^{-1}\mathbf{z}}-\frac{\text{vec}\left(\mathbf{p}\mathbf{p}^H\right)^H}{\mathbf{p}^H\boldsymbol{\Sigma}^{-1}\mathbf{p}}\right)\\ &\times \left(\vartheta_1\left(\boldsymbol{\Sigma}^T\otimes\boldsymbol{\Sigma}\right)^{-1}+\vartheta_2\text{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\text{vec}\left(\boldsymbol{\Sigma}^{-1}\right)^H\right)\\ &\times \left(\frac{\text{vec}\left(\mathbf{z}\mathbf{p}^H\right)}{\mathbf{p}^H\boldsymbol{\Sigma}^{-1}\mathbf{z}}+\frac{\text{vec}\left(\mathbf{p}\mathbf{z}^H\right)}{\mathbf{z}^H\boldsymbol{\Sigma}^{-1}\mathbf{p}}-\frac{\text{vec}\left(\mathbf{p}\mathbf{p}^H\right)}{\mathbf{p}^H\boldsymbol{\Sigma}^{-1}\mathbf{p}}\right).\end{aligned} \tag{4.16}$$

Using Eqs. (A.1) and (A.8) one can easily show that

$$\frac{\text{vec}(\mathbf{a}\mathbf{b}^H)^H}{\mathbf{a}^H\boldsymbol{\Sigma}^{-1}\mathbf{b}}\left(\boldsymbol{\Sigma}^T\otimes\boldsymbol{\Sigma}\right)^{-1}\frac{\text{vec}(\mathbf{c}\mathbf{d}^H)}{\mathbf{d}^H\boldsymbol{\Sigma}^{-1}\mathbf{c}}=\frac{\mathbf{a}^H\boldsymbol{\Sigma}^{-1}\mathbf{c}\,\mathbf{d}^H\boldsymbol{\Sigma}^{-1}\mathbf{b}}{\mathbf{a}^H\boldsymbol{\Sigma}^{-1}\mathbf{b}\,\mathbf{d}^H\boldsymbol{\Sigma}^{-1}\mathbf{c}}$$

and

$$\frac{\text{vec}(\mathbf{a}\mathbf{b}^H)^H}{\mathbf{a}^H\boldsymbol{\Sigma}^{-1}\mathbf{b}}\text{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\text{vec}\left(\boldsymbol{\Sigma}^{-1}\right)^H\frac{\text{vec}(\mathbf{c}\mathbf{d}^H)}{\mathbf{d}^H\boldsymbol{\Sigma}^{-1}\mathbf{c}}=1.$$

After some mathematical manipulations, one can obtain the final results in Eq. (4.10). Applying $\boldsymbol{\Gamma}$ instead of $\boldsymbol{\Gamma}_1$ in Eq. (4.16), one obtains the results in Eq. (4.12). Analogously, starting with

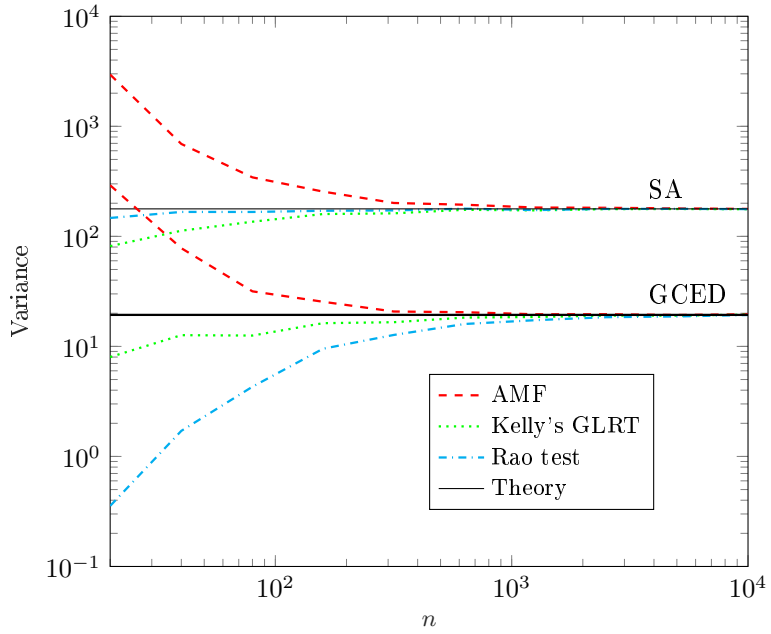
$$\Lambda'_{\text{ANMF}} = \Lambda_{\text{ANMF}}\left(\frac{\partial a}{a}-\frac{\partial b}{b}\right), \tag{4.17}$$

one obtains

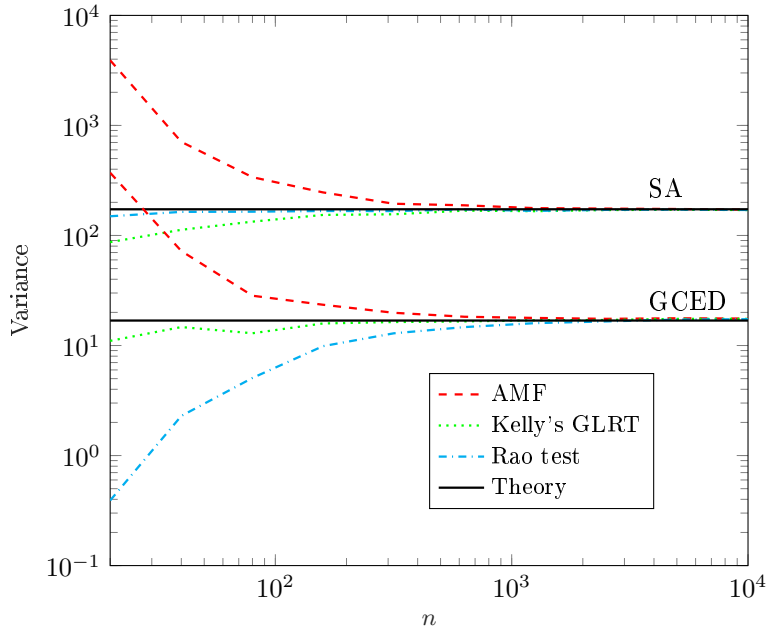
- ANMF: $\Lambda'_{\text{ANMF}} = \dfrac{\text{vec}\left(\mathbf{z}\mathbf{p}^H\right)^H}{\mathbf{z}^H\boldsymbol{\Sigma}^{-1}\mathbf{p}}+\dfrac{\text{vec}\left(\mathbf{p}\mathbf{z}^H\right)^H}{\mathbf{p}^H\boldsymbol{\Sigma}^{-1}\mathbf{z}}-\dfrac{\text{vec}\left(\mathbf{p}\mathbf{p}^H\right)^H}{\mathbf{p}^H\boldsymbol{\Sigma}^{-1}\mathbf{p}}-\dfrac{\text{vec}\left(\mathbf{z}\mathbf{z}^H\right)^H}{\mathbf{z}^H\boldsymbol{\Sigma}^{-1}\mathbf{z}},$ (4.18)

which leads to the results from Theorem 4.3.6 and concludes the proof.

□



(a) Student's M-estimator



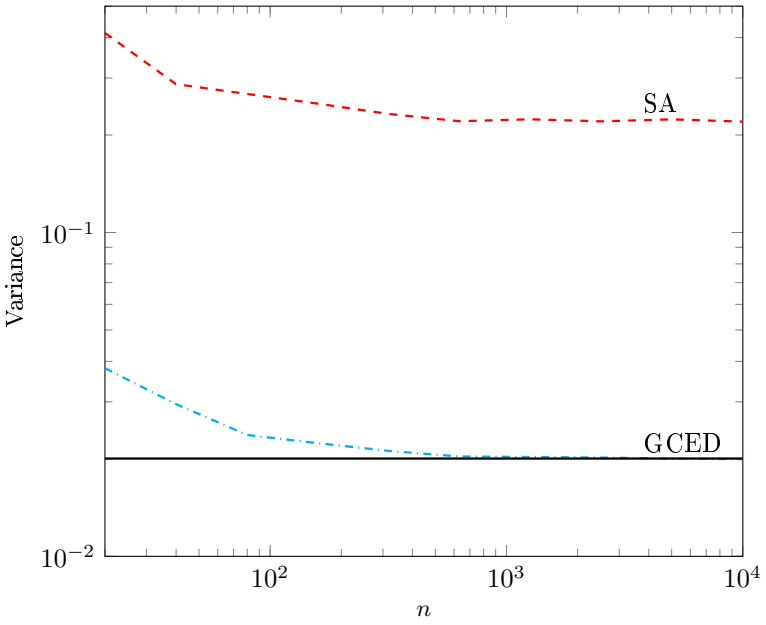
(b) Tyler's M-estimator

Figure 4.3 – Empirical variances of the robust detectors in standard asymptotic regime and when centered around their GCEDs, compared to the corresponding theoretical results; t-distributed data with $\nu = 4$ and $p = 10$.

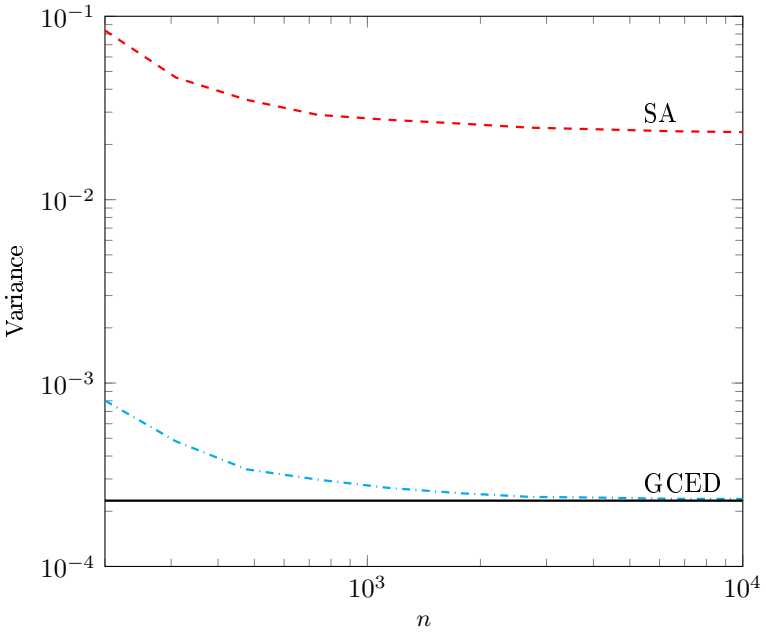
4.3.2 Experiments

In this part, some simulations are provided in order to support the theoretical results and demonstrate the practical interest of proposed theorems.

The simulations have been carried out with complex CES-distributed secondary data.



(a) $p = 10$



(b) $p = 100$

Figure 4.4 – Empirical variances of the robust TyE-ANMF (ANMF built with the Tyler's estimator) in the SA regime (red curve) and when centered around its GCED (blue curve), compared to the theoretical results.

The scatter matrix $\mathbf{\Sigma}$ is Toeplitz with a correlation coefficient $\rho = 0.5$. Empirical means are computed using Monte Carlo runs. In experiments, the parameters for M-estimators (DoF parameter, shape, scale, etc.) are usually set to be equal to the ones used to simulate the CES data, unless differently specified (mismatch scenarios). In practice, it is possible to estimate these parameters and then use appropriate M-estimators. However, this analysis

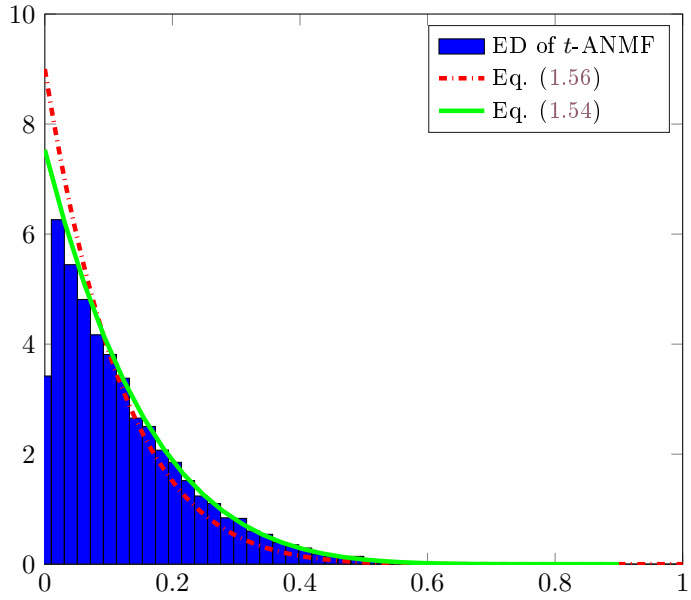


Figure 4.5 – Empirical distribution of the t-ANMF versus the theoretical distribution of NMF (Eq. (1.56)) in red and theoretical approximative distribution (Eq. (1.54)) in green; $p = 10$.

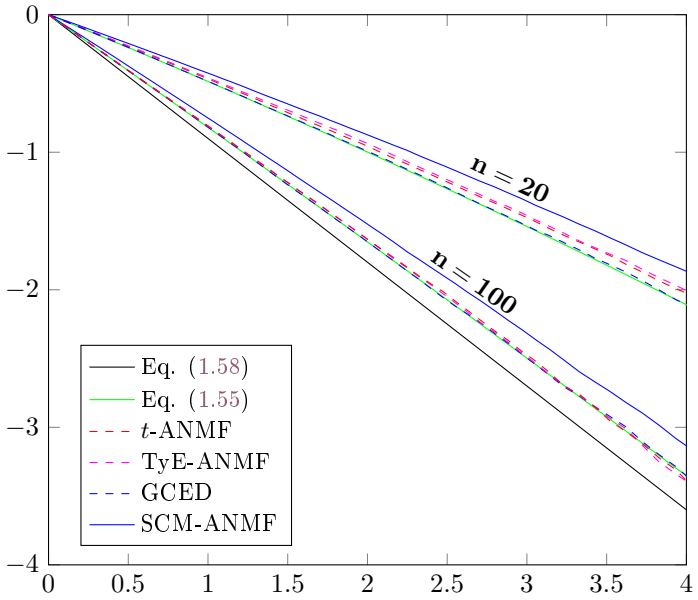


Figure 4.6 – Comparison between $P_{fa} - \lambda$ relationships for the t-ANMF, TyE-ANMF and SCM-ANMF with the empirical and theoretical results for the GCED (Eq. (1.55)) and the NMF(Eq. (1.58)); Student t-distributed data with $\nu = 2$, $p = 10$.

is beyond the scope of this thesis.

Figure 4.3 illustrates the theoretical result of Theorems 4.3.3 and 4.3.5. The empirical variances of robust detectors in the SA regime, i.e. when centering around the detectors

obtained with the true scatter matrix, are plotted and compared to the theoretical results (Theorem 4.3.5). Moreover, the empirical variances of the difference between the robust detectors and their GCEDs are also given together with the corresponding theoretical result (Theorem 4.3.5). First, one can see that the simulations validate theoretical results. Then, one can note that the variances when comparing to the GCEDs are significantly lower than the ones in the SA, which **justifies the proposed approximation of the behavior of robust detectors with the one of their GCEDs**. One observes that the error decreases very fast as the number n of samples increases. Furthermore, simulations show that **the approximation is also valid for small** n.

In Figure 4.4 the corresponding results for the ANMF built with the Tyler's estimator (TyE-ANMF) for $p = 10$ and $p = 100$ are plotted. Once again, one can note that the empirical variance of TyE-ANMF when compared to the GCED is remarkably smaller than the one in SA. In addition, one can notice that **this difference is even smaller for higher dimension, which is completely in agreement with theoretical results**.

Since the variance between a robust detector and its GCED is remarkably smaller that the one in the SA regime, one can conclude that **the distribution of the robust detector can be well-approximated with the one of GCED**. Figure 4.5 illustrates this point. The histogram on Figure 4.5 represents the empirical distribution of t-ANMF that is compared to the theoretical distribution of NMF given by Eq. (1.56) (red curve) and theoretical distribution of GCWE given by Eq. (1.54) (green curve), where t-ANMF is the ANMF test built with Student's M-estimator. One can note that the red curve mismatches the empirical distribution of t-ANMF, while the green one borders the plot area of the histogram showing that Eq. (1.56) gives a good approximation of the t-ANMF's behavior. This is also valid for others detection statistics.

Having this in mind, one can analyze the $P_{fa} - \lambda$ relationship for the robust detectors. In Figure 4.6 we observe this relationship for the robust ANMF for different number of secondary data, $n = 20$ and $n = 100$. The empirical results for t-ANMF and TyE-ANMF have been plotted and compared to the theoretical ones given by Eqs. (1.58) and (1.55). The empirical results for the corresponding GCWE-ANMF (that is not available in practice) and the SCM-ANMF built with observed data are also provided. First, one can notice a good match of the empirical distributions of both t-ANMF and TyE-ANMF with the theoretical and empirical distribution of the GCWE-ANMF. This shows that **the behavior of robust detectors is better approximated with the one of the GCWE-ANMF (green curve) than with the corresponding NMF (black curve)**. Secondly, one can see that this claim is **even more obvious for small** n **since all curves approach when** n **increases**. Finally, one can note that the SCM built with secondary data does not satisfy the relationship Eq. (1.55) anymore. This is expected since **the SCM is calculated with a non-Gaussian data and its performance is remarkably degraded, which highly supports the use of** M**-estimators in this context**.

Finally, for a constant P_{fa}, we can study the P_d of the robust ANMF for a given SNR δ. Figure 4.7 shows the empirical P_d for TyE-ANMF where the detection threshold is computed empirically for $P_{fa} = 0.001$ (TyE-ANMF$_{\text{emp}}$) together with the empirical P_d obtained with the threshold computed using Eq. (1.55) (TyE-ANMF$_{\text{the}}$). The empirical results have been compared to the corresponding theoretical results for the GCWE-ANMF given by Eq. (1.53) (green curves) for $n = 20, 50, 200$. Finally, the theoretical results for NMF (Eq. (1.57)) have been plotted. The figure is revealing in several ways. First, the results for TyE-ANMF$_{\text{emp}}$ match perfectly the ones for TyE-ANMF$_{\text{the}}$, **meaning that we obtain the same** P_d **for empirically and theoretically computed threshold which can significantly reduce the computational cost of the method**. In addition, both empirical results coincide with the theoretical one for the GCED as stated. Finally, the

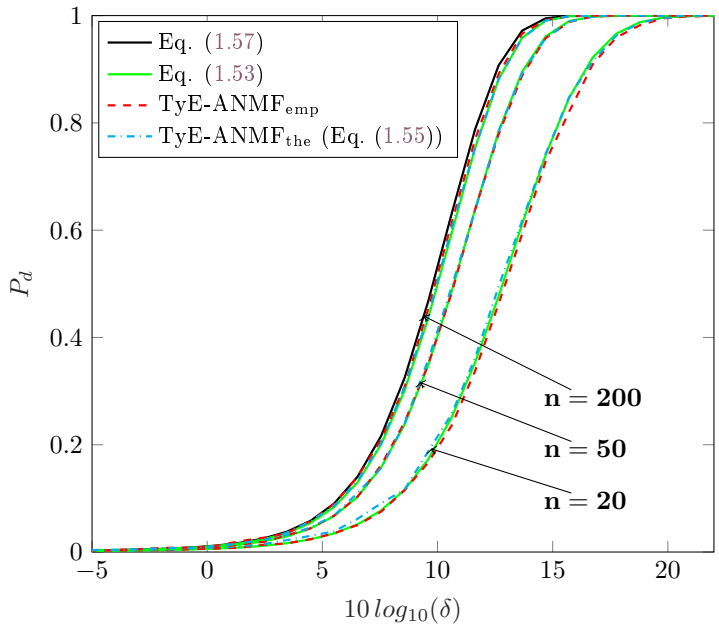


Figure 4.7 – Probability of Detection of TyE-ANMF for $P_{fa} = 0.001$: Empirical results obtained with an empirically computed threshold (TyE-ANMF$_{\text{emp}}$) and theoretically computed threshold using Eq. (1.55) (TyE-ANMF$_{\text{the}}$), with the theoretical results for GCED Eq. (1.53) and NMF (1.57); t-distributed data with $\nu = 2$.

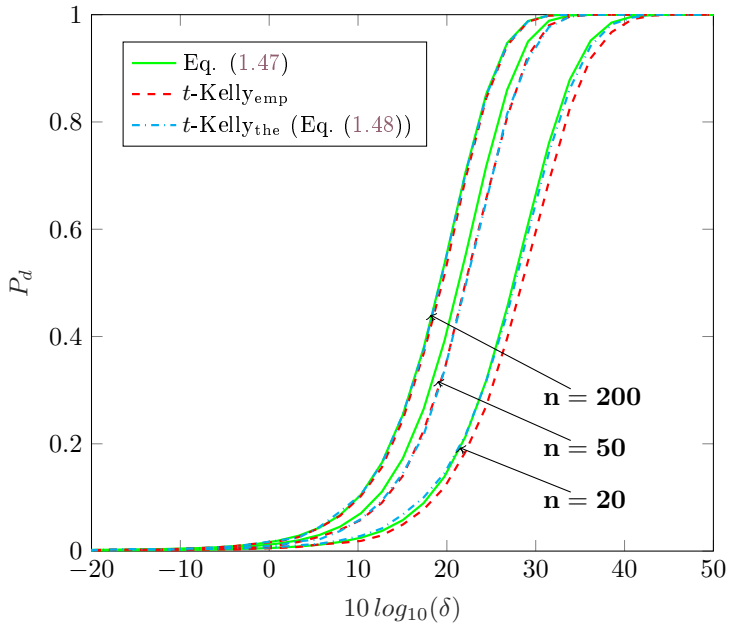


Figure 4.8 – Probability of Detection of t-Kelly for $P_{fa} = 0.001$: Comparison of the empirical results obtained for empirically computed threshold (t-Kelly$_{\text{emp}}$) and with the theoretically computed threshold using Eq. (1.48) (t-Kelly$_{\text{the}}$), with the theoretical result for GCED given in Eq. (1.47); t-distributed data with $\nu = 2$.

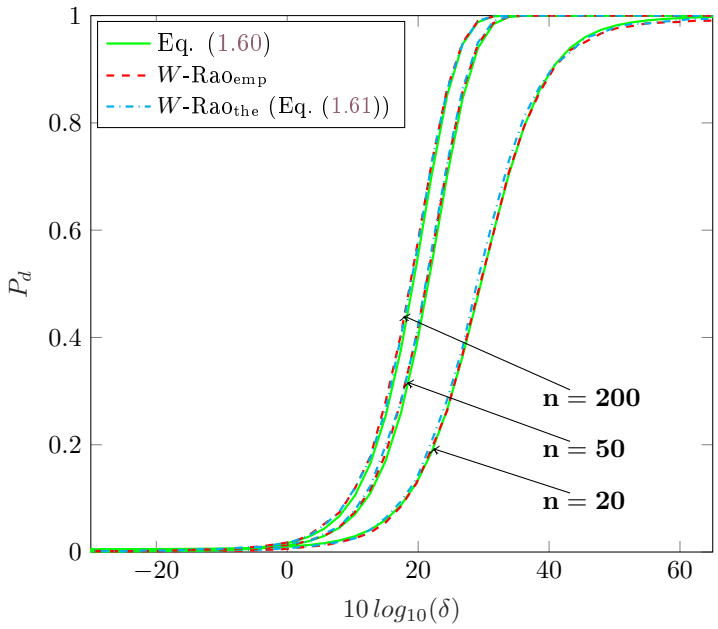


Figure 4.9 – Probability of Detection of W-Rao for $P_{fa} = 0.001$: Comparison of the empirical results obtained for empirically computed threshold (W-Rao$_{\text{emp}}$) and theoretically computed threshold using Eq. (1.61) (W-Rao$_{\text{the}}$), with the theoretical result for GCED given in Eq. (1.60); W-distributed data with $s = 0.5$.

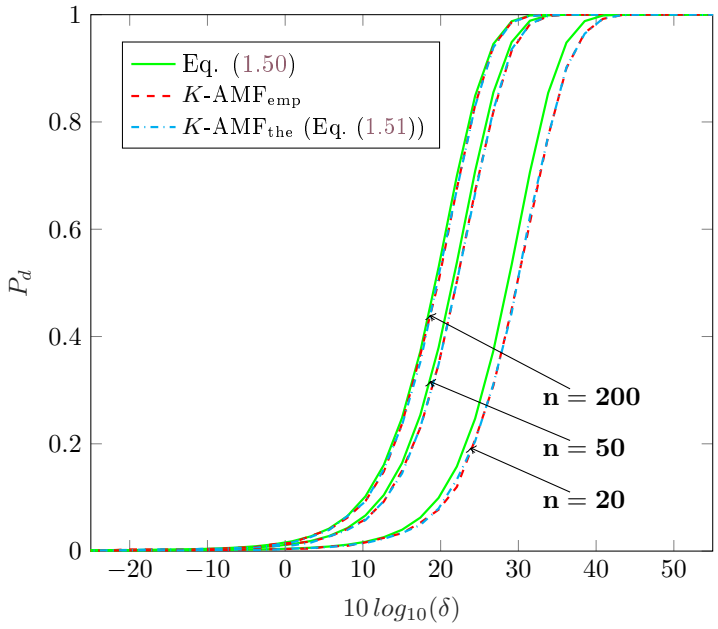


Figure 4.10 – Probability of Detection of K-AMF for $P_{fa} = 0.001$: Comparison of the empirical results obtained for empirically computed threshold (K-AMF$_{\text{emp}}$) and theoretically computed threshold using Eq. (1.51), with the theoretical result for the GCED given in Eq. (1.50); K-distributed data with $\nu = 2$.

approximation is valid even for small n approaching the results for NMF when n increases. **Taken together, these results suggest that one can use M-estimators to compute the value of detection statistic and precompute the detection threshold using the theoretical results for GCED**.

Figures 4.8, 4.9 and 4.10 summarize the same quantities for Kelly's GLRT, Rao statistics and AMF for different M-estimators.

Figure 4.8 plots the results for the Kelly's GLRT computed with Student's t-estimator (t-Kelly). The data is t-distributed with $\nu = 2$. We can draw the same conclusions as for the TyE-ANMF. In this case, we notice the slight deviation between the empirical results and the theoretical ones for small values of n. However, when n increases all the curves approaches.

Figure 4.9 plots the results for the Rao statistics computed with Weibull M-estimator (W-Rao). The data is W-distributed with $s = 0.5$.

Finally, Figure 4.10 plots the results for the AMF computed with K M-estimator (K-AMF). The data is K-distributed with $\nu = 2$. As for the t-Kelly (Figure 4.8), one observes a minor deviation for $n = 20$, which for $n = 50$ becomes completely insignificant.

In summary, the comments drawn for the TyE-ANMF are also valid for others robust detectors, supporting the proposed approximation and showing the interest of the theoretical results.

4.4 Robust low-rank detection

Let us consider LR detection problem defined as follows

$$\begin{cases} H_0 : \mathbf{z} = \widetilde{\mathbf{c}} & \mathbf{z}_i = \widetilde{\mathbf{c}}_i, \quad i = 1, \ldots, n, \\ H_1 : \mathbf{z} = \alpha \mathbf{p} + \widetilde{\mathbf{c}} & \mathbf{z}_i = \widetilde{\mathbf{c}}_i, \quad i = 1, \ldots, n, \end{cases} \tag{4.19}$$

where $\widetilde{\mathbf{c}}_i$ are i.i.d. with $\widetilde{\mathbf{c}}_i \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}})$ with $\mathbf{\Sigma} = \mathbf{\Sigma}_r + \gamma^2 \mathbf{I}_p$ that is low-rank structured as in Eq. (3.30).

4.4.1 Asymptotics of the robust LR-ANMF

Let us consider the operator $\mathcal{R}_r\{.\}$ defined in Eq. (3.31). Let $\widehat{\mathbf{\Sigma}}$ be an M-estimator built with the sample $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ from Eq. (4.19) and let $\widehat{\mathbf{\Sigma}}_{\text{GCWE}}$ be its GCWE from Definition 2.1.2. We have the corresponding principal subspaces defined as in Eq. (3.32)

$$\begin{cases} \mathbf{\Pi}_r^{\perp} = \mathbf{I} - \mathcal{R}_r\{\mathbf{\Sigma}\}, \\ \widehat{\mathbf{\Pi}}_r^{\perp M} = \mathbf{I} - \mathcal{R}_r\{\widehat{\mathbf{\Sigma}}\}, \\ \widehat{\mathbf{\Pi}}_r^{\perp \text{GCWE}} = \mathbf{I} - \mathcal{R}_r\{\widehat{\mathbf{\Sigma}}_{\text{GCWE}}\}. \end{cases} \tag{4.20}$$

Finally, let $\Lambda_{\text{LR}}(\cdot)$ be the LR-ANMF defined in Eq. (1.66). Then, we can define

$$\begin{cases} \Lambda_{\text{LR}} = \Lambda_{\text{LR}}\left(\mathbf{\Pi}_r^{\perp}\right), \\ \widehat{\Lambda}_{\text{LR}}^{M} = \Lambda_{\text{LR}}\left(\widehat{\mathbf{\Pi}}_r^{\perp M}\right), \\ \widehat{\Lambda}_{\text{LR}}^{\text{GCED}} = \Lambda_{\text{LR}}\left(\widehat{\mathbf{\Pi}}_r^{\perp \text{GCWE}}\right). \end{cases} \tag{4.21}$$

Now, one can state the following theorems.

Theorem 4.4.1: SA of the robust LR-ANMF

Let us consider the LR-ANMF test defined by Eq. (1.66). Thus, conditionally to the distributions of $\mathbf{z}$, the asymptotic distribution of $\widehat{\Lambda}_{\mathrm{LR}}^{M}$ is given by

$$\sqrt{n}\left(\widehat{\Lambda}_{\mathrm{LR}}^{M}-\Lambda_{\mathrm{LR}}\right)_{\mathbf{z}} \xrightarrow{d} \mathcal{N}\left(0, 2\vartheta_1\gamma^2\left(1-\Lambda_{\mathrm{LR}}\right)\left(b-a\Lambda_{\mathrm{LR}}\right)\right), \tag{4.22}$$

where

$$a=\frac{\mathbf{z}^H\mathbf{A}\mathbf{p}}{\mathbf{z}^H\mathbf{\Pi}_{\mathbf{c}}^{\perp}\mathbf{p}}+\frac{\mathbf{p}^H\mathbf{A}\mathbf{z}}{\mathbf{p}^H\mathbf{\Pi}_{\mathbf{c}}^{\perp}\mathbf{z}} \quad \text{and} \quad b=\frac{\mathbf{z}^H\mathbf{A}\mathbf{z}}{\mathbf{z}^H\mathbf{\Pi}_{\mathbf{c}}^{\perp}\mathbf{z}}+\frac{\mathbf{p}^H\mathbf{A}\mathbf{p}}{\mathbf{p}^H\mathbf{\Pi}_{\mathbf{c}}^{\perp}\mathbf{p}},$$

with $\mathbf{A}$ defined in Theorem 3.3.1.

Note that the variance of the robust LR-ANMF, as in full-rank case, depends only on the first scale factor. However, this property stands also for other detection characteristics, such as Kelly's GLRT, AMF and Rao test, since the scale-invariance is brought by the properties of the used projector.

Proof. In order to prove the theorem, we will use Lemma 3.1.1. Analogously to Eq. (4.18), the derivative of Λ_{LR} w.r.t. $\mathbf{\Pi}_r^{\perp}$ is

$$\Lambda_{\mathrm{LR}}'=\Lambda_{\mathrm{LR}}\left(\frac{\mathrm{vec}\left(\mathbf{z}\mathbf{p}^H\right)^H}{\mathbf{z}^H\mathbf{\Pi}_r^{\perp}\mathbf{p}}+\frac{\mathrm{vec}\left(\mathbf{p}\mathbf{z}^H\right)^H}{\mathbf{p}^H\mathbf{\Pi}_r^{\perp}\mathbf{z}}-\frac{\mathrm{vec}\left(\mathbf{p}\mathbf{p}^H\right)^H}{\mathbf{p}^H\mathbf{\Pi}_r^{\perp}\mathbf{p}}-\frac{\mathrm{vec}\left(\mathbf{z}\mathbf{z}^H\right)^H}{\mathbf{z}^H\mathbf{\Pi}_r^{\perp}\mathbf{z}}\right). \tag{4.23}$$

Then, to compute the asymptotic covariance matrix of $\widehat{\mathbf{\Pi}}^{\perp M}$, one can use the quantity $\widehat{\mathbf{\Pi}}_r^{\perp M}=\mathbf{\Pi}_r^{\perp}-\delta\mathbf{\Pi}_r$ and similar steps as in the proof of Theorem 3.3.1. One can show that $\mathbf{\Gamma}_{\mathbf{\Pi}^{\perp}}=\mathbf{\Gamma}_{\mathbf{\Pi}}$. Consequently, to compute the variance of $\widehat{\Lambda}_{\mathrm{LR}}$, one needs to compute

$$\begin{aligned}\Lambda_{\mathrm{LR}}'\mathbf{\Gamma}_{\mathbf{\Pi}}\left(\Lambda_{\mathrm{LR}}'\right)^H &= \left(\frac{\mathrm{vec}\left(\mathbf{z}\mathbf{p}^H\right)^H}{\mathbf{z}^H\mathbf{\Pi}_r^{\perp}\mathbf{p}}+\frac{\mathrm{vec}\left(\mathbf{p}\mathbf{z}^H\right)^H}{\mathbf{p}^H\mathbf{\Pi}_r^{\perp}\mathbf{z}}-\frac{\mathrm{vec}\left(\mathbf{p}\mathbf{p}^H\right)^H}{\mathbf{p}^H\mathbf{\Pi}_r^{\perp}\mathbf{p}}-\frac{\mathrm{vec}\left(\mathbf{z}\mathbf{z}^H\right)^H}{\mathbf{z}^H\mathbf{\Pi}_r^{\perp}\mathbf{z}}\right)\\ &\times \vartheta_1\Lambda_{\mathrm{LR}}^2\left(\mathbf{A}^T\otimes\mathbf{B}+\mathbf{B}^T\otimes\mathbf{A}\right)\\ &\times \left(\frac{\mathrm{vec}\left(\mathbf{z}\mathbf{p}^H\right)}{\mathbf{p}^H\mathbf{\Pi}_r^{\perp}\mathbf{z}}+\frac{\mathrm{vec}\left(\mathbf{p}\mathbf{z}^H\right)}{\mathbf{z}^H\mathbf{\Pi}_r^{\perp}\mathbf{p}}-\frac{\mathrm{vec}\left(\mathbf{p}\mathbf{p}^H\right)}{\mathbf{p}^H\mathbf{\Pi}_r^{\perp}\mathbf{p}}-\frac{\mathrm{vec}\left(\mathbf{z}\mathbf{z}^H\right)}{\mathbf{z}^H\mathbf{\Pi}_r^{\perp}\mathbf{z}}\right).\end{aligned} \tag{4.24}$$

Furthermore, since $\mathbf{A}=\mathbf{U}_r\left(\gamma^2\mathbf{\Lambda}_r^{-2}+\mathbf{\Lambda}_r^{-1}\right)\mathbf{U}_r^H$ and $\mathbf{B}=\gamma^2\mathbf{\Pi}_r^{\perp}$

$$\begin{aligned}&\frac{\mathrm{vec}(\mathbf{a}\mathbf{b}^H)^H}{\mathbf{a}^H\mathbf{\Pi}_r^{\perp}\mathbf{b}}\left(\mathbf{A}^T\otimes\mathbf{B}+\mathbf{B}^T\otimes\mathbf{A}\right)\frac{\mathrm{vec}(\mathbf{c}\mathbf{d}^H)}{\mathbf{d}^H\mathbf{\Pi}_r^{\perp}\mathbf{c}}\\ &\quad=\frac{\mathbf{a}^H\mathbf{A}\mathbf{c}\,\mathbf{d}^H\mathbf{B}\mathbf{b}}{\mathbf{a}^H\mathbf{\Pi}_r^{\perp}\mathbf{b}\,\mathbf{d}^H\mathbf{\Pi}_r^{\perp}\mathbf{c}}+\frac{\mathbf{a}^H\mathbf{B}\mathbf{c}\,\mathbf{d}^H\mathbf{A}\mathbf{b}}{\mathbf{a}^H\mathbf{\Pi}_r^{\perp}\mathbf{b}\,\mathbf{d}^H\mathbf{\Pi}_r^{\perp}\mathbf{c}}\\ &=\gamma^2\frac{\mathbf{a}^H\mathbf{A}\mathbf{c}\,\mathbf{d}^H\mathbf{\Pi}_r^{\perp}\mathbf{b}}{\mathbf{a}^H\mathbf{\Pi}_r^{\perp}\mathbf{b}\,\mathbf{d}^H\mathbf{\Pi}_r^{\perp}\mathbf{c}}+\gamma^2\frac{\mathbf{a}^H\mathbf{\Pi}_r^{\perp}\mathbf{c}\,\mathbf{d}^H\mathbf{A}\mathbf{b}}{\mathbf{a}^H\mathbf{\Pi}_r^{\perp}\mathbf{b}\,\mathbf{d}^H\mathbf{\Pi}_r^{\perp}\mathbf{c}}.\end{aligned} \tag{4.25}$$

Then, taking various combinations of $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ and $\mathbf{d}$ in Eq. (4.24), after some fastidious, but simple computations, one obtains the final result.

□

Theorem 4.4.2: GCED of the robust LR-ANMF

Let us consider the LR-ANMF test defined by Eq. (1.66). Thus, conditionally to the distributions of $\mathbf{z}$, the asymptotic distribution of $\widehat{\Lambda}_{\mathrm{LR}}^{M}$ is given by

$$\sqrt{n}\left(\widehat{\Lambda}_{\mathrm{LR}}^{M}-\widehat{\Lambda}_{\mathrm{LR}}^{\mathrm{GCED}}\right)_{\mathbf{z}} \xrightarrow{d} \mathcal{N}\left(0, 2\sigma_1\gamma^2\left(1-\Lambda_{\mathrm{LR}}\right)\left(b-a\Lambda_{\mathrm{LR}}\right)\right). \tag{4.26}$$

Remark 4.4.2

- One can note that, similarly to $\widehat{\Lambda}_{\mathrm{ANMF}}^{M}$, the variance of $\widehat{\Lambda}_{\mathrm{LR}}^{M}$ when compared to $\widehat{\Lambda}_{\mathrm{LR}}^{\mathrm{GCED}}$ (Eq. (4.26)) differs from the one in the SA regime (Eq. (4.22)) only in the scale factors σ_1 and ϑ_1. Consequently, since $\sigma_1 < \vartheta_1$, the behavior of $\widehat{\Lambda}_{\mathrm{LR}}^{M}$ is better characterized with the behavior of $\widehat{\Lambda}_{\mathrm{LR}}^{\mathrm{GCED}}$ than with the one of Λ_{LR}.

Proof. The results can be obtained using the results from Theorem 4.4.1 and common steps for the difference between the SA and GCED regime. □

4.4.2 Experiments

The following simulations validate the proposed results. In Figure 4.11, the results for the LR-ANMF built with the Student's M-estimator (t-LR-ANMF) are presented. The data follow the complex t-distribution with $\nu = 2$. The empirical variance of t-LR-ANMF and the one of the difference between t-LR-ANMF and the GCED, compared to the theoretical results of Theorems 4.4.1 and 4.4.2, are plotted. One can draw the same conclusions as in the full-rank detection problem, i.e., that the variance is significantly smaller in the GCED regime than in the SA one and consequently, the distribution of t-LR-ANMF can be well-approximated by the one of its GCED.

However, in this case, the exact distribution of the corresponding GCED is unknown. The approximation given by Eq. (1.67) could be used, but it should be noted that does not provide accurate results for small n. Nevertheless, we can compare the empirical distributions of t-LR-ANMF and its GCED. From Figure 4.12 we observe a great match between these two empirical distributions, which supports our claims.

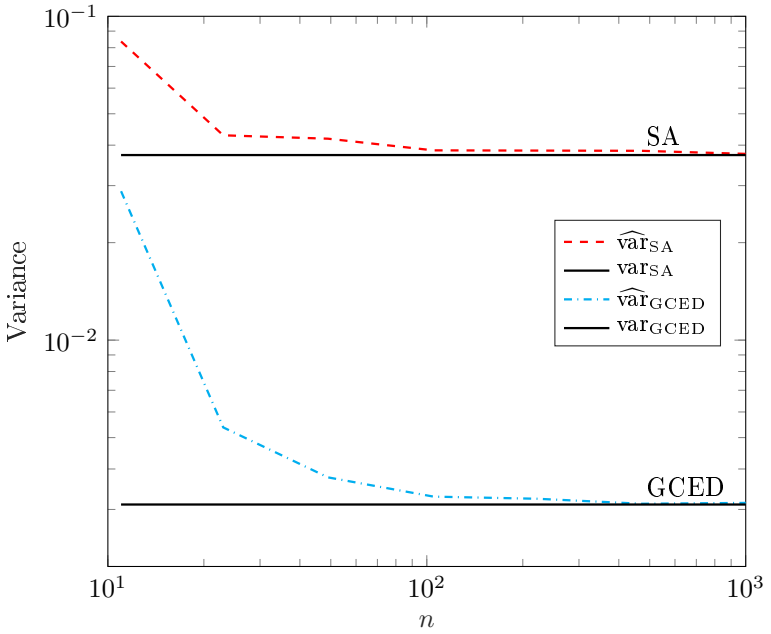


Figure 4.11 – Empirical variances of the robust t-LR-ANMF in SA regime and when centered around its GCED, compared to the theoretical results (Theorems 4.4.1 and 4.4.2), t-distributed data, $\nu = 2$, $p = 10$.

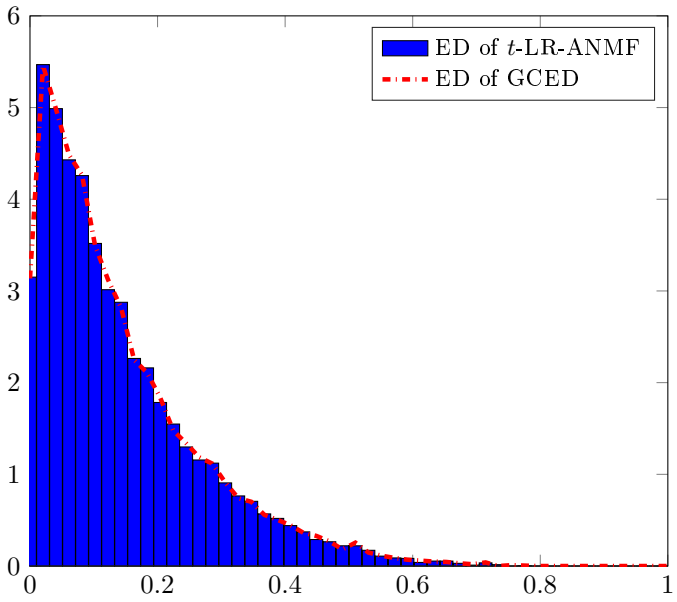


Figure 4.12 – Empirical distribution of the t-LR-ANMF (in blue) versus the empirical distribution of the GCED (in red); $p = 10$.

4.5 Conclusion

In this chapter we have studied the signal detection problem when the secondary data is CES-distributed. The behaviour of the robust Mahalanobis distance and robust detectors have been analyzed. The robust statistics in this context correspond to the traditional statistics derived for the Gaussian-distributed data computed with an M-estimator plugged-in instead of the SCM.

In order to analyze the distribution of the robust statistics, we have introduced their Gaussian-Core Equivalents (GCE), i.e. GCEM and GCED, thanks to the GCWE defined in Chapter 2. Finally, we have derived the asymptotic distributions in the SA regime and GCE one.

We have seen that the distribution of the robust statistic can be well-approximated by the one of the corresponding GCE. The application of these results has been demonstrated in the context of the detection threshold computation. We have seen that the PFA-threshold relationship of a robust detector coincides well with the theoretical one of the corresponding GCED. Consequently, one can use an M-estimator to obtain an accurate estimation of the scatter matrix in the CES context and rely on the theoretical PFA-threshold relationship of the GCED to compute the detection threshold for a given PFA. This has been empirically illustrated for several detection statistics with different set-ups. The theoretical detection derivation can significantly reduce the computational cost of the robust detection algorithms.

Finally, it should be noted that the theoretical results for GCEDs are valid when the primary data are Gaussian-distributed. We have shown that an M-estimator can be plugged-in instead of the traditional SCM without changing the behavior of the statistics. Consequently, if the primary data have a particular CES distribution, the approximative distribution of the robust statistics can be derived using the approximative Wishart distribution of the used M-estimator.

In conclusion, the analysis done in this chapter additionally supports the previously obtained results and leads to the same conclusion: *One should always use the robust estimators in order to mitigate the impact of outlying, missing and/or heterogeneous data, while relying on the statistical properties of Gaussian-based estimators.*

Chapter 5

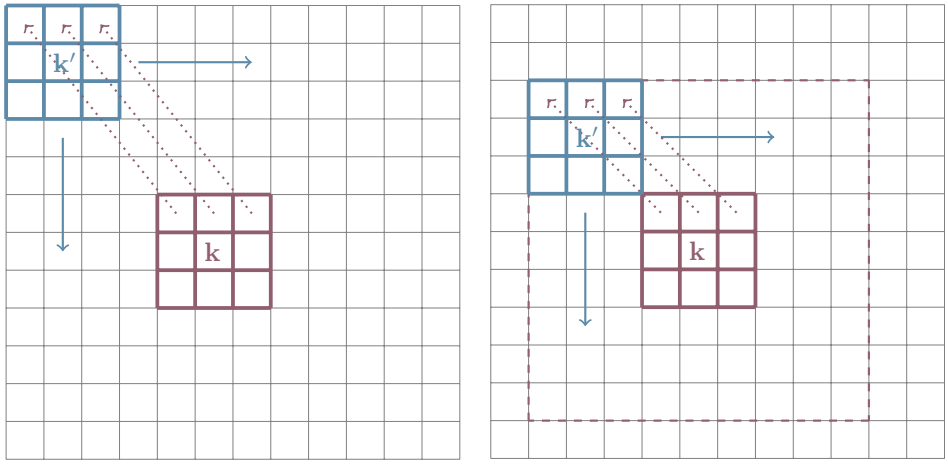
Robust NL-means approach for PolSAR image denoising

In this chapter we propose a new method for polarimetric synthetic aperture radar (PolSAR) denoising, named M-NL. More precisely, we seek to address a new statistical approach for weights computation in non-local (NL) approaches. The aim is to present a simple criterion to detect similar pixels in a PolSAR image, which is based on the new statistical properties of M-estimators derived in Chapter 2. A binary hypothesis test is used to select similar pixels which will be used for covariance matrix estimation together with associated weights. The method is then compared to an advanced state-of-the-art PolSAR denoising method, NL-SAR method. The filter performances are measured by a set of different indicators, including relative errors on incoherent target decomposition parameters, coherences, polarimetric signatures, and edge preservation on a set of simulated PolSAR images. The results reveal that M-NL outperforms the NL-SAR method in most cases. Moreover, the results obtained for RADARSAT-2 PolSAR data are presented. The M-NL method offers efficient speckle reduction in homogeneous regions with good edge preservation.

Publications associated with this chapter:

[J4] **G. Draškovi**ć, F. Pascal, and F.Tupin, "*M*-NL: Robust NL-means approach for PolSAR image denoising," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 6, pp. 997–1001, 2019.

[C8] **G. Draškovi**ć, F. Pascal et F. Tupin, "Débruitage des images polarimétriques avec le *M*-NL robuste," *colloque GRETSI sur le traitement du signal et des images*, Lille, France, Août 2019.



(a) Non-local means principle (b) "Local" non-local means

Figure 5.1 – Patch comparison in non-local means methods: comparing $\mathbf{k}$ to $\mathbf{k}'$ reduces to comparing each pixel from the patch centered in $\mathbf{k}$ to the corresponding one from the patch centred in $\mathbf{k}'$

5.1 Non-local means methods

Let us introduce the general methodology of non-local means denoising methods and their adaptation to PolSAR image despeckling.

Contrary to local window filters, which use the values of a group of pixels surrounding a target pixel to filter the image, in non-local means (NLM) methods all pixels in the image are used during the processing and are weighted with their similarity to the target pixel. Moreover, instead of a simple pixel comparison, the patch comparison is proposed to be robust to noise. In order to obtain the similarity between two pixels, the method compares the values in small local windows centered in the pixels to be compared. Then, the final value in the central pixel of the patch is obtained as a weighted mean based on the computed (dis)similarities.

The patch comparison in NLM methods permits to find similar geometrical configurations in the whole image and even with a high level of noise provide better results. However, these methods demand more time resources. In order to reduce the computational cost and improve the performances, the search area is limited to a large window [51], as shown in Figure 5.1.

The NLM methods can be also considered as a generalization of the bilateral filter where the dissimilarities between two single points in the bilateral kernel are replaced with a patch-based term ignoring the geometric distance between pixels.

The NLM method was originally introduced for digital images [21]. Later, the methodology used in NLM methods has been successfully adapted to the despeckling of PolSAR data [25, 47, 185, 192]. The dissimilarities are computed using the covariance matrix comparison between pixels. These methods principally concentrate on how to compute the similarities between pixels, select similar pixels and better calculate the weights.

In the following, we present NL-SAR [46], a unified nonlocal framework for the despeckling of amplitude, polarimetric SAR, and/or interferometric SAR images. However, in this chapter we are interested only in PolSAR image despeckling. In Section 5.3, we propose a new approach to compute weights using M-estimators of scatter and their statistical properties.

5.2 NL-SAR

NL-SAR (non-local SAR) is an advanced NLM method for radar image denoising. It contains the following steps:

- **Pre-estimation** A first step of pre-filtering is applied before the patch comparison when the noise level is very high. For a given scale s, this step is performed by convolution with a truncated Gaussian envelope defined by $c_g \exp\left(-\frac{x^2+y^2}{(s+0.5)^2}\right)$ if $-\mathrm{s} \leq x,y \leq \mathrm{s}$ and 0 otherwise, where c_g is a normalization constant [1].
- **Weight computation:** To measure the similarity between two empirical (pre-estimated with the previous step) independent matrices $\widehat{\mathbf{\Sigma}}_1$ and $\widehat{\mathbf{\Sigma}}_2$, [2] the generalized likelihood ratio (GLR) criterion obtained under the Gaussian assumption is used

$$\mathcal{L}_{\mathrm{GLR}}\left(\widehat{\mathbf{\Sigma}}_1, \widehat{\mathbf{\Sigma}}_2\right) = \frac{\left|\widehat{\mathbf{\Sigma}}_1\right|^L \left|\widehat{\mathbf{\Sigma}}_2\right|^L}{\left|\frac{1}{2}\left(\widehat{\mathbf{\Sigma}}_1 + \widehat{\mathbf{\Sigma}}_2\right)\right|^{2L}},$$

where L is referred to as the number of looks (see Eq. (1.81)) .
The dissimilarity between two patches centered at the two pixels l and l' is then:

$$\Delta_{\mathrm{NLSAR}}(l,l') = \sum_{\tau} -\ln\left(\mathcal{L}_{\mathrm{GLR}}\left(\widehat{\mathbf{\Sigma}}(l+\tau), \widehat{\mathbf{\Sigma}}(l'+\tau)\right)\right),$$

where $\tau \in [-\mathrm{p},\mathrm{p}]^2$ is a 2-D shift indicating the location within each patch of size $\mathbb{P} = (2\mathrm{p}+1)\times(2\mathrm{p}+1)$.
Finally, the weights are computed using the following learned kernels

$$\omega_{\mathrm{NLSAR}}(l,l') = \begin{cases} \exp\left(-\frac{\left|\mathbb{G}^{-1}\{\mathbb{F}[\Delta_{\mathrm{NLSAR}}(l,l')]\} - c_{\mathrm{NLSAR}}\right|}{h}\right) & \text{if } l \neq l', \\ 1 & \text{if } l = l', \end{cases} \tag{5.1}$$

where $c_{\mathrm{NLSAR}} = \mathbb{E}[\mathbb{G}^{-1}|\{\mathbb{F}[\Delta_{\mathrm{NLSAR}}(l,l')]\}|H_0]$. The parameter h is set to 1/3 and $\mathbb{G}$ is the χ^2-distribution with 49 degrees (up to a change of variables, this corresponds to the set of parameters for Gaussian NL-means used in [23] with 7×7 patches). Patches are extracted from a homogeneous area in order to sample the probability distribution of $\Delta_{\mathrm{NLSAR}}(l,l')$. Dissimilarity values $\Delta_{\mathrm{NLSAR}}(l,l')$ are estimated off-line for all pairs of patches $\Delta_{\mathrm{NLSAR}}(l,l')$. These dissimilarities are stored in a separate table sorted in increasing order for each possible choice of the patch size p and pre-estimation scale s. On-line, when the weight $\omega_{\mathrm{NLSAR}}(l,l')$ must be evaluated from the dissimilarity $\Delta_{\mathrm{NLSAR}}(l,l')$ evaluated with given parameters p and s, the closest value to $\Delta_{\mathrm{NLSAR}}(l,l')$ is found in the corresponding sorted table by binary search. The rank of that value in the table (i.e., its index) gives the corresponding quantile $\mathbb{F}[\Delta_{\mathrm{NLSAR}}(l,l')]$.
- **Weighted MLE:** The weighted MLE is given by the weighted means

$$\widehat{\mathbf{\Sigma}}_{\mathrm{NL}}(l) = \frac{\sum_{l'} \omega_{\mathrm{NLSAR}}(l,l')\mathbf{k}'\mathbf{k}'^H}{\sum_{l'} \omega_{\mathrm{NLSAR}}(l,l')}, \tag{5.2}$$

where the weights $\omega_{\mathrm{NLSAR}}(l,l')$ are defined as in Eq. (5.1) and the sum is carried out over all pixels $\mathbf{k}'$ in the search window of size $\mathbb{W} = (2\mathrm{w}+1)\times(2\mathrm{w}+1)$ centered at pixel $\mathbf{k}$. The weighted MLE in Eq. (5.2) is computed for various triples of (s, p, w).

1. The case s = 0 indicates that the pre-estimation is off.
2. The matrices, associated to each pixels, are estimated by using the neighbor pixels in a given window $\mathbb{S} = (2\mathrm{s}+1)\times(2\mathrm{s}+1)$. Note that numerous parameters/effects will affect the results, e.g., the window size, the patch size, the resolution, the fact that estimated matrices that are spatially close are not statistically independent (since same pixels are used several times), etc.

- **Bias reduction:** The algorithm tends to over-smooth the image around bright targets that have intensities significantly larger than their surrounding background. Thus, a bias reduction step is performed on each estimate $\widehat{\mathbf{\Sigma}}_{\mathrm{NL}}$. The Non-Local Reduced Bias (NLRB) estimate is obtained as a convex combination between the NL estimation and the noisy empirical covariance:

$$\widehat{\mathbf{\Sigma}}_{\mathrm{NLRB}}(l) = \widehat{\mathbf{\Sigma}}_{\mathrm{NL}}(l) + \alpha_{\mathrm{RB}} \left(\mathbf{C}(l) - \widehat{\mathbf{\Sigma}}_{\mathrm{NL}}(l) \right) \tag{5.3}$$

  with $\mathbf{C} = \mathbf{k}\mathbf{k}^H$ for single look images and

$$\alpha_{\mathrm{RB}} = \max_j \left[\max \left(0, \frac{\widehat{\mathrm{Var}}\left[I_j\right]_{\mathrm{NL}}(l) - \widehat{I}_{j\mathrm{NL}}(l)^2/L}{\widehat{\mathrm{Var}}\left[I_j\right]_{\mathrm{NL}}(l)} \right) \right], \tag{5.4}$$

  where $I_j(x) = \mathbf{C}_{j,j}(x)$ and $\widehat{I}_{j\mathrm{NL}}(l) = \left[\widehat{\mathbf{\Sigma}}_{\mathrm{NL}}\right]_{j,j}(l)$, $j = 1, 2, 3$. Note that the value of α_{RB} is chosen using the strategy of the LLMMSE introduced in [106].

- **Aggregation:** The final ("best") estimation is locally selected basing on the variance reduction procedure introduced in [154]. The equivalent number of looks (Eq. (1.82)) after the bias reduction step is equal to

$$\widehat{L}_{\mathrm{NLRB}}(l) = \frac{\widehat{L}_{\mathrm{NL}}(l)}{(1-\alpha_{\mathrm{RB}})^2 + \left(\alpha^2 + \frac{2\alpha_{\mathrm{RB}}(1-\alpha_{\mathrm{RB}})}{\sum_{l'} \omega_{\mathrm{NLSAR}}(l,l')}\right) \widehat{L}_{\mathrm{NL}}(l)}. \tag{5.5}$$

  Finally, the estimate with the maximal value of $\widehat{L}_{\mathrm{NLRB}}$ is selected as the best estimate.

5.3 M-NL

In this section we present "step-by-step" the proposed method for improving the weight computation in NL-SAR and discuss about its benefits.

5.3.1 Robust pre-estimation

In order to compute the (dis)similarity between two pixels, one needs to compute a pre-estimation of the scatter matrix. In NL-SAR, this pre-estimation is done using truncated Gaussian on patches of size $\mathbb{S} = (2\mathrm{s}+1) \times (2\mathrm{s}+1)$. In this method, we propose to use an M-estimator instead. It is important to choose an estimator that estimates the shape of the matrix but also keeps the information on scale. If not, two matrices with same shape but different scales will be classified as similar, i.e. in the same class (e.g. in a clustering process). Consequently, we can propose an M-estimator that represents a sort of trade-off between the SCM and the Tyler's M-estimator. In Chapter 2, we have seen that all analyzed M-estimators meet this requirement for specific values of their parameters. For instance, we can use the Student's M-estimator (Definition 1.40) given as the solution of

$$\widehat{\mathbf{\Sigma}}_t = \frac{p+\nu/2}{\mathrm{S}} \sum_{i=1}^{\mathrm{S}} \frac{\mathbf{k}_i \mathbf{k}_i^H}{\nu/2 + \mathbf{k}_i^H \widehat{\mathbf{\Sigma}}_t^{-1} \mathbf{k}_i}, \tag{5.6}$$

where $\mathrm{S} = (2\mathrm{s}+1)^2$, $p = 3$ is the vector dimension and ν is the DoF parameter. We recall that for $\nu \to \infty$, this estimator produces the SCM, while when $\nu = 0$ the Tyler's estimator is obtained.

5.3.2 Pixel selection

Using the pre-estimated values, neighboring samples are selected around each pixel. The central pixel at location l is compared to all pixels in a circular window following a spiral path (see [46] for more details). To compute the dissimilarities between two pixels instead of classical generalized likelihood ratio tests (GLRT), we propose to use the Box's M-test defined as

$$\mathcal{L}_{\text{Box}} = \frac{\left|\widehat{\mathbf{\Sigma}}_1\right|^{\text{S}/2}\left|\widehat{\mathbf{\Sigma}}_2\right|^{\text{S}/2}}{\left|\widehat{\mathbf{\Sigma}}\right|^{\text{S}}},$$

where $\widehat{\mathbf{\Sigma}}_1$ is obtained by Eq. (5.6) with a sample $\mathbf{k}^{(1)} = (\mathbf{k}_1, \ldots, \mathbf{k}_{\text{S}})$, $\widehat{\mathbf{\Sigma}}_2$ with $\mathbf{k}^{(2)} = (\mathbf{k}_{\text{S}+1}, \ldots, \mathbf{k}_{2\text{S}})$ and $\widehat{\mathbf{\Sigma}}$ with $\mathbf{k} = (\mathbf{k}^{(1)}, \mathbf{k}^{(2)})$. This statistic has values between 0 and 1, where the values close to 0 reject the hypothesis that the matrices $\widehat{\mathbf{\Sigma}}_1$ and $\widehat{\mathbf{\Sigma}}_2$ are equal and values close to 1 accept it. By modifying the statistic $\mathcal{L}_{\text{Box}}$, Box has obtained the approximated χ^2 distribution [17]

$$u = -2(1-\beta)\ln\left(\mathcal{L}_{\text{Box}}\right) \sim \chi^2(p(p+1)/2)$$

with $\beta = \frac{3}{2\text{S}}\frac{2p^2+3p-1}{6(p+1)}$, where $p = 3$ is the size of the scattering vector. This result has been obtained for Wishart-distributed matrices. Thanks to the results obtained in Chapter 2, we propose to use this approximation for the statistic computed with M-estimators. We then propose to compute the similarity between two patches centered in pixels l and l' as

$$\Delta_{\text{MNL}}(l, l') = \sum_{\tau} u\left[(l+\tau), (l'+\tau)\right], \tag{5.7}$$

where $\tau \in [-\text{p}, \text{p}]$ is a 2-D shift indicating the location within each patch of size $\mathbb{P} = (2\text{p}+1) \times (2\text{p}+1)$. We then compare the dissimilarities to a threshold in order to select similar pixels. Under the hypothesis H_0 (the two patches follow the same distribution) Δ has the χ^2 distribution with $d = 6(2\text{p}+1)^2$ DoF. The critical region of the test is then given by

$$\left\{R_c = \Delta_{\text{MNL}}, \Delta_{\text{MNL}} > \chi^2_{P_{fa}}(d)\right\}$$

with P_{fa} the probability of false alarm and $\chi^2_{P_{fa}}(d)$ the quantile of order $1 - P_{fa}$ of the $\chi^2(d)$.

5.3.3 Weight computation

Once similar pixels are chosen, we proceed to the weights computation. To define the weights from the dissimilarity measure $\Delta_{\text{MNL}}(l, l')$, we propose to use an exponential kernel

$$\omega_{\text{MNL}}(l, l') = \begin{cases} \exp\left(-\frac{\left|\Delta_{\text{MNL}}(l,l') - c_{\text{MNL}}\right|}{\lambda_{\text{MNL}}}\right) & \text{if } l \neq l', \\ 1 & \text{if } l = l'. \end{cases} \tag{5.8}$$

The parameter $c_{\text{MNL}} = \mathbb{E}\left[\Delta_{\text{MNL}}(l, l') | H_0\right]$ is the expected dissimilarity of two patches under H_0 and the threshold λ_{MNL} can be computed as $\lambda_{\text{MNL}} = F^{-1}_{\chi^2(d)}(1 - P_{fa})$. This mapping from the (dis)similarities to the weights prevents any pixel from having a larger weight than the central pixel. The parameter c_{MNL} has been introduced in order to give a weight close to 1 when the compared pixels comes from the same distribution while preventing the noise enforcing. We normalize the quantity with λ_{MNL} in order to obtain comparable weight values for different values of p.

Finally, the weighted maximum likelihood estimator is computed using Eq. (5.2).

Algorithm 2: M-NL method [J4]

Initialization: $\mathbb{W}, \mathbb{P}, \mathbb{S}, \lambda_{\text{MNL}}, c_{\text{MNL}}, \nu$
forall x, y **do**
  for s $\in \mathbb{S}$ *(scale size)* **do**
    Pre-estimation with Eq. (5.6)
forall x, y *(coordinates of pixel l)* **do**
  for w $\in \mathbb{W}$ *(search window size)* **do**
    Compute Δx and Δy
    $x' = x + \Delta x$
    $y' = y + \Delta y$; *(coordinates of pixel l')*
    for s $\in \mathbb{S}$ **do**
      for p $\in \mathbb{P}$ *(patch size)* **do**
        Compute $\Delta_{\text{MNL}}(l, l')$ with Eq. (5.7)
        if $\Delta_{\text{MNL}}(l, l') \leq \lambda_{\text{MNL}}$[p] **then**
          Compute $\omega_{\text{MNL}}(l, l')$ with Eq. (5.8)
        else
          $\omega_{\text{MNL}}(l, l') \leftarrow 0$
  forall s, p, w **do**
    Compute $\widehat{\mathbf{\Sigma}}_{\text{NL}}$ with Eq. (5.2)
    Bias-reduction step $\rightarrow \widehat{\mathbf{\Sigma}}_{\text{NLRB}}$ Eq. (5.3)
  return The best estimate

The method is recapped in Algorithm 2. First, the maximum sizes of search windows $\mathbb{W}$, patch $\mathbb{P}$ and pre-estimation scale $\mathbb{S}$ are set together with the threshold λ_{MNL} and the constant c that differs for each patch size. Then, the pre-estimation is performed for all pixels in the image and for all values of s $\in \mathbb{S}$ where s=0 means the pre-estimation is off, i.e. the matrix is equal to $\mathbf{k}\mathbf{k}^H$ for the pixel at location l with coordinates (x, y). Then, for all window sizes the central pixel is compared to all pixels in the window using the pre-estimations corresponding to different values of s and different sizes of patches to perform the patch comparison. Then, for each triple of (s, p, w) an estimate $\widehat{\mathbf{\Sigma}}_{\text{NL}}$ is computed. Afterwards, for each $\widehat{\mathbf{\Sigma}}_{\text{NL}}$ the bias-reduced estimate $\widehat{\mathbf{\Sigma}}_{\text{NLRB}}$ is obtained and finally, the best one (with highest number of looks) is selected for each pixel giving the final filtered image. This final part, enclosed within the box, is the same as in NL-SAR.

5.4 Experiments: implementation and evaluation

5.4.1 Simulated data

In this section the results obtained for simulated and PolSAR data are presented. The simulated images have been generated using a Markov Random Field (MRF) following a Gibbs distribution as in [56]. Then, a polarimetric behavior has been assigned to the different parts of the designed images. The polarimetric signatures have been sampled from the NASA/JPL AIRSAR, four-looks, L-band POLSAR data set of San Francisco and are taken as the centroids of standard partition of the $H/\overline{\alpha}$ plane. The polarimetric signatures are presented in Table 5.1 [56].

The procedure is the following: First, a random number C of polarimetric classes is chosen between 3 and 5, $C-1$ classes for distributed scatterers and the last class for point

C1	$\begin{pmatrix} 5.56 & -0.03 - j0.36 & 0.47 - j0.24 \\ -0.03 + j0.36 & 6.64 & 0.24 - j0.20 \\ 0.47 + j0.24 & 0.24 + j0.20 & 4.53 \end{pmatrix}$
C2	$\begin{pmatrix} 7.79 & -0.03 - j0.50 & 0.56 - j0.30 \\ -0.03 + j0.50 & 5.38 & 0.20 - j0.17 \\ 0.56 + j0.30 & 0.20 + j0.17 & 4.38 \end{pmatrix}$
C3	$\begin{pmatrix} 14.69 & 2.59 - j0.92 & 1.98 - j0.85 \\ 2.59 + j0.92 & 25.394.55 + j0.20 & \\ 1.98 + j0.85 & 4.55 - j0.20 & 5.12 \end{pmatrix}$
C4	$\begin{pmatrix} 10.95 & 0.42 - j0.89 & 1.17 - j0.65 \\ 0.42 + j0.89 & 7.51 & 0.83 \\ 1.17 + j0.65 & 0.83 & 3.29 \end{pmatrix}$
C5	$\begin{pmatrix} 10.99 & -0.45 - j0.69 & 0.85 - j0.73 \\ -0.45 + j0.69 & 3.38 & 0.21 + j0.01 \\ 0.85 + j0.73 & 0.21 - j0.01 & 2.05 \end{pmatrix}$
C6	$\begin{pmatrix} 29.95 & 23.04 + j0.79 & 4.83 - j2.47 \\ 23.04 - j0.79 & 29.99 & 5.21 - j3.0 \\ 4.83 + j2.47 & 5.21 + j3.0 & 3.23 \end{pmatrix}$
C7	$\begin{pmatrix} 5.40 & -1.14 - j0.34 & 0.27 - j0.33 \\ -1.14 + j0.34 & 0.56 & -0.01 + j0.09 \\ 0.27 + j0.33 & -0.01 - j0.09 & 0.16 \end{pmatrix}$
C8	$\begin{pmatrix} 990.02 & 4.97 & 7.04 \\ 4.97 & 0.02 & 0.04 \\ 7.04 & 0.04 & 0.05 \end{pmatrix}$

Table 5.1 – Coherency matrix for each simulated class. Class C8 is the point target class

scatters (targets). A ground truth is generated using an MRF and the targets correspond to squares of sizes varying between 2×2 and 5×5 pixels. For each distributed scatterer, one of the seven possible polarimetric signatures is randomly assigned and the Gaussian speckle noise is generated according to them. Finally, the targets generated using the remaining eighth polarimetric signature are added to the speckle noise.

Figure 5.2 details different possible scenarios. First column represents the ground truths obtained with MRF following a Gibbs distribution. The second one displays the corresponding Gaussian speckles to be processed.

After the denoising, the set of following parameters has been evaluated:

- Radiometric parameters σ_r that correspond to the diagonal elements of the estimated scatter matrix containing the power information.
- Complex correlation parameters ρ_r, derived from the three complex off-diagonal terms, containing the complex correlation between three polarimetric channels.
- Incoherent decomposition parameters: Entropy (H) Eq. (1.78), Anisotropy (A) Eq. (1.79) and the mean Alpha angle ($\overline{\alpha}$) Eq. (1.80), containing the information about the physical nature of the scattering mechanism within the resolution cell.
- Co-polar and cross-polar polarization signatures (PS) that contain the information about the polarization synthesis capability of PolSAR data [193].
- Edge preservation (EP).

For the first three groups of parameters the estimated value $\widehat{\theta}$ is obtained as the corresponding mean value from the pixels for every scattering class and for every simulated image, given the filter. Then, the absolute relative bias of the estimated parameter is computed as

$$b_\theta = \frac{\left|\widehat{\theta} - \theta\right|}{\theta}. \tag{5.9}$$

The final (average) value is chosen as the median across all the simulated images and the various scattering classes. In the case of the radiometric parameters and the complex correlation coefficients an additional median operator is applied to provide the final estimated value.

Then, the co- and cross-polar signatures of the average scatter matrix of the pixels belonging to a given scattering class of every simulated image are obtained. The corresponding absolute relative bias, for the pixels of every scattering class and for every simulated image, given the filter is

$$b_\theta = \underset{\phi_i,\varphi_i,\phi_j,\varphi_j}{\text{median}} \left\{ \frac{\left|\widehat{\theta}(\phi_i,\varphi_i,\phi_j,\varphi_j) - \theta(\phi_i,\varphi_i,\phi_j,\varphi_j)\right|}{\theta(\phi_i,\varphi_i,\phi_j,\varphi_j)} \right\}, \tag{5.10}$$

where ϕ and φ are the orientation and ellipticity angles describing the polarization ellipse, whereas i and j refer to the receive and transmit waves, respectively. Then, as in previous case, the median across all the images and classes is obtained. The final parameter is obtained as the mean value of the absolute relative bias of the the co- and cross-polar signatures.

Finally, edge preservation is measured on the boundary positions between extended targets. First, the gradient preservation (GP) is obtained as the average ratio between the observed gradient values on the power bands of the filtered power band image $\widehat{I_j}$ to the gradient values on the ground truth image I_j over the scattering classes

$$GP(i) = \frac{1}{C}\sum_{k=1}^{C} \frac{\sum_{cl(\mathbf{k})=k} \left|\nabla \widehat{I_j}(\mathbf{k})\right|}{\sum_{cl(\mathbf{k})=k} |\nabla I_j(\mathbf{k})|}, \tag{5.11}$$

where ∇ represents the Sobel gradient operator and $cl(\mathbf{k})$ is the class label for the pixel $\mathbf{k}$. Then, a simple mapping from GP to EP is performed , as follows

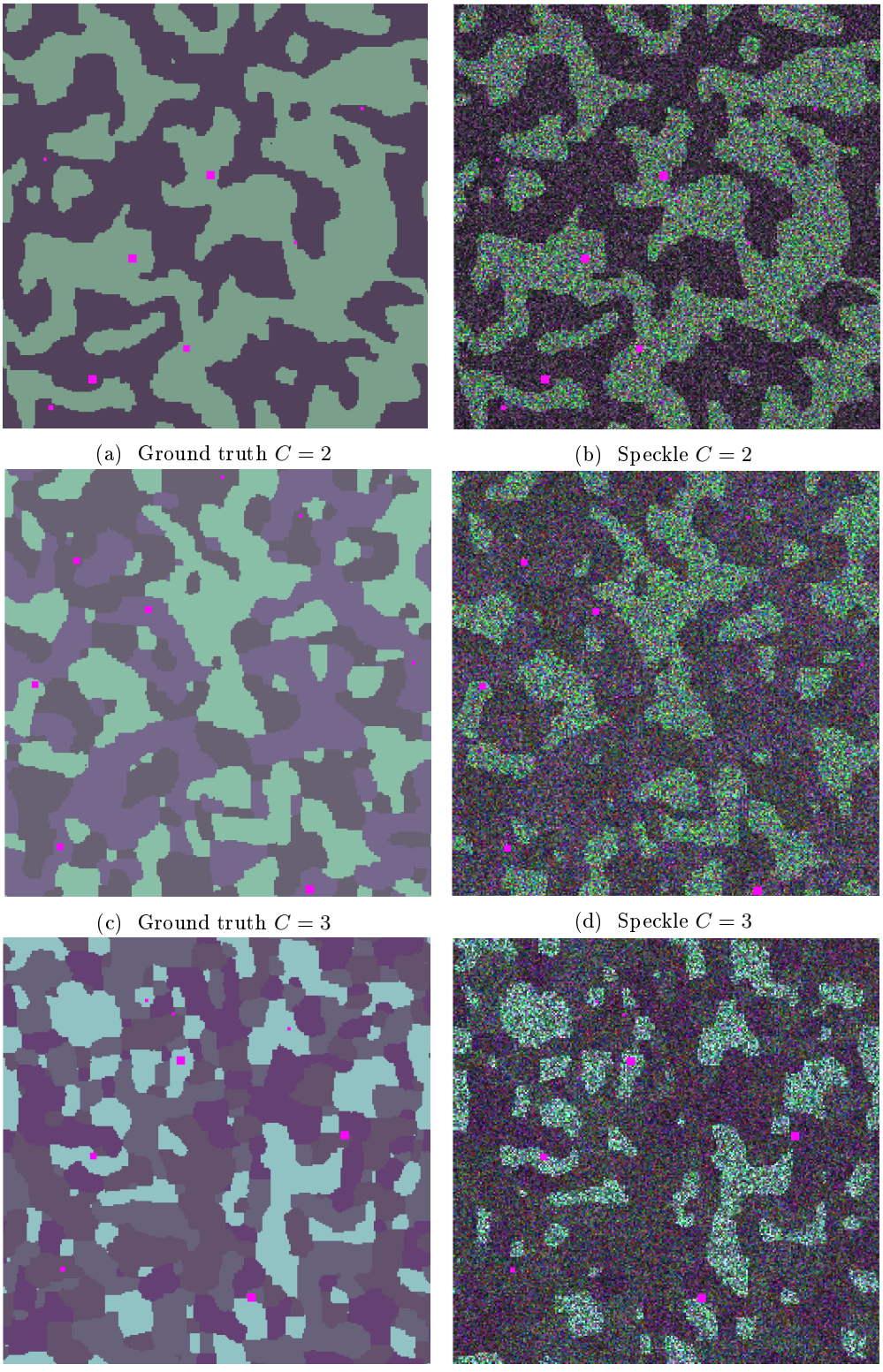
$$EP(i) = \begin{cases} 1 - |1 - GP(i)|, & GP(i) < 2, \\ 0 & GP(i) \geq 2, \end{cases} \tag{5.12}$$

in order to give a measure close to 0 in the case of edge oversmoothing or undersmoothing and values close to 1 for good edge preservation. The final parameter is obtained as the average value over the diagonal elements of the scatter matrix.

In order to perform the comparison, we have simulated one hundred 128×128 artificial PolSAR images as described previously. The set of parameters used in both methods is: window size: $\mathbb{W} \in \{3^2, 5^2, \ldots, 25^2\}$, patch size: $\mathbb{P} \in \{3^2, 5^2, \ldots, 11^2\}$ and scale: $\mathbb{S} \in \{0, 1, 2\}$. Since the speckle is Gaussian, we have chosen ν to be big enough ($\nu = 100$) in order to preserve the information about the texture and ensure the convergence of the solution in the pre-estimation step. The values for λ have been computed using the corresponding formula from Section 5.3.3.

Figure 5.3 shows the results for one realization of simulated images. The images are presented in the following order, from left to right and from top to bottom: original image (ground truth), speckle, NL-SAR results and M-NL results. The ground truth image on Figure 5.3a contains two classes for distributed scatterers. As can be seen from Figure 5.3c et Figure 5.3d the homogeneous areas are much better smoothed with M-NL than with NL-SAR while the edges are better captured and less blurred.

In Figure 5.4 we can see results for another realization. Figure 5.4a consists of three scattering classes. Figures 5.4c and 5.4d display the results obtained with NL-SAR and M-NL, respectively. In this case, we display the filtered images together with the difference of the filtered image and corresponding ground truth in order to better visualize the results. From Figure 5.4e and Figure 5.4f, one can see that in some parts of the images the noise is apparently more reduced with M-NL. This is the most visible at the image borders (green

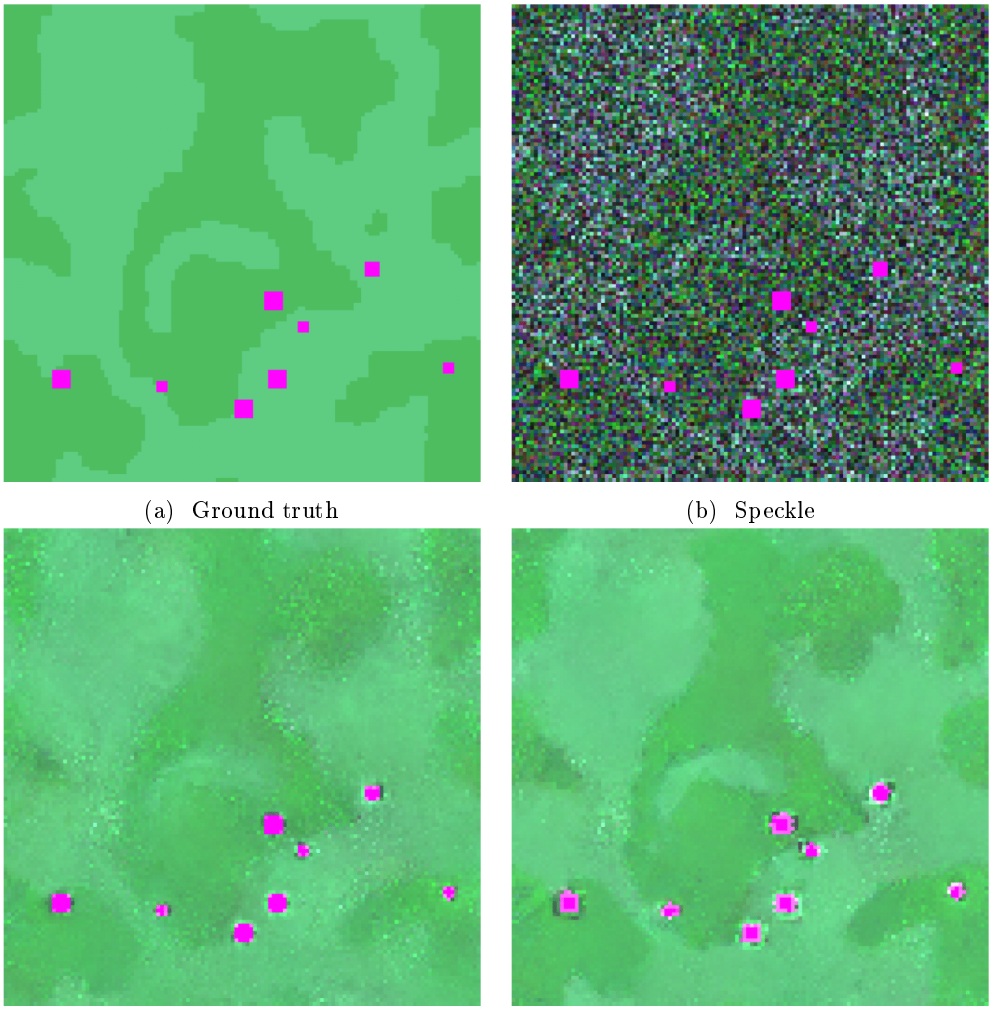


(a) Ground truth $C = 2$ (b) Speckle $C = 2$

(c) Ground truth $C = 3$ (d) Speckle $C = 3$

(e) Ground truth $C = 4$ (f) Speckle $C = 4$

Figure 5.2 – Simulated 256×256 PolSAR images with ground truth (left) and PolSAR speckle (right)



(a) Ground truth (b) Speckle

(c) NL-SAR (d) M-NL

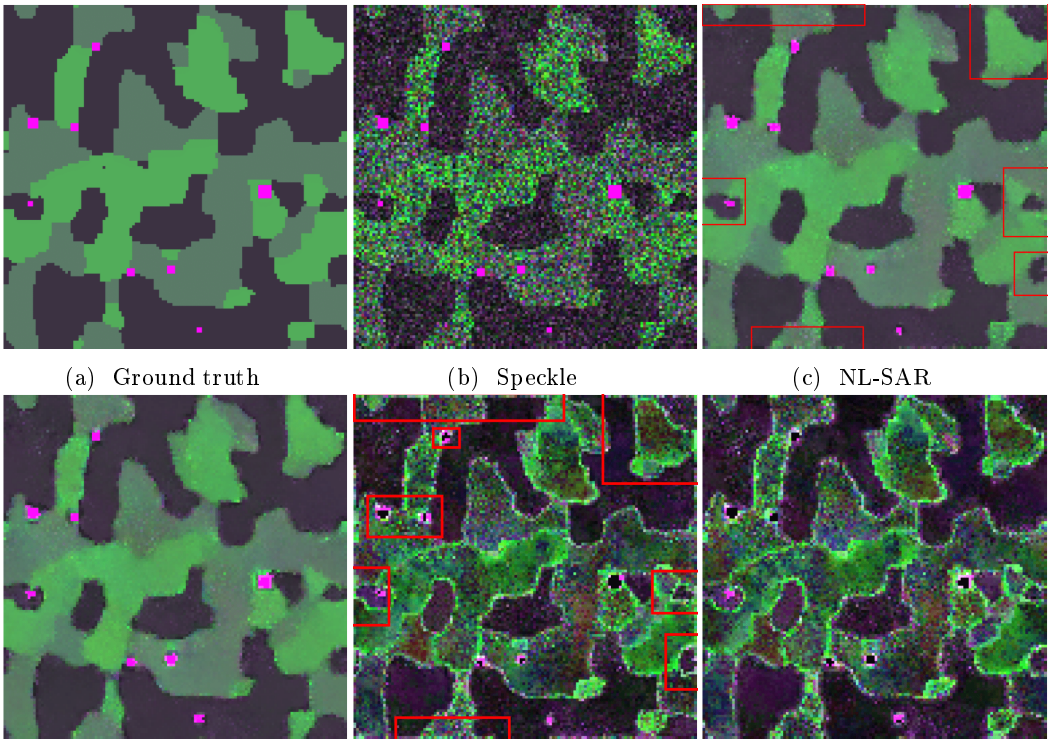
Figure 5.3 – Application to the simulated data - 128×128 images: (a) ground truth, (b) speckle, (c) results obtained with NL-SAR, (d) results obtained with M-NL.

points), one can look for instance the top border and corners of the images. Obviously, the M-NL gives an estimation closer to the ground truth in both cases. The visualization of the difference also reveals that most of the targets are better estimated with M-NL, which can not be seen directly from Figure 5.4c and Figure 5.4d. Some of them are marked in red squares.

Filters	σ	$\lvert\rho\rvert$	$\angle\rho$	H	A	$\overline{\alpha}$	PS	EP
NL-SAR	2.21	**7.47**	**11.96**	14.51	35.84	**10.51**	1.15	0.45
M-NL	**1.56**	9.10	14.47	**14.49**	**33.96**	10.97	**1.05**	**0.56**

Table 5.2 – Filtering results for simulated data: all measures but EP (EP $\in [0, 1]$) are absolute relative errors in %.

Table 5.2 lists the evaluation parameters defined above. Numerical results have been computed over the set of simulated PolSAR images and the final values are compared. One can note that M-NL outperforms NL-SAR in almost all measures except ρ and $\overline{\alpha}$. Thus,



(a) Ground truth (b) Speckle (c) NL-SAR

(d) M-NL (e) NL-SAR difference (f) M-NL difference

Figure 5.4 – Application to the simulated data - 128×128 images: (a) ground truth, (b) speckle, (c) results obtained with NL-SAR, (d) results obtained with M-NL, (e) difference between the results and ground truth for NL-SAR, (f) difference between the results and ground truth for M-NL.

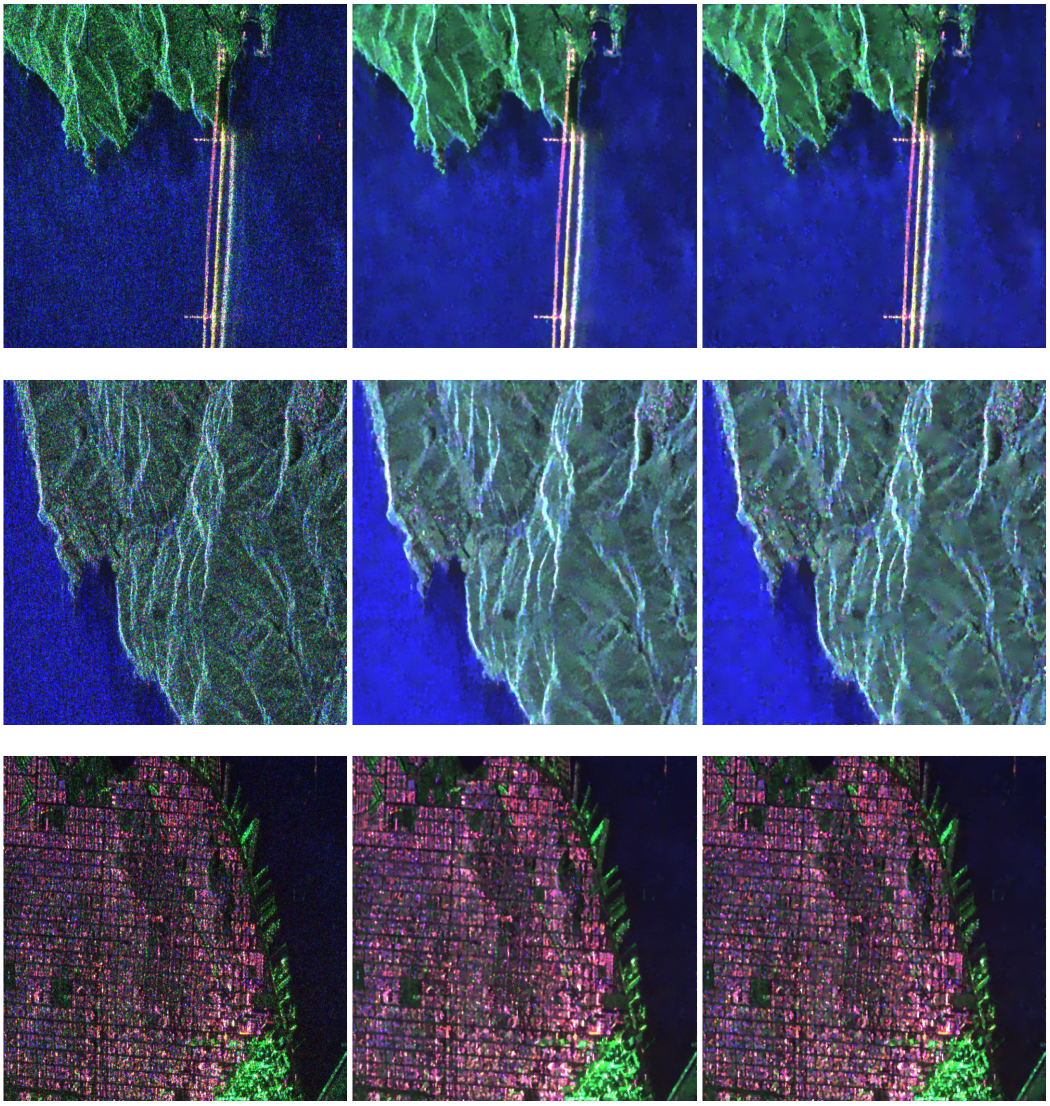
it could be more convenient to use NL-SAR for terrain classification based on correlation coefficient or when measuring soil moisture implementing the alpha angle parameter. On the other hand, M-NL gives better estimation of radiometric parameters, almost all incoherent decomposition parameters, polarization signatures and edge preservation parameters. A significant improvement in edge preservation is also visually noticeable on the simulated images, thus these results are not surprising. Finally, it should be pointed out that all the results are obtained in the case of Gaussian-distributed speckle, which is the most favorable scenario for NL-SAR.

5.4.2 RADARSAT-2 PolSAR data

The results for real data are given in Figure 5.5. Three different parts of San Francisco Bay are presented from top to bottom, representing different scenarios in PolSAR images such as water, vegetation and urban areas. In this case, we do not dispose of any information about the ground truth, thus one can analyze the results only visually. First, one can note that M-NL better smooths the homogeneous areas, while preserving well the edges in textured scenarios. It can also be noted that, as in the case of simulated data, M-NL gives results with higher contrast in comparison to NL-SAR.

5.4.3 Robust kernels

In order to compute the weights we could propose to use different kernels for dissimilarities to weight mapping in Eq. (5.8). Recall the desired characteristics of a kernel function



Speckle NL-SAR M-NL

Figure 5.5 – Real data: San Francisco Bay - 512×512 PolSAR images. From left to right: speckle, results obtained with NL-SAR and results obtained with M-NL.

$\omega(x)$: it is a monotonic function, decreasing with $\max_x \omega(x) = \omega(0)$ and $\omega(x) = 0$ when $x \to \infty$.

In the theory of robust statistics, many functions satisfying the preceding conditions have been proposed [73]. In this work, we propose to analyze the following kernel functions

$$\begin{cases} \omega_1(x) = e^{-x} \\ \omega_2(x) = 1/(1 + x/c_1) \\ \omega_3(x) = e^{-x^2} \\ \omega_4(x) = 1/(1 + (x/c_2)^2) \end{cases} \tag{5.13}$$

These functions are illustrated in Figure 5.6 for x between 0 and 1 (because $\frac{|\Delta_{\mathrm{MNL}}(l,l') - c|}{\lambda_{\mathrm{MNL}}} < 1$). The constants c_1 and c_2 are set to 0.5 and 0.9, respectively. The

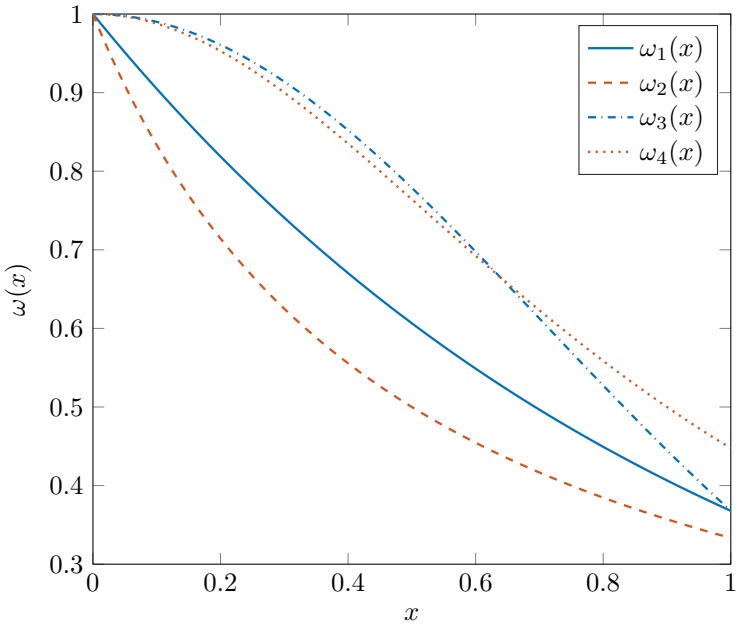


Figure 5.6 – Kernel functions proposed in (5.13) with $c_1 = 0.5$ et $c_2 = 0.9$

function $\omega_1(x) = e^{-x}$ is generally used and this idea comes from the Gaussian hypothesis where the dissimilarities are often calculated as the squared (Euclidean) distance between two vectors.

In the context of PolSAR images, we do not compare two vectors, but the pre-estimated matrices. Therefore, we propose three other kernels different from that of the original algorithm. The $\omega_2(x)$ function has a steeper slope than $\omega_1(x)$ and the weights decrease faster with x. On the other hand, $\omega_3(x)$ and $\omega_4(x)$ put more weight on similar pixels, then from one point, the weight values drop very quickly. This slope is steeper for $\omega_3(x)$, while the values of $\omega_4(x)$ remain quite high.

ω	σ	$\lvert\rho\rvert$	$\angle\rho$	H	$\overline{\alpha}$	A	EP
ω_1	**1.56**	**9.10**	14.47	14.49	10.97	33.96	**0.56**
ω_2	1.6	9.24	14.03	14.49	**10.77**	34.08	0.55
ω_3	**1.56**	9.32	**13.82**	**14.45**	10.79	**33.95**	0.55
ω_4	1.61	9.17	14.49	14.46	10.93	33.98	0.55

Table 5.3 – Results for different kernel functions for simulated data.

It can be noted that ω_1 and ω_3 give a good estimate of the radiometric parameters. ω_1 gives the best estimate of the amplitude of the correlation parameters (PCs) and best preserves the contours. On the other hand, ω_3 gives the best estimate of PC phase, entropy and anisotropy. The best estimate of the alpha angle is given by ω_2. From these results, the kernels ω_1 and ω_3 are preferable to ω_2 and ω_4. In addition, the ω_3 function gives better results than the ω_1 function. As a result, the kernel ω_1 is replaced by ω_3 for the application to real data (Figure 5.7). The images on the left-hand side correspond to the 512×512 speckle images of the San Francisco Bay. The right-hand side displays the results of the despeckling with M-NL. One can note that the speckle is visibly reduced, especially on the homogeneous zones (the water, the forest), while the edges (buildings) are preserved.

Finally, the computational cost of the algorithm haven't been analyzed yet. The whole

method has been coded by modifying the original code of NL-SAR method, in order to well preserve all common parts of the codes. However, we can briefly comment the pros and cons of M-NL in comparison to NL-SAR. In M-NL the weight computation is accelerated thanks to the pixel selection (Section 5.3.2), while in NL-SAR for each pixel the binary search is performed in order to obtain the weight value. However, a major drawback of M-NL is that M-estimators need to be iteratively computed. Also, the formulas include matrix inversion that can be very expensive. For large PolSAR images, the pre-estimation step can take lot of time and thus, this part needs to be optimized.

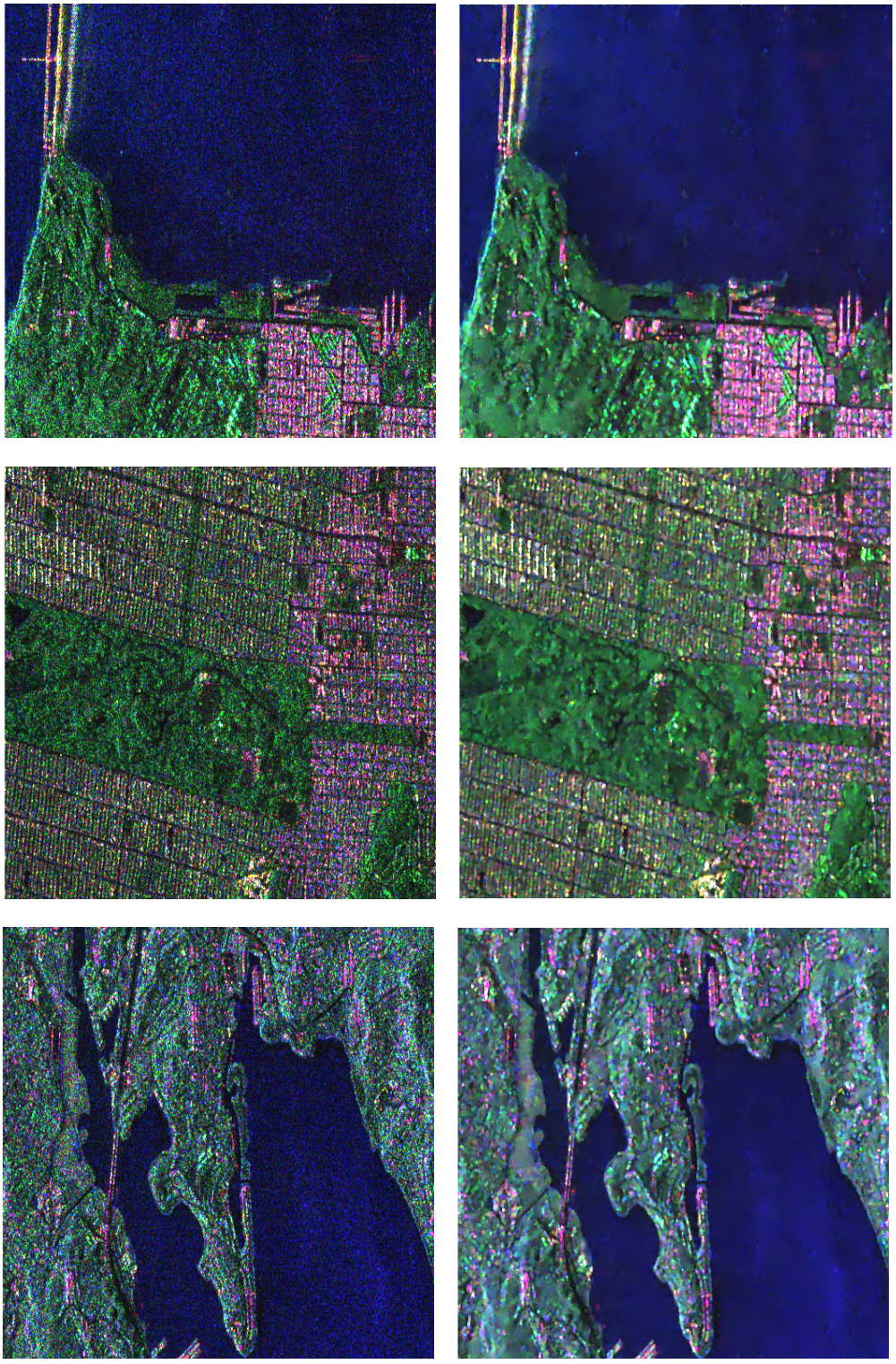
5.5 Conclusion

In this chapter we have introduced a new statistical approach for PolSAR image despeckling, named M-NL. The proposed method relies on M-estimators and has been compared to the NL-SAR method showing better results.

The method is non-local means (NLM) based with common steps with NL-SAR. Yet, the computation of weighted NLM estimates, the crucial part of any NLM denoising method, is based on the statistical behavior of M-estimators. The pre-estimated matrices are obtained using an M-estimator (Student's M-estimator in our experiments) and compared using the criteria obtained for Wishart-distributed matrices. This is possible because of the strong proximity between M-estimators and Wishart-distributed matrices demonstrated in Chapter 2. Several kernel functions are analyzed, starting with the standard exponential kernel. The results have been evaluated on simulated PolSAR data based on Markov Random Field with Gibbs distribution. Various parameters have been compared such as radiometric parameters, channel correlation, entropy, anisotropy etc. We have seen that the M-NL provides smaller relative bias for most parameters showing better performances.

However, the computational cost introduced by the iteratively reweighting algorithm of M-estimators in the pre-estimation step of M-NL can be prohibitive. This cost is even more important for large PolSAR images and higher value of the scale parameter. This problem could be possibly solved by replacing the M-estimator by its "one-step approximation", given by an explicit weighted mean estimator.

Nevertheless, we have seen that M-estimators can be successfully applied in PolSAR image processing. Depending on application, different M-estimators can be employed. The Tyler's M-estimator could be the best choice when only the shape of the matrix is important, as proposed in several works. However, if the information contained in the scale of the matrix needs to be preserved, one could use another M-estimator more adapted to the data model. One of the perspectives for future works could be based on choosing the "best" M-estimator according to the considered problematic.



Speckle M-NL

Figure 5.7 – Real data: San Francisco Bay - 512×512 PolSAR images. PolSAR speckle (left) and the results obtained with M-NL with the kernel ω_3

Conclusions and perspectives

Conclusions

In this thesis, we have dealt with the general problem of covariance matrix estimation for several signal and image processing applications. Under standard assumptions, namely classical Gaussian modeling, the well-known sample covariance matrix (SCM) is computed as it is the efficient Maximum-Likelihood estimator (MLE). This Wishart-distributed estimator is particularly convenient and widely used thanks to its simplicity and explicit definition form. However, with high resolution brought by modern technologies and/or with outliers contained in most of recent datasets, Gaussian models turn out to be incapable to describe the underlying physics of collected data. Indeed, by construction, they present an important lack of robustness and flexibility. Other models, such as complex elliptically symmetric (CES) distributions, come into use as appropriate alternatives. These distributions permit to model various scenarios of highly heterogeneous, outlying and/or missing data. In these circumstances, the classical tools adapted to the Gaussian framework, such as the SCM, exhibit significantly degraded performances. Therefore, the Gaussian-based methods have to be replaced by more robust ones.

In term of covariance matrix estimation, a common solution to this problem is to use the MLE corresponding to the assumed CES distribution. In the context of CES distributions, the covariance matrix is not always defined. In order to obtain the second-order moments, one seek at estimating the scatter matrix that is equal to the scaled covariance matrix when the latter exists. The scatter matrix reflects the correlations between observation vector components, and is always well-defined, even when the distribution turns out to have infinite variance, justifying the interest of focusing on this parameter.

Nevertheless, observed data cannot be always perfectly modeled by a particular CES distribution. The observations can follow different patterns corresponding to various models. Hence, the data can be simply considered as Gaussian corrupted by a certain amount of outliers to reject. This is similar to the problem of considering a heavy-tailed distribution, with unknown shape. In this framework, the robust M-estimators, designed to mitigate the influence of the outlying data, are of great interest. More precisely, M-estimators, by definition, provide robust solutions to both problems: data with outliers and mis-modeling data distributions. These estimators, obtained by iterative reweighting algorithms, assign different importance to each observation depending on its magnitude. However, the statistical behavior of these estimators stayed unknown because of their implicit definition. Consequently, the main goal of this thesis has been to confront this challenge and to provide new insights into the statistical properties of M-estimators, i.e. to characterize their behavior better than any other analysis in the literature. Then, it has to be pointed out that the proposed methodology to answer this key problem has been generalized to various problems in signal and image processing, e.g. to robust detectors, robust distances (Mahalanobis distance), robust eigenvalues/eigenvectors estimators and also robust denoising methods. Of course, the obtained results are in favor of extensions of this framework in other applications.

In this work, it has been shown that the statistical properties of an M-estimator can be well-described with a Wishart distribution. Moreover, it has been revealed that the eigenvalue/eigenvector decomposition (EVD) parameters of an M-estimators behave similarly to the corresponding EVD parameters of a Wishart-distributed matrix. The same conclusions have been drawn for the principal subspaces obtained from M-estimators. Several applications of these results, such as optimal low-rank filtering, rank estimation, have been proposed. Moreover, the results have been extended to the adaptive signal detectors. It has been demonstrated that the properties of a robust detector in the CES framework, obtained with an M-estimator plugged-in instead of the traditional SCM, behaves as the corresponding Gaussian-based detection statistics. Finally, the application of the new statistical properties of M-estimators to polarimetric synthetic aperture radar (PolSAR) image despeckling has been presented. Following previous elements, one can draw a general conclusion for this thesis: **Always prefer the use of robust techniques since they exhibit good performances when applied to real datasets and can be accurately characterized thanks to very simple models.**

Chapter 1 has been dedicated to the state of the art. In the first part, the main principles of covariance matrix estimation with the application to the signal detection problem and polarimetric SAR imaging have been recapitulated. After some basic definitions and assumption made throughout this thesis, the classical Gaussian approach was presented. The SCM and its properties were recalled. Then, the CES distributions were presented with several examples and discussions on their properties. The robust framework together with detailed analysis of different M-estimators has also been presented. We have seen that studied M-estimators can be considered as a particular trade-off between the SCM and the Tyler's M-estimator. This trade-off corresponds to the classical compromise between efficiency and robustness in general statistics problems. The second part introduced the signal detection problem. After a brief introduction about different signal detection scenarios, several Gaussian-based detection statistics were itemized. Their statistical properties in single steering vector case were recalled followed by the corresponding detection probability and probability of false alarm expression. Finally, the last part started with recapitulating the basics of SAR technology and SAR polarimetry principles. We have seen that the CES distributions are widely used to model PolSAR data, since they nature can be well-described with multiplicative models (texture $\times$ speckle). Finally, a concise overview of PolSAR despeckling methods concluded the first chapter.

In Chapter 2, we have introduced a new approach to analyze the statistical properties of M-estimators. To that end, we have proposed a new *Gaussian core* representation of CES data. We have defined a *Gaussian-Core Wishart Equivalent (GCWE)* as the SCM built with Gaussian cores of the observed CES data. Although this theoretical tool cannot be observed and computed, it plays a key role in the analysis. Indeed, we have derived the asymptotic distribution of the difference between an M-estimator built with CES data and its GCWE. The second-order statistics of this difference turn out to be significantly lower than the ones of the M-estimator in the standard asymptotic regime, i.e. when compared to the true scatter matrix parameter. These results revealed that the statistical properties of an M-estimator can be well-approximated with the ones of its GCWE, i.e. with a Wishart distribution. Moreover, the results for special cases of M-estimators have been derived to support these claims and have been validated by various simulation experiments.

In Chapter 3, an extension of these results to the EVD parameters and principal subspaces obtained from M-estimators has been proposed. Several cases, based on the underlying structure of the scatter matrix, have been analyzed. Under the assumption that the scatter matrix contains strictly different eigenvalues, the distribution of eigenvalues and eigenvectors obtained from M-estimators has been analyzed. In the case when

the eigenvalues of the scatter matrix can be grouped in few blocks containing equal eigenvalues, a robust method for grouping eigenvalues, named *eFusion*, has been proposed. The method was first introduced for Gaussian-distributed data, and then, thanks to the previously derived results, extended to CES data. This method has shown an important improvement on simulated data, compared to existing methods. Finally, in the case when the scatter matrix contains few high-valued eigenvalues (e.g. radar clutter), the principal subspace corresponding to these eigenvalues has been analyzed. Its estimation obtained from M-estimators has been studied, leading to the same conclusions as previously. The theoretical results have been validated on simulated data with various settings.

In Chapter 4, we have addressed the problem of robust signal detection. Robust detectors were defined as traditional detection statistics built with M-estimators instead of the classical SCM. Following the idea of Chapter 2, a *Gaussian-Core Equivalent Detector (GCED)* of a robust detector has been introduced. Then, in addition to standard asymptotic distribution of robust detectors, the asymptotic distribution of the difference between a robust detector and its GCED has been analyzed. The results revealed that the behavior of a robust detector can be well-approximated with the one of its GCED. This is of particular interest, since it allows one to use an M-estimator to obtain an accurate estimation of the scatter matrix of CES data, while relying on the statistical properties of its GCED when tuning the detection parameters, e.g. the detection threshold. Robust detection analysis has been performed in the classical settings as well as in the low-rank setting, that is of particular interest when dealing with high-dimensional datasets. Finally, the same conclusions obtained for Full-Rank and Low-Rank detectors, also hold for robust Mahalanobis distances used in anomalies detection, data clustering, etc.

Finally, in Chapter 5, we have focused on non-local means (NLM) PolSAR despekling. As explained in Chapter 1, the PolSAR data are usually modeled by various CES distributions. Basing on the results obtained in Chapter 2, we have proposed to use M-estimators in order to estimate the second-order statistics of the data during the pre-estimation step in NLM methods. In particular, we have used the Student's M-estimator and statistical tools derived for Wishart-distributed matrices while searching for similar pixels in an image. Student's M-estimator was chosen as an example of M-estimators that keeps information on both shape and scale of the scatter matrix. It is also important to notice that the contribution of this chapter also relies on the use of the well-known Box's test for deriving similarities between pixels, which is possible because of the derived properties of the M-estimators. The new, M-NL, method showed good performances in speckle reduction on both simulated and real PolSAR images, suggesting that M-estimators can be successfully applied in various PolSAR image applications.

The main conclusion, arising from the analyses done in each chapter of this thesis, is that one should always use techniques robust to different deviations from the assumed data model, i.e., heterogeneity, outliers, missing data etc., and rely on the statistical tools adapted to Gaussian-distributed data. In addition to various experiments supporting this claim, this thesis has provided a unified theoretical framework to justify this approach.

Pespectives

Following previous conclusions, the results of this thesis can be extended to different problematics and can be applied in numerous signal processing and machine learning applications. As mentioned previously, this work provided non-exhaustive results. Indeed, for instance, in several machine learning problems, it is essential to characterize distances

behavior. This is the case for clustering problems. This short-term perspective also applies for detection problems where numerous detector statistics can be found in the literature. This is also the case for problems involving EVD parameters, applied in dimension reduction problems, as detailed hereafter.

Nowadays, the analyzed data sets are usually very large meaning that, in most cases, the data have a high dimensional nature. In this context, the analysis done in this thesis could be extended to the estimators adapted to high dimensional scenarios. In particular, one could analyze regularized covariance matrix estimators obtained from different penalization functions. This could potentially provide new ideas on how to tune the regularization parameters, in order to obtain specific performances of regularized M-estimators. The regularization parameters estimation is a difficult problem involved in many domains and could be tackled with the proposed framework. Furthermore, all the results in this thesis have been obtained assuming a fixed data vector dimension. One could possibly consider the cases when the data dimension grows together with the sample size. Even though some results for this set-up have been already obtained using random matrix theory, one could possibly try to exploit similar ideas in order to improve the existing methods. In general, following high-dimensional settings, one could also exploit the structure of the data in order to work in a low dimensional manifold and derive new robust estimation solutions.

Another aspects of this work that has not received an important attention is the characterization of the loss brought by robust approaches over the optimal ones. This loss over optimal efficient estimation methods can be derived from the proposed characterization of M-estimators properties, by considering the asymptotic coefficients obtained in this thesis. Indeed, these coefficients can be viewed as a notion of distance to the Gaussian tools and could be exploited to characterize this loss. This is of particular importance when dealing with problems of mis-modeling.

The results of Chapter 3 can be of a great interest in machine learning applications. In particular, the EVD parameters of the scatter matrix are widely used in dimensionality reduction problems and data visualization. As various data clustering methods, these techniques search to measure similarities between observations. Including robust techniques in these algorithms could help to capture more useful information by low-dimensional projection of the data and improve clustering performances while rejecting the outlying data. The results obtained for the Mahalanobis distance could be possible extended to a wide range of distances (between vectors/matrices) used in literature, such as, for instance, the Riemannian manifold-oriented distance analyzed in Chapter 3.

As briefly discussed in Chapter 5, an important problem to be tackled is the computational cost induced by iterative reweighting algorithms used to compute the M-estimators. The convergence speed to obtain any M-estimator depends on a particular choice of tuning parameters as well as on the nature of the underlying data. Consequently, depending on the application problem different solutions may be proposed. For instance, when the data are supposed to be identically distributed with a certain amount of outliers and/or missing data, one could try to derive analogous algorithms with a reduced number of iterations or, for instance, derive explicit weighted estimators with weights close to the ones obtained in the final iteration of the corresponding M-estimator. In machine learning techniques, such as data clustering, these weights should contain information about similarity between different observations. Originally, M-estimators were introduced to reduce the influence of outlying values with high magnitude. In clustering methods, the outlying data to be rejected correspond to the data with high dissimilarity when compared to the data to be processed. Similarly to the proposed despeckling method in Chapter 5, one could build weighted covariance matrix estimators using different robust techniques to compute the weights.

Still on the computational cost reduction, recent works have suggested to work in the appropriate manifold and derive new estimators. Since covariance matrices are, by definition, Hermitian (or symmetric) positive definite matrices, they naturally live in a Riemannian manifold. Tackling the estimation problems, and consequently the fixed-point algorithms, by using appropriate distances (non-Euclidean ones) opens a wide range of new problems for long-term research perspectives.

Moreover, different modifications, potentially improvements, of the M-NL method can be studied. The results have been obtained using statistical tools derived for independent Wishart-distributed matrices, which, in general, is not true for pre-estimated scatter matrices. Thus, according to the "level" of correlation, one could modify the value of equivalent number of looks, use different statistical models to compute the detection threshold, etc. More generally, the assumption of independent data should be relaxed.

To conclude, the perspectives of this work are various and quite diverse. The problem of non-Gaussianity is currently emerging in many areas and robust tools are of crucial importance. This study is only one step towards enhancing our understanding of their behavior...

Appendix A

Matrix operations

In the following we provide some quantities that has been used during the results derivations in this thesis [144]:

$$\begin{aligned}
\mathrm{vec}(\mathbf{ABC}) &= \left(\mathbf{C}^T \otimes \mathbf{A}\right) \mathrm{vec}(\mathbf{B}), && \text{(A.1)}\\
(\mathbf{A} \otimes \mathbf{C})(\mathbf{B} \otimes \mathbf{D}) &= (\mathbf{AB} \otimes \mathbf{CD}), && \text{(A.2)}\\
\mathrm{vec}(\mathbf{A})^H \mathbf{K} &= \mathrm{vec}(\mathbf{A})^T, && \text{(A.3)}\\
\mathbf{K}\left(\mathbf{A} \otimes \mathbf{B}\right) &= \left(\mathbf{B} \otimes \mathbf{A}\right)\mathbf{K}, && \text{(A.4)}\\
\partial \mathbf{A}^{-1} &= -\mathbf{A}^{-1}\partial\mathbf{A}\mathbf{A}^{-1}, && \text{(A.5)}\\
\partial \,\mathrm{Tr}\left(\mathbf{A}\right) &= \mathrm{Tr}\left(\partial\mathbf{A}\right), && \text{(A.6)}\\
\mathrm{Tr}\left(\mathbf{AB}\right) &= \mathrm{Tr}\left(\mathbf{BA}\right), && \text{(A.7)}\\
\mathrm{Tr}\left(\mathbf{A}^H\mathbf{B}\right) &= \mathrm{vec}^H\left(\mathbf{A}\right)\mathrm{vec}\left(\mathbf{B}\right), && \text{(A.8)}\\
\left(\mathbf{A} \otimes \mathbf{B}\right)^{-1} &= \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}, && \text{(A.9)}
\end{aligned}$$

Appendix B

Real Elliptically Symmetric distributions

In this part, we define the RES distributions and present the relationship between them and CES distributions.

Definition B.0.1. *Real elliptically symmetric (RES) distribution*
A random p-dimensional vector $\mathbf{z}$ is said to be elliptically distributed if and only if there exists a vector $\boldsymbol{\mu} \in \mathbb{R}^p$, a positive semidefinite matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{p \times p}$, and a function $\phi : \mathbb{R}_+ \to \mathbb{R}$ such that the characteristic function $t \to \varphi_{z-\mu}(t)$ of $\mathbf{z} - \boldsymbol{\mu}$ corresponds to $t \to \phi(t^T \boldsymbol{\Sigma} t)$, $t \in \mathbb{R}^p$. This case is denoted as $\mathcal{ES}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$.

Similarly to the complex case, the RES-distributed vectors admit the following stochastic representation.

Theorem B.0.1. *Stochastic representation*
An r.v. $\mathbf{z} \sim \mathcal{ES}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ *if and only if it can be represented as*

$$\mathbf{z} \overset{d}{=} \boldsymbol{\mu} + \sqrt{\mathcal{Q}} \mathbf{A} \mathbf{u}^{(k)}, \tag{B.1}$$

where the non-negative real random variable $\mathcal{Q}$*, called the modular variate, is independent of the r.v.* $\mathbf{u}^{(k)}$ *that is uniformly distributed on the unit* k*-hypersphere* $\mathcal{S}^{k-1}$ *with*

$$\mathcal{S}^{k-1} = \left\{ \mathbf{u} \in \mathbb{R}^k : \|\mathbf{u}\|_2 = 1 \right\} \tag{B.2}$$

and $\mathbf{A} \in \mathbb{R}^{p \times k}$ *with* $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T$.

In the absolutely continuous case, we have the following definition.

Theorem B.0.2. *Let* $\mathbf{z} \sim \mathcal{ES}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$*. Let* $\boldsymbol{\Sigma}$ *be positive definite and the c.d.f. of* $\mathcal{R}$ *be absolutely continuous. Then the p.d.f. of* $\mathbf{z}$ *can be written as*

$$f_{\mathbf{z}}(\mathbf{z}) = c_{m,g} |\boldsymbol{\Sigma}|^{-1/2} g_{\mathbf{z}} \left((\mathbf{z} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \boldsymbol{\mu}) \right) \tag{B.3}$$

with

$$\begin{cases} c_{p,g} = (s_p \mu_{p-1,g})^{-1}, \\ s_p = \dfrac{(2\pi)^{p/2}}{\Gamma(p/2)}, \\ \mu_{p,g} = \displaystyle\int_0^\infty t^p g(t^2) dt. \end{cases} \tag{B.4}$$

This case is denoted as $\mathbf{z} \sim \mathcal{ES}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g_{\mathbf{z}})$.

Definition B.0.2. *Generalized complex elliptically symmetric (GCES) distribution*
An r.v. $\mathbf{z} = \mathbf{x} + j\mathbf{y} \in \mathbb{C}^p$ is said to have a GCES distribution if $\mathbf{z}_R = \left(\mathbf{x}^T, \mathbf{y}^T\right)^T \in \mathbb{R}^{2p}$ has a RES distribution.

Then, the p.d.f. of $\mathbf{z}_R$ is given by Eq. (B.3) with

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_\mathbf{x} \\ \boldsymbol{\mu}_\mathbf{y} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{xx}} & \boldsymbol{\Sigma}_{\mathbf{xy}} \\ \boldsymbol{\Sigma}_{\mathbf{yx}} & \boldsymbol{\Sigma}_{\mathbf{yy}} \end{pmatrix}. \tag{B.5}$$

Then

$$\begin{aligned} \boldsymbol{\mu}_\mathbf{z} &= \boldsymbol{\mu}_\mathbf{x} + j\boldsymbol{\mu}_\mathbf{y}, \\ \boldsymbol{\Theta} &= \boldsymbol{\Sigma}_{\mathbf{xx}} + \boldsymbol{\Sigma}_{\mathbf{yy}} - j\left(\boldsymbol{\Sigma}_{\mathbf{yx}} - \boldsymbol{\Sigma}_{\mathbf{xy}}\right), \\ \boldsymbol{\Phi} &= \boldsymbol{\Sigma}_{\mathbf{xx}} - \boldsymbol{\Sigma}_{\mathbf{yy}} + j\left(\boldsymbol{\Sigma}_{\mathbf{yx}} + \boldsymbol{\Sigma}_{\mathbf{xy}}\right). \end{aligned} \tag{B.6}$$

Then, we can write

$$\tilde{\mathbf{z}} = \begin{pmatrix} \mathbf{z} \\ \mathbf{z}^* \end{pmatrix}, \quad \tilde{\boldsymbol{\mu}} = \begin{pmatrix} \boldsymbol{\mu}_{\boldsymbol{\mu}} \\ \boldsymbol{\mu}_{\boldsymbol{\mu}}^* \end{pmatrix}, \quad \tilde{\boldsymbol{\Sigma}} = \begin{pmatrix} \boldsymbol{\Theta} & \boldsymbol{\Phi} \\ \boldsymbol{\Phi}^* & \boldsymbol{\Theta}^* \end{pmatrix}. \tag{B.7}$$

Finally, the p.d.f. of a GCES-distributed vector $\mathbf{z}$, denoted $\mathbf{z} \sim \mathcal{GCES}(\boldsymbol{\mu}, \boldsymbol{\Theta}, \boldsymbol{\Phi}, g_\mathbf{z})$, can be written as

$$f_\mathbf{z}(\mathbf{z}) = c_{p,g} 2^p |\tilde{\boldsymbol{\Sigma}}|^{-1/2} g_\mathbf{z}\left(2\left(\tilde{\mathbf{z}} - \tilde{\boldsymbol{\mu}}\right)^H \tilde{\boldsymbol{\Sigma}}^{-1}\left(\tilde{\mathbf{z}} - \tilde{\boldsymbol{\mu}}\right)\right). \tag{B.8}$$

Definition B.0.3. *CES distribution*
If an r.v. $\mathbf{z}$ is GCES-distributed with a null pseudo-covariance matrix, i.e. $\mathbf{z} \sim \mathcal{GCES}(\boldsymbol{\mu}, \boldsymbol{\Theta}, \mathbf{0}, g_\mathbf{z})$, then it is said to have a CES distribution, denoted as $\mathbf{z} \sim \mathcal{CES}(\boldsymbol{\mu}, \boldsymbol{\Theta}, g_\mathbf{z})$.

Appendix C

Real-valued M-estimators

In this appendix, the asymptotics of real-valued M-estimators (M-estimators built with real-valued data) are provided. First, we recall the asymptotic distributions of the real sample covariance matrix (SCM) and real M-estimators when centering about the true scatter matrix. Then, we derive in Theorem C.0.1 the relationships between two types of estimators. This result represents the real-valued version of the one of Theorem 2.2.1. Then, only the main lines of the proof, that differ from the complex case, are given. Finally, we briefly discuss how these results can be useful to obtain results of Theorem 2.2.1.

Before turning to the main result, let us remind the important definitions, as well as existing results, in the real case. For the sake of simplicity, we will use the same abbreviations and notations as in the complex case.

Definition C.0.1. *Sample Covariance Matrix (SCM)*
Let $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ be an n-sample of p-dimensional complex i.i.d. vectors with $\mathbf{x}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$. The real-valued SCM is defined by

$$\widehat{\mathbf{\Sigma}}_{\text{SCM}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T. \tag{C.1}$$

The SCM (up to a normalization factor) is Wishart-distributed and its asymptotic distribution is given by [13]

$$\sqrt{n}\text{vec}(\widehat{\mathbf{\Sigma}}_{\text{SCM}} - \mathbf{\Sigma}) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, (\mathbf{I} + \mathbf{K})(\mathbf{\Sigma} \otimes \mathbf{\Sigma})\right). \tag{C.2}$$

Definition C.0.2. *M-estimator*
Let $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ be an n-sample of p-dimensional complex i.i.d. vectors with $\mathbf{z}_i \sim \mathcal{ES}(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}})$. An M-estimator, denoted by $\widehat{\mathbf{\Sigma}}$, is defined by the solution of the following M-estimating equation

$$\widehat{\mathbf{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n} u(\mathbf{z}_i^T \widehat{\mathbf{\Sigma}}^{-1} \mathbf{z}_i) \mathbf{z}_i \mathbf{z}_i^T, \tag{C.3}$$

where u is any real-valued weight function on $[0, \infty)$ that respects Maronna's conditions [119]. Then

$$\sqrt{n}\text{vec}(\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}_\sigma) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{C}_M),$$

where

$$\mathbf{C}_M = \vartheta_1 (\mathbf{I} + \mathbf{K}) (\mathbf{\Sigma}_\sigma \otimes \mathbf{\Sigma}_\sigma) + \vartheta_2 \text{vec}(\mathbf{\Sigma}_\sigma)\text{vec}(\mathbf{\Sigma}_\sigma)^T, \tag{C.4}$$

with $\mathbf{\Sigma}_\sigma = \sigma^{-1}\mathbf{\Sigma}$ and the parameter σ is the solution of $\mathbb{E}\left[u(\sigma\mathcal{Q})\sigma\mathcal{Q}\right] = p$ with $\mathcal{Q}$ from Eq. (B.1). The constants ϑ_1 and ϑ_2 are given in [169].

Analogously to the complex case, we introduce Gaussian-core model and the Gaussian Core Wishart Equivalent (GCWE) of an M-estimator.

Definition C.0.3. *Gaussian-core representation*
A real random vector $\mathbf{z} \sim \mathcal{ES}(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}})$ can be represented as

$$\mathbf{z} \overset{d}{=} \frac{\sqrt{\mathcal{Q}}}{\|\mathbf{g}\|}\mathbf{A}\mathbf{g} \tag{C.5}$$

with $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $\mathcal{Q} \overset{d}{=} \mathbf{z}^T\mathbf{\Sigma}^{-1}\mathbf{z}$ and $\mathbf{\Sigma} = \mathbf{A}\mathbf{A}^T$ is a factorization of $\mathbf{\Sigma}$.

Definition C.0.4. *GCWE*
Let n measurements $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ be drawn as $\mathbf{z}_i \sim \mathcal{ES}(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}})$ and denote $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ their Gaussian cores as $\mathbf{z}_i = \sqrt{\mathcal{Q}_i}/\|\mathbf{g}_i\|\mathbf{A}\mathbf{g}_i$ with $\mathbf{x}_i = \mathbf{A}\mathbf{g}_i$ (cf. Definition C.0.3), $i = 1, \ldots, n$. Let $\widehat{\mathbf{\Sigma}}$ be an M-estimator built with $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ using Eq. (C.3). The SCM built from the Gaussian cores, i.e.

$$\widehat{\mathbf{\Sigma}}_{\text{GCWE}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^T \tag{C.6}$$

is referred to as the GCWE of $\widehat{\mathbf{\Sigma}}$.

Now, one has the state the following theorem.

Theorem C.0.1. *GCWE regime for real-valued M-estimators*
Let $\widehat{\mathbf{\Sigma}}$ and $\widehat{\mathbf{\Sigma}}_{\text{GCWE}}$ be defined by Eqs. (C.3) and (C.6), respectively. The asymptotic distribution of $\sigma\widehat{\mathbf{\Sigma}} - \widehat{\mathbf{\Sigma}}_{\text{GCWE}}$ is given by

$$\sqrt{n}\text{vec}(\sigma\widehat{\mathbf{\Sigma}} - \widehat{\mathbf{\Sigma}}_{\text{GCWE}}) \overset{d}{\to} \mathcal{N}(\mathbf{0}, \mathbf{\Gamma}) \tag{C.7}$$

where $\mathbf{\Gamma}$ is defined by

$$\mathbf{\Gamma} = \sigma_1(\mathbf{I} + \mathbf{K})(\mathbf{\Sigma} \otimes \mathbf{\Sigma}) + \sigma_2\text{vec}(\mathbf{\Sigma})\text{vec}(\mathbf{\Sigma})^T \tag{C.8}$$

with

$$\begin{aligned}\sigma_1 &= \frac{a_M p(p+2) + c_M(c_M - 2b_M)}{c_M^2},\\ \sigma_2 &= \frac{a_M - p^2}{(c_M - p^2)^2} - \frac{a_M(p+2)}{c_M^2} + 4\frac{p(c_M - b_M)}{c_M(c_M - p^2)},\end{aligned} \tag{C.9}$$

where $a_M = \mathbb{E}[\Psi^2(\sigma\mathcal{Q})]$, $b_M = \mathbb{E}[\Psi(\sigma\mathcal{Q})\|\mathbf{g}\|^2]$, $c_M = 2\mathbb{E}[\Psi'(\sigma\mathcal{Q})\sigma\mathcal{Q}] + p^2$ with $\Psi(\sigma\mathcal{Q}) = u(\sigma\mathcal{Q})\sigma\mathcal{Q}$.

Remark C.0.1. *Notice that the asymptotic covariance matrix $\mathbf{\Gamma}$ differs from the one in the complex case (Theorem 2.2.1) not only in terms of the scale factors values , but also in terms of its structure. The structure is the same as in classical asymptotic results (Eqs. (C.2) and (C.4)) but the coefficients are different. In the case of the identity matrix as covariance matrix, this structure involves only three non-null elements d_1, d_2 and d_3 at the positions (i, j) and equal to:*

- *$d_1 = 2\sigma_1 + \sigma_2$ for $i = j = q + p(q-1)$ with $q = 1, \ldots, p$,*
- *$d_2 = \sigma_1$ for $i = j = q + p(q'-1)$ and $i = q + p(q'-1), j = q' + p(q-1)$ with $q \neq q'$ and $q, q' = 1, \ldots, p$,*
- *$d_3 = \sigma_2$ for $i = q + p(q-1), j = q' + p(q'-1)$ with $q \neq q'$ and $q, q' = 1, \ldots, p$.*

Remark C.0.2. *Note that the results of Theorem 2.2.1 can be obtained using the results for real M-estimators and vector/matrix complex-to-real mapping [130]. Indeed, following the same steps as in [117] one can easily prove that* $\mathbf{\Gamma}_{CM} = (\mathbf{l}^T \otimes \mathbf{l}^H)\mathbf{\Gamma}(\mathbf{l}^T \otimes \mathbf{l}^H)^H$ *and* $\mathbf{\Omega}_{CM} = \mathbf{\Gamma}_{CM}\mathbf{K}$ *where* $\mathbf{\Gamma}$ *is defined by Eq. (C.8) for* $2p$*-dimensional vectors and* $\mathbf{l}^T = (\mathbf{I}_p, -j\mathbf{I}_p)$ *(*$j = \sqrt{-1}$ *here), which after some mathematical manipulations leads to the final results.*

Proof. If we proceed as in the proof for the complex case, we obtain

$$\mathbf{\Gamma} = \mathbf{\Gamma}_1(\mathbf{\Sigma}) - 2\mathbf{\Gamma}_2(\mathbf{\Sigma}) + \mathbf{\Gamma}_3(\mathbf{\Sigma}), \tag{C.10}$$

where the matrices $\mathbf{\Gamma}_1(\mathbf{\Sigma})$ and $\mathbf{\Gamma}_3(\mathbf{\Sigma})$ are given by Eqs. (C.4) and (C.1), respectively. Analogically to the complex case, one has

$$\mathbf{\Gamma}_2(\mathbf{\Sigma}) = \mathbf{D}_1^{-1}(\mathbf{\Sigma})\mathbf{B}(\mathbf{\Sigma})\left(\mathbf{D}_2^{-1}(\mathbf{\Sigma})\right)^T,$$

where $\mathbf{D}_1(\mathbf{\Sigma}) = \mathbb{E}\left[d\{\mathrm{vec}\left(\Psi_1(\mathbf{\Sigma})\right)\}/d\{\mathrm{vec}(\mathbf{\Sigma})\}\right]$, $\mathbf{B}(\mathbf{\Sigma}) = \mathrm{cov}\left(\mathrm{vec}\left(\Psi_1(\mathbf{\Sigma})\right), \mathrm{vec}\left(\Psi_2(\mathbf{\Sigma})\right)\right)$ and $\mathbf{D}_2(\mathbf{\Sigma}) = \mathbb{E}\left[d\{\mathrm{vec}\left(\Psi_2(\mathbf{\Sigma})\right)\}/d\{\mathrm{vec}(\mathbf{\Sigma})\}\right]$ with $\Psi_1(\mathbf{\Sigma}) = \sigma u(\mathbf{z}^T(\sigma^{-1}\mathbf{\Sigma})^{-1}\mathbf{z})\mathbf{z}\mathbf{z}^T - \mathbf{\Sigma}$ and $\Psi_2(\mathbf{\Sigma}) = \mathbf{x}\mathbf{x}^T - \mathbf{\Sigma}$.

Assuming w.l.o.g. that $\mathbf{\Sigma} = \mathbf{I}$, one has a real-valued expression obtained in [169]

$$\mathbf{D}_1^{-1}(\mathbf{I}) = \alpha_1\mathbf{I} + \alpha_2\mathrm{vec}(\mathbf{I})\mathrm{vec}(\mathbf{I})^T,$$

where $\alpha_1 = -\dfrac{p(p+2)}{c_M}$ and $\alpha_2 = \dfrac{p(c_M - p^2 - 2p)}{c_M(c_M - p^2)}$ with $c_M = 2\mathbb{E}\left[\sigma\mathcal{Q}\Psi'(\sigma\mathcal{Q})\right] + p^2$. Moreover, it is simple to show that $\mathbf{D}_2(\mathbf{I})^{-1} = -\mathbf{I}$.

Analogously to the complex case, one can obtain

$$\mathbf{B}_2(\mathbf{I}) = \beta_1(\mathbf{I} + \mathbf{K}) + \beta_2\mathrm{vec}(\mathbf{I})\mathrm{vec}(\mathbf{I})^T$$

with $\beta_1 = \dfrac{b_M}{p(p+2)}$ and $\beta_2 = \beta_1 - 1$. After some mathematical manipulations, one obtains

$$\mathbf{\Gamma}_2(\mathbf{I}) = \gamma_1(\mathbf{I} + \mathbf{K}) + \gamma_2\mathrm{vec}(\mathbf{I})\mathrm{vec}(\mathbf{I})^T$$

with

$$\begin{aligned} \gamma_1 &= -\alpha_1\beta_1 = \frac{b_M}{c_M}, \\ \gamma_2 &= -(\alpha_1\beta_2 + 2\alpha_2\beta_1 + p\alpha_2\beta_2) = \frac{2p(b_M - c_M)}{c_M(c_M - p^2)}. \end{aligned} \tag{C.11}$$

This leads to the final expression of $\mathbf{\Gamma}_2(\mathbf{\Sigma})$

$$\mathbf{\Gamma}_2(\mathbf{\Sigma}) = \gamma_1(\mathbf{I} + \mathbf{K})(\mathbf{\Sigma} \otimes \mathbf{\Sigma}) + \gamma_2\mathrm{vec}(\mathbf{\Sigma})\mathrm{vec}(\mathbf{\Sigma})^T. \tag{C.12}$$

Finally, combining Eq. (C.12) together with Eqs. (C.4) and (C.2) in Eq. (C.10), one obtains the final result. □

Appendix D

Proof of Theorem 2.2.1

To prove the statement let us rewrite the right hand side of Eq. (2.4) as follows:

$$\begin{aligned}\sqrt{n}\left(\mathrm{vec}\left(\sigma\widehat{\mathbf{\Sigma}}-\widehat{\mathbf{\Sigma}}_{\mathrm{GCWE}}\right)\right) &= \sqrt{n}\left(\mathrm{vec}\left(\sigma\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma}-\widehat{\mathbf{\Sigma}}_{\mathrm{GCWE}}+\mathbf{\Sigma}\right)\right)\\ &= \left[\mathbf{1},-\mathbf{1}\right]\begin{bmatrix}\sqrt{n}\mathrm{vec}\left(\sigma\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma}\right)\\ \sqrt{n}\mathrm{vec}\left(\widehat{\mathbf{\Sigma}}_{\mathrm{GCWE}}-\mathbf{\Sigma}\right)\end{bmatrix}.\end{aligned}$$

Therefore one has $\mathbf{\Gamma}^{(n)}=\mathbf{\Gamma}_1^{(n)}-2\mathbf{\Gamma}_2^{(n)}+\mathbf{\Gamma}_3^{(n)}$ with

$$\begin{aligned}\mathbf{\Gamma}_1^{(n)} &= \mathbb{E}\left[n\mathrm{vec}\left(\sigma\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma}\right)\mathrm{vec}\left(\sigma\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma}\right)^H\right],\\ \mathbf{\Gamma}_2^{(n)} &= \mathbb{E}\left[n\mathrm{vec}\left(\sigma\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma}\right)\mathrm{vec}\left(\widehat{\mathbf{\Sigma}}_{\mathrm{GCWE}}-\mathbf{\Sigma}\right)^H\right],\\ \mathbf{\Gamma}_3^{(n)} &= \mathbb{E}\left[n\mathrm{vec}\left(\widehat{\mathbf{\Sigma}}_{\mathrm{GCWE}}-\mathbf{\Sigma}\right)\mathrm{vec}\left(\widehat{\mathbf{\Sigma}}_{\mathrm{GCWE}}-\mathbf{\Sigma}\right)^H\right].\end{aligned}$$

One has now

$$\mathbf{\Gamma}^{(n)}\xrightarrow[n\to+\infty]{}\mathbf{\Gamma}=\mathbf{\Gamma}_1\left(\mathbf{\Sigma}\right)-2\mathbf{\Gamma}_2\left(\mathbf{\Sigma}\right)+\mathbf{\Gamma}_3\left(\mathbf{\Sigma}\right), \tag{D.1}$$

where the matrices $\mathbf{\Gamma}_1\left(\mathbf{\Sigma}\right)$ and $\mathbf{\Gamma}_3\left(\mathbf{\Sigma}\right)$ are given by Eqs. (1.32) and (1.11), respectively.

Following the similar ideas used in [119, 170], we provide a more general result that allows to compute a corelation between two estimators

$$\mathbf{\Gamma}_2^{(n)}\xrightarrow[n\to+\infty]{}\mathbf{\Gamma}_2\left(\mathbf{\Sigma}\right)=\mathbf{D}_1^{-1}\left(\mathbf{\Sigma}\right)\mathbf{B}\left(\mathbf{\Sigma}\right)\mathbf{D}_2^{-1}\left(\mathbf{\Sigma}\right),$$

where

$$\mathbf{D}_1\left(\mathbf{\Sigma}\right)=\mathbb{E}\left[d\{\mathrm{vec}\left(\Psi_1\left(\mathbf{\Sigma}\right)\right)\}/d\{\mathrm{vec}\left(\mathbf{\Sigma}\right)\}\right], \tag{D.2}$$

$$\mathbf{B}\left(\mathbf{\Sigma}\right)=\mathbb{E}\left[\mathrm{vec}\left(\Psi_1\left(\mathbf{\Sigma}\right)\right)\mathrm{vec}\left(\Psi_2\left(\mathbf{\Sigma}\right)\right)^H\right], \tag{D.3}$$

$$\mathbf{D}_2\left(\mathbf{\Sigma}\right)=\mathbb{E}\left[d\{\mathrm{vec}\left(\Psi_2\left(\mathbf{\Sigma}\right)\right)\}/d\{\mathrm{vec}\left(\mathbf{\Sigma}\right)\}\right], \tag{D.4}$$

with

$$\Psi_1\left(\mathbf{\Sigma}\right) = \sigma u\left(\mathbf{z}^H(\sigma^{-1}\mathbf{\Sigma})^{-1}\mathbf{z}\right)\mathbf{z}\mathbf{z}^H-\mathbf{\Sigma}, \tag{D.5}$$

$$\Psi_2\left(\mathbf{\Sigma}\right) = \mathbf{x}\mathbf{x}^H-\mathbf{\Sigma}. \tag{D.6}$$

W.l.o.g., we will assume that $\mathbf{\Sigma}=\mathbf{I}$. Indeed, one has that

$$\mathbf{\Gamma}_2\left(\mathbf{\Sigma}\right)=\left(\mathbf{\Sigma}^{T/2}\otimes\mathbf{\Sigma}^{1/2}\right)\mathbf{\Gamma}_2\left(\mathbf{I}\right)\left(\mathbf{\Sigma}^{T/2}\otimes\mathbf{\Sigma}^{1/2}\right)^H.$$

In order to determine the final result, we will derive the expression for $\mathbf{\Gamma}_2(\mathbf{I})$. One can show that

$$\mathbf{D}_1^{-1}(\mathbf{I}) = \alpha_1\mathbf{I} + \alpha_2 \mathrm{vec}(\mathbf{I})\,\mathrm{vec}(\mathbf{I})^T,$$

where $\alpha_1 = -\dfrac{p(p+1)}{c_M}$ and $\alpha_2 = \dfrac{p(c_M^2 - p^2 - p)}{c_M(c_M - p^2)}$ with $c_M = \mathbb{E}\left[\sigma\mathcal{Q}\Psi'(\sigma\mathcal{Q})\right] + p^2$. Moreover, it is simple to show that $\mathbf{D}_2(\mathbf{I})^{-1} = -\mathbf{I}$.

Then, basing on Theorem 2 from [170], one can derive more general result

$$\mathbf{B}_2(\mathbf{I}) = \beta_1\mathbf{I} + \beta_2 \mathrm{vec}(\mathbf{I})\mathrm{vec}(\mathbf{I})^T,$$

where

$$\begin{aligned}
\beta_1 &= \mathrm{cov}\left[\Psi_1(\mathbf{I})_{jk}\Psi_2(\mathbf{I})_{jk}\right] \\
&= \mathbb{E}\left[u(\sigma\mathcal{Q})\sigma\mathcal{Q}\|\mathbf{g}\|^2 u_j^2 u_k^2\right] \\
&= \frac{\mathbb{E}\left[\Psi(\sigma\mathcal{Q})\|\mathbf{g}\|^2\right]}{p(p+1)} = \frac{b_M}{p(p+1)}
\end{aligned}$$

and

$$\begin{aligned}
\beta_2 &= \mathrm{cov}\left[\Psi_1(\mathbf{I})_{jj}\Psi_2(\mathbf{I})_{kk}\right] \\
&= \beta_1 - \mathbb{E}\left[\Psi(\sigma\mathcal{Q})\right]\mathbb{E}\left[\|\mathbf{g}\|^2\right]/p^2 \\
&= \beta_1 - 1
\end{aligned}$$

since $u_i^2 \sim \beta(1, p-1)$, $\mathbb{E}[u_j^2] = 1/p$ and $\mathbb{E}[u_j^2 u_k^2] = 1/(p(p+1))$. After some mathematical manipulations, one obtains

$$\mathbf{\Gamma}_2(\mathbf{I}) = \gamma_1\mathbf{I} + \gamma_2 \mathrm{vec}(\mathbf{I})\mathrm{vec}(\mathbf{I})^T$$

with

$$\begin{aligned}
\gamma_1 &= -\alpha_1\beta_1 = \frac{b_M}{c_M}, \\
\gamma_2 &= -(\alpha_1\beta_2 + \alpha_2\beta_1 + p\alpha_2\beta_2) = \frac{p(b_M - c_M)}{c_M(c_M - p^2)}.
\end{aligned} \tag{D.7}$$

This leads to the final expression of $\mathbf{\Gamma}_2(\mathbf{\Sigma})$

$$\mathbf{\Gamma}_2(\mathbf{\Sigma}) = \gamma_1\mathbf{\Sigma}^T \otimes \mathbf{\Sigma} + \gamma_2 \mathrm{vec}(\mathbf{\Sigma})\mathrm{vec}(\mathbf{\Sigma})^H. \tag{D.8}$$

Combining Eq. (D.8) together with Eqs. (1.32) and (1.11) in Eq. (D.1), one obtains the coefficients σ_1 and σ_2 as follows

$$\begin{aligned}
\sigma_1 &= \vartheta_1 - 2\gamma_1 + 1 \\
&= \frac{a_M p(p+1) + c_M(c_M - 2b_M)}{c_M^2}
\end{aligned}$$

and

$$\begin{aligned}
\sigma_2 &= \vartheta_2 - 2\gamma_2 \\
&= \frac{a_M - p^2}{(c_M - p^2)^2} - \frac{a_M(p+2)}{c_M^2} + 2\frac{p(c_M - b_M)}{c_M(c_M - p^2)}.
\end{aligned}$$

Finally, one can easily prove that $\mathbf{\Omega} = \mathbf{\Gamma}\mathbf{K}$ [117], which leads to the final results and concludes the proof.

Appendix E

Proof of Theorem 2.2.2

First, notice that for the Tyler's estimator $\sigma = 1$ thanks to the employed normalization. Following the same idea as in the proof of Theorem 2.2.1 one can write

$$\mathbf{\Gamma}_{\mathrm{T}}^{(n)} = \mathbf{\Gamma}_{\mathrm{T1}}^{(n)} - 2\mathbf{\Gamma}_{\mathrm{T2}}^{(n)} + \mathbf{\Gamma}_{\mathrm{T3}}^{(n)} \tag{E.1}$$

with

$$\begin{aligned}
\mathbf{\Gamma}_{\mathrm{T1}}^{(n)} &= \mathbb{E}\left[n\mathrm{vec}\left(\widehat{\mathbf{\Sigma}}_{\mathrm{T}} - \mathbf{\Sigma}\right)\mathrm{vec}\left(\widehat{\mathbf{\Sigma}}_{\mathrm{T}} - \mathbf{\Sigma}\right)^{H}\right], \\
\mathbf{\Gamma}_{\mathrm{T2}}^{(n)} &= \mathbb{E}\left[n\mathrm{vec}\left(\widehat{\mathbf{\Sigma}}_{\mathrm{T}} - \mathbf{\Sigma}\right)\mathrm{vec}\left(\widehat{\mathbf{\Sigma}}_{\mathrm{GCWE}} - \mathbf{\Sigma}\right)^{H}\right], \\
\mathbf{\Gamma}_{\mathrm{T3}}^{(n)} &= \mathbb{E}\left[n\mathrm{vec}\left(\widehat{\mathbf{\Sigma}}_{\mathrm{GCWE}} - \mathbf{\Sigma}\right)\mathrm{vec}\left(\widehat{\mathbf{\Sigma}}_{\mathrm{GCWE}} - \mathbf{\Sigma}\right)^{H}\right],
\end{aligned} \tag{E.2}$$

where $\mathbf{\Gamma}_{\mathrm{T1}}^{(n)} \xrightarrow[n\to+\infty]{} \mathbf{\Gamma}_{\mathrm{T1}}$ equal to $\mathbf{C}_{\mathrm{T}}$ from Eq. (1.36) and $\mathbf{\Gamma}_{\mathrm{T3}}^{(n)} \xrightarrow[n\to+\infty]{} \mathbf{\Gamma}_{\mathrm{T3}}$ equal to $\mathbf{C}_{\mathrm{SCM}}$ from Eq. (1.11). Let us now introduce some notations (see [137] for details)

- $\widehat{\mathbf{\Sigma}} = \mathbf{\Sigma} + \delta\mathbf{\Sigma}$ where $\widehat{\mathbf{\Sigma}}$ denotes an estimate of $\mathbf{\Sigma}$,
- $\mathbf{\Delta} = \mathbf{\Sigma}^{-1/2}\widehat{\mathbf{\Sigma}}, \mathbf{\Sigma}^{-1/2} - \mathbf{I}$,
- $\delta = \mathrm{vec}(\mathbf{\Delta})$,
- theoretical Tyler's (TT) estimator of $\mathbf{\Sigma}$ is equal to

$$\widehat{\mathbf{\Sigma}}_{\mathrm{TT}} = \frac{p}{n}\sum_{i=1}^{n}\frac{\mathbf{z}_i\mathbf{z}_i^H}{\mathbf{z}_i^H\mathbf{\Sigma}^{-1}\mathbf{z}_i}. \tag{E.3}$$

In the sequel, these quantities will be indexed according to the studied estimator: GCWE, TT and T. Thanks to Eq. (A.1), we have

$$\begin{aligned}
\mathbf{\Gamma}_{\mathrm{T2}}^{(n)} &= \mathbb{E}\left[n\mathrm{vec}\left(\mathbf{\Sigma}^{1/2}\mathbf{\Delta}_{\mathrm{T}}\mathbf{\Sigma}^{1/2}\right)\mathrm{vec}\left(\mathbf{\Sigma}^{1/2}\mathbf{\Delta}_{\mathrm{GCWE}}\mathbf{\Sigma}^{1/2}\right)^{H}\right] \\
&= \left(\mathbf{\Sigma}^{T/2}\otimes\mathbf{\Sigma}^{1/2}\right)\mathbb{E}\left[n\delta_{\mathrm{T}}\delta_{\mathrm{GCWE}}^{H}\right]\left(\mathbf{\Sigma}^{T/2}\otimes\mathbf{\Sigma}^{1/2}\right)^{H}.
\end{aligned}$$

To derive $\mathbb{E}\left[n\delta_{\mathrm{T}}\delta_{\mathrm{GCWE}}^{H}\right]$ we can use the statement that $\sqrt{n}\delta_{\mathrm{T}}$ and $\frac{p+1}{p}\sqrt{n}\delta_{\mathrm{TT}}$ share the same asymptotic distribution for all CES distributions (see [137]). Now, it remains to derive the quantity $\mathbb{E}\left[n\delta_{\mathrm{TT}}\delta_{\mathrm{GCWE}}^{H}\right]$.

Let $\mathbf{v}_i = \mathbf{\Sigma}^{-1/2}\mathbf{z}_i$ and $\mathbf{g}_i = \mathbf{\Sigma}^{-1/2}\mathbf{x}_i$. Then, for large n, one can write

$$\begin{aligned}
\mathbb{E}\left[n\delta_{\mathrm{TT}}\delta_{\mathrm{GCWE}}^H\right] &= \mathbb{E}\left[n\mathrm{vec}\left(\frac{p}{n}\sum_{i=1}^{n}\left(\frac{\mathbf{v}_i\mathbf{v}_i^H}{\mathbf{v}_i^H\mathbf{v}_i}\right)-\mathbf{I}\right)\mathrm{vec}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{g}_i\mathbf{g}_i^H\right)-\mathbf{I}\right)^H\right] \\
&= \frac{p}{n}\mathbb{E}\left[\mathrm{vec}\left(\sum_{i=1}^{n}\left(\frac{\mathbf{v}_i\mathbf{v}_i^H}{\mathbf{v}_i^H\mathbf{v}_i}\right)\right)\mathrm{vec}\left(\sum_{i=1}^{n}\left(\mathbf{g}_i\mathbf{g}_i^H\right)\right)^H\right]-n\mathrm{vec}\left(\mathbf{I}\right)\mathrm{vec}\left(\mathbf{I}\right)^T \\
= \frac{p}{n}\left(\sum_{i=1}^{n}\mathbb{E}\left[\mathrm{vec}\left(\frac{\mathbf{v}_i\mathbf{v}_i^H}{\mathbf{v}_i^H\mathbf{v}_i}\right)\mathrm{vec}\left(\mathbf{g}_i\mathbf{g}_i^H\right)^H\right]+\sum_{j\neq k}\mathbb{E}\left[\mathrm{vec}\left(\frac{\mathbf{v}_k\mathbf{v}_k^H}{\mathbf{v}_k^H\mathbf{v}_k}\right)\mathrm{vec}\left(\mathbf{g}_j\mathbf{g}_j^H\right)^H\right]\right) & \\
-n\mathrm{vec}(\mathbf{I})\mathrm{vec}(\mathbf{I})^T & \\
= p\mathbb{E}\left[\mathrm{vec}\left(\frac{\mathbf{v}\mathbf{v}^H}{\mathbf{v}^H\mathbf{v}}\right)\mathrm{vec}\left(\mathbf{g}\mathbf{g}^H\right)^H\right]+(n-1)\mathbb{E}\left[\mathrm{vec}\left(p\frac{\mathbf{v}\mathbf{v}^H}{\mathbf{v}^H\mathbf{v}}\right)\right]\mathbb{E}\left[\mathrm{vec}\left(\mathbf{g}\mathbf{g}^H\right)^H\right] & \\
-n\mathrm{vec}(\mathbf{I})\mathrm{vec}(\mathbf{I})^T & \\
= p\mathbb{E}\left[\mathrm{vec}\left(\frac{\mathbf{v}\mathbf{v}^H}{\mathbf{v}^H\mathbf{v}}\right)\mathrm{vec}\left(\mathbf{g}\mathbf{g}^H\right)^H\right]-\mathrm{vec}(\mathbf{I})\mathrm{vec}(\mathbf{I})^T, &
\end{aligned}$$

where $\mathbf{v}_i = \sqrt{\mathcal{Q}_i}\mathbf{u}_i \sim \mathcal{CES}(\mathbf{0},\mathbf{I},g)$ and $\mathbf{g}_i = \|\mathbf{g}_i\|\mathbf{u}_i \sim \mathcal{CN}(\mathbf{0},\mathbf{I})$ with $\mathcal{Q}_i$ being an r.va. whose p.d.f. is unkown, $\|\mathbf{g}\|^2 \sim (1/2)\chi^2_{2p}$ and $\mathbf{u}_i \sim \mathcal{U}(\mathcal{CS}^{p-1})$. Then, focusing on the following variable

$$\mathbf{P} = \mathbb{E}\left[\mathrm{vec}\left(\frac{\mathbf{v}\mathbf{v}^H}{\mathbf{v}^H\mathbf{v}}\right)\mathrm{vec}\left(\mathbf{g}\mathbf{g}^H\right)^H\right], \tag{E.4}$$

each element of matrix $\mathbf{P}$ becomes

$$P_{kl} = \mathbb{E}\left[u_q u_r^* u_{q'}^* u_{r'}\|\mathbf{g}\|^2\right] \tag{E.5}$$

with $k = q + p(r-1)$ and $l = q' + p(r'-1)$.
Now, let us define the 4th-order moments of a complex random vector $\mathbf{u}$ by

$$\alpha_{i_1,i_2;j_1,j_2} = \mathbb{E}\left[u_{i_1}u_{i_2}u_{j_1}^*u_{j_1}^*\right]. \tag{E.6}$$

By the circular symmetry properties all even-order central moments vanish in the cases that the sets (i_1,i_2) and (j_1,j_2) differ. In our case, that means that $P_{kl} = 0$ except for the following indices:

- $k = q + p(q-1)$,
- $k = q + p(q-1)$, $l = q' + p(q'-1)$ and $q \neq q'$,
- $k = q + p(q'-1)$, $l = q + p(q'-1)$ and $q \neq q'$.

Now, since $\mathbf{u} \sim \mathcal{U}(\mathcal{CS}^{p-1})$ one has $|u_i|^2 \sim \beta(1,p-1)$, $\mathbb{E}\left[|u_i|^4\right] = 2/p(p+1)$ and $\mathbb{E}\left[|u_i|^2|u_j|^2\right] = 1/p(p+1)$. Also, as $\|\mathbf{g}\|^2 \sim (1/2)\chi^2_{2p}$ one obtains the following results:

- $P_{q+p(q-1),q+p(q-1)} = 2/(p+1)$,
- $P_{q+p(q-1),q'+p(q'-1)} = 1/(p+1)$,
- $P_{q+p(q'-1),q+p(q'-1)} = 1/(p+1)$,

and thus

$$\mathbb{E}\left[n\delta_{\mathrm{TT}}\delta_{\mathrm{GCWE}}^H\right] \xrightarrow[n\to+\infty]{} \mathbf{B}, \tag{E.7}$$

where

$$\mathbf{B} = \frac{p}{p+1}\left(\mathbf{I} - \frac{1}{p}\mathrm{vec}(\mathbf{I})\mathrm{vec}(\mathbf{I})^T\right). \tag{E.8}$$

Therefore

$$\mathbb{E}\left[n\delta_{\mathrm{T}}\delta_{\mathrm{GCWE}}^H\right] \xrightarrow[n\to+\infty]{} \frac{p+1}{p}\mathbf{B}. \tag{E.9}$$

Since the covariance matrix is Hermitian and using the property Eq. (A.2), one has $\mathbf{\Gamma}_{\mathrm{T2}}^{(n)} \xrightarrow[n\to+\infty]{} \mathbf{\Gamma}_{\mathrm{T2}}$ where

$$\begin{aligned}\mathbf{\Gamma}_{\mathrm{T2}} &= \left(\mathbf{\Sigma}^{T/2}\otimes\mathbf{\Sigma}^{1/2}\right)\left(\mathbf{\Sigma}^{T/2}\otimes\mathbf{\Sigma}^{1/2}\right)^{H} - \frac{1}{p}\left(\mathbf{\Sigma}^{T/2}\otimes\mathbf{\Sigma}^{1/2}\right)\mathrm{vec}(\mathbf{I})\mathrm{vec}(\mathbf{I})^{T}\left(\mathbf{\Sigma}^{T/2}\otimes\mathbf{\Sigma}^{1/2}\right)^{H} \\ &= \left(\mathbf{\Sigma}^{T}\otimes\mathbf{\Sigma}\right) - \frac{1}{p}\mathrm{vec}(\mathbf{\Sigma})\mathrm{vec}(\mathbf{\Sigma})^{H}. \end{aligned} \tag{E.10}$$

Finally, we obtain the expression of $\mathbf{\Gamma}_{\mathrm{T}}$

$$\mathbf{\Gamma}_{\mathrm{T}} = \mathbf{\Gamma}_{\mathrm{T1}} - 2\mathbf{\Gamma}_{\mathrm{T2}} + \mathbf{\Gamma}_{\mathrm{T3}} = \frac{1}{p}\left(\mathbf{\Sigma}^{T}\otimes\mathbf{\Sigma}\right) + \frac{p-1}{p^{2}}\mathrm{vec}(\mathbf{\Sigma})\mathrm{vec}(\mathbf{\Sigma})^{H}. \tag{E.11}$$

The asymptotic pseudo-covariance matrix $\mathbf{\Omega}$ is defined as

$$\mathbf{\Omega} = \mathbb{E}\left[n\mathrm{vec}\left(\widehat{\mathbf{\Sigma}}_{\mathrm{T}} - \widehat{\mathbf{\Sigma}}_{\mathrm{GCWE}}\right)\mathrm{vec}\left(\widehat{\mathbf{\Sigma}}_{\mathrm{T}} - \widehat{\mathbf{\Sigma}}_{\mathrm{GCWE}}\right)^{T}\right]. \tag{E.12}$$

As derived in [117], $\mathbf{\Omega}_{\mathrm{T}} = \mathbf{\Gamma}_{\mathrm{T}}\mathbf{K}$ which thanks to Eq. (A.3) leads to the result of Theorem 2.2.1 and concludes the proof.

Appendix F

Synthèse

F.1 Introduction

Les statistiques de second ordre jouent un rôle de première importance dans les applications de traitement du signal. Leurs estimations ont récemment suscité un intérêt croissant dans de nombreux communautés [130, 135, 27, 182, 117, 162, 129]. Sous l'hypothèse gaussienne, l'estimateur classique est la matrice de covariance empirique (SCM), distribuée selon une loi de Wishart [13] (à distance finie, i.e. pour un nombre d'échantillons n et une dimension p des observations fixés). En tant qu'estimateur du maximum de vraisemblance (MLE) dans ce contexte, la SCM possède de bonnes propriétés statistiques. Cependant, lorsque la distribution des données n'est plus gaussienne, ou quand les données sont corrompues par des données aberrantes (*outliers*), les performances de la SCM peuvent se dégrader fortement.

La théorie de l'estimation robuste offre dans ce cas une alternative grâce aux M-estimateurs, étudiés dans le cas réel dans [119]. Ces résultats ont été récemment étendus au cas complexe, plus approprié aux applications de traitement du signal, dans [117]. La plupart des travaux sur l'estimation robuste de la matrice de covariance ont été menés dans le cadre de distributions symétriques elliptiques complexes (CES) (voir e.g., [130]). Ces M-estimateurs donnent de très bons résultats quand ils sont utilisés à la place de la SCM. Malheureusement, leur définition est implicite, ce qui rend leur analyse statistique difficile. Cette thèse vise donc à caractériser le comportement des M-estimateurs plus finement que par une analyse asymptotique classique [117, 130]. Les résultats obtenus révèlent que les propriétés statistiques des M-estimateurs peuvent être bien approximées par une distribution de Wishart.

Grâce à ces résultats, nous analysons la décomposition de la matrice de covariance en éléments propres. Selon l'application, la matrice de covariance peut posséder une structure particulière impliquant valeurs propres multiples contenant les informations d'intérêt. Nous abordons ainsi divers scénarios rencontrés dans la pratique et proposons des procédures robustes basées sur des M-estimateurs.

Nous montrerons aussi l'intérêt des résultats proposés pour le traitement de données radar. Nous analysons deux problèmes souvent rencontrés dans le traitement radar : détection adaptative et traitement des images radar.

La détection adaptative de signaux corrompus par un bruit additif est un problème omniprésent en traitement statistique du signal. Ce problème a fait l'objet de nombreuses études dans le contexte des perturbations gaussiennes. Plusieurs statistiques de décision ont été proposées, telles que le test du rapport de vraisemblance généralisé (GLRT) (détecteur de Kelly) [90], le filtre adapté adaptatif (AMF ou GLRT à deux niveaux) [150], son équivalent normalisé (filtre adapté normalisé adaptatif, ANMF ou *Adaptive Cosine Estimator*, ACE) [39, 97] et le test de Rao [118]. Les détecteurs associés ont été caractérisés en termes de PD et de PFA, propriété de taux de fausse alarme constant (CFAR) (voir, par exemple, [132] et

les références contenues dans cet article) et les performances dans des scénarios *mismatch* [68]. Pour détecter un signal, ces détecteurs ont besoin de connaître les caractéristiques de second ordre des données du bruit seul. Lorsque la matrice de covariance des données est inconnue, la SCM la remplace généralement dans la formulation des détecteurs adaptatifs. Dans le cas non-gaussien, un détecteur robuste peut être construit comme un détecteur classique en remplaçant la SCM par un M-estimateur. Dans ce contexte, les résultats proposés pour caractériser les M-estimateurs permettront d'analyser finement le comportement de ces détecteurs robustes.

Les images de radar à synthèse d'ouverture polarimétrique (PolSAR) sont très utilisés pour la classification du terrain, la détection de cible, etc. Dans les images PolSAR, chaque pixel correspond à un vecteur complexe, formée par la réponse des signaux rétrodiffusés dans différentes combinaisons de polarisations linéaires reçues et transmises, appelé vecteur de cible. Étant donné la position inconnue des diffuseurs et en raison de la cohérence des systèmes PolSAR, ce vecteur peut être modélisé comme aléatoire selon le modèle de Goodman [72]. Ce phénomène, appelé *speckle*, dégrade considérablement la qualité de l'image ainsi que les performances de nombreux traitements. Par conséquent, afin de déterminer les paramètres physiques d'intérêt, une étape de filtrage du speckle est généralement appliquée, dans le but de réduire les fluctuations dues à ce bruit. Le filtrage de speckle consiste à estimer la matrice de covariance du vecteur de cible dans chaque pixel de l'image. Dans cette thèse, nous analysons l'intérêt d'utiliser les techniques robustes, particulièrement les M-estimateurs, pour des problématiques de filtrage d'images PolSAR.

F.2 Etat de l'art

Cette section présente le contexte général de la thèse. Elle est divisée en trois parties. La première partie porte sur l'estimation de la matrice de covariance, commençant par l'approche gaussienne classique, suivi du cadre elliptique et des méthodes robustes. La deuxième partie traite la détection de signaux. Le problème de la détection d'un signal complexe dans un environnement homogène et les détecteurs associés sont présentés. Enfin, la dernière partie traite des images PolSAR.

F.2.1 Estimation de la matrice de covariance

Dans cette partie, nous définissons les termes nécessaires à notre étude sur l'estimation robuste.

Definition F.2.1. *Matrice de covariance*
La matrice de covariance $\mathbf{C} \in \mathcal{H}$ du vecteur complexe $\mathbf{z} = \mathbf{a} + j\mathbf{b}$ est définie comme

$$\mathbf{C} = \mathbb{E}\left[\mathbf{z}\mathbf{z}^H\right] = \mathbb{E}\left[\mathbf{a}\mathbf{a}^T\right] + \mathbb{E}\left[\mathbf{b}\mathbf{b}^T\right] + j\left(\mathbb{E}\left[\mathbf{b}\mathbf{a}^T\right] - \mathbb{E}\left[\mathbf{a}\mathbf{b}^T\right]\right). \tag{F.1}$$

Definition F.2.2. *Matrice de pseudo-covariance*
La matrice de pseudo-covariance $\mathbf{P} \in \mathcal{CS}$ du vecteur complexe $\mathbf{z} = \mathbf{a} + j\mathbf{b}$ est définie comme

$$\mathbf{P} = \mathbb{E}\left[\mathbf{z}\mathbf{z}^T\right] = \mathbb{E}\left[\mathbf{a}\mathbf{a}^T\right] - \mathbb{E}\left[\mathbf{b}\mathbf{b}^T\right] + j\left(\mathbb{E}\left[\mathbf{b}\mathbf{a}^T\right] + \mathbb{E}\left[\mathbf{a}\mathbf{b}^T\right]\right). \tag{F.2}$$

Dans le traitement de signal les signaux sont très souvent considérés comme circulaires du second ordre.

Definition F.2.3. *Circularité du second-ordre*
Le vecteur complexe $\mathbf{z}$ est dit circulaire du second-ordre lorsque $\mathbf{P} = 0$.

Hypothèse gaussienne

Traditionnellement, la plupart des applications de traitement du signal considèrent les données comme étant gaussiennes. Dans ce cas, l'estimateur de maximum vraisemblance est

la matrice de covariance empirique (SCM). Soit $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ un n-échantillon de vecteurs complexes indépendants de taille p avec $\mathbf{x}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$. Alors, la SCM donnée par

$$\widehat{\mathbf{C}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^H \tag{F.3}$$

suit au facteur $1/n$ près une distribution de Wishart complexe, notée $\mathcal{CW}_p(n, \mathbf{C})$. Sa distribution asymptotique est donnée par

$$\sqrt{n}\mathrm{vec}\left(\widehat{\mathbf{C}} - \mathbf{C}\right) \xrightarrow{d} \mathcal{GCN}\left(\mathbf{0}, \mathbf{C}^T \otimes \mathbf{C}, \left(\mathbf{C}^T \otimes \mathbf{C}\right)\mathbf{K}\right),$$

où $\mathcal{GCN}$ dénote la loi normale généralisée.

Distributions CES

Une généralisation naturelle de la distribution gaussienne est donnée par la classe des distributions elliptiques symétriques complexes (CES), qui permettent de modéliser des queues plus lourdes et plus légères que la distribution gaussienne.

Definition F.2.4. *Distribution symétrique elliptique complexe (CES)*
Un vecteur $\mathbf{z}$ aléatoire complexe circulaire de taille p suit une distribution CES, notée $\mathcal{CES}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g_{\mathbf{z}})$, si sa densité de probabilité (p.d.f.) peut s'écrire

$$f_{\mathbf{z}}(\mathbf{z}) = C_{p,g} |\boldsymbol{\Sigma}|^{-1} g_{\mathbf{z}}\left((\mathbf{z} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \boldsymbol{\mu})\right) \tag{F.4}$$

où $C_{p,g}$ est une constante, $g_{\mathbf{z}} : [0, \infty) \to [0, \infty)$ est une fonction telle que l'Eq. (F.4) définit une p.d.f., $\boldsymbol{\mu}$ est l'espérance de $\mathbf{z}$ et $\boldsymbol{\Sigma}$ sa matrice de dispersion. $\boldsymbol{\Sigma}$ contient la structure de la matrice de covariance de $\mathbf{z}$, i.e., cette matrice de covariance (si elle existe) est proportionnelle à $\boldsymbol{\Sigma}$. Dans ce travail, nous supposons $\boldsymbol{\mu} = \mathbf{0}$ (hypothèse commune en traitement statistique du signal).

Théorème F.2.1. *Représentation stochastique des CES [187]*
Un vecteur aléatoire $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \boldsymbol{\Sigma}, g_{\mathbf{z}})$ peut être représenté par

$$\mathbf{z} \overset{d}{=} \sqrt{\mathcal{Q}}\mathbf{A}\mathbf{u} \tag{F.5}$$

où $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^H$ est une factorisation de $\boldsymbol{\Sigma}$ et $\mathbf{u} \sim \mathcal{U}(\mathcal{CS}^{p-1})$ (distribution uniforme sur la sphère unité). $\mathcal{Q}$ est une variable aléatoire réelle non négative, indépendante de $\mathbf{u}$ avec une p.d.f. dépendant uniquement de $g_{\mathbf{z}}$.

Une sous-classe très importante de distributions CES dans les applications de traitement du signal sont les distributions gaussienne-composée complexes, également appelés *Spherically Invariant Random Vectors* (SIRV). Ces distributions sont largement utilisées pour modélisation du fouillis radar.

Definition F.2.5. *Distribution gaussienne-composée complexe*
Un vecteur aléatoire $\mathbf{z}$ suit une distribution gaussienne-composée complexe (de moyenne nulle) s'il peut être représenté par

$$\mathbf{z} \overset{d}{=} \sqrt{\tau}\mathbf{g} \tag{F.6}$$

où τ est une variable aléatoire réelle non négative appelée la texture et $\mathbf{g} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$ est le speckle.

Exemples des distributions CES

- **La loi gaussienne $\mathcal{CN}(\mathbf{0}, \mathbf{\Sigma})$:** Elle s'obtient avec $g(x) = \exp(-x)$ et $C_{p,g} = \pi^{-p}$.
- **La t-distribution $\mathcal{C}t_\nu(\mathbf{0}, \mathbf{\Sigma})$:** La t-distribution avec ν degrés de liberté s'obtient avec $g(x) = (1 + 2x/\nu)^{-(2p+\nu)/2}$ et $C_{p,g} = 2^p\Gamma(\frac{2p+\nu}{2})/[(\pi\nu)^p\Gamma(\frac{\nu}{2})]$.
- **La K-distribution $\mathcal{CK}_\nu(\mathbf{0}, \mathbf{\Sigma})$:** La K-distribution avec le paramètre de forme ν s'obtient avec $g(x) = x^{(\nu-p)/2}K_{\nu-p}(2\sqrt{\nu x})$ et $C_{p,g} = 2\nu^{(\nu+p)/2}/(\pi^p\Gamma(p))$.

Definition F.2.6. *M-estimateur*
Soit $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ un n-échantillon de vecteurs complexes indépendants de taille p avec $\mathbf{z}_i \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g_\mathbf{z})$. Un M-estimateur de $\mathbf{\Sigma}$, noté $\widehat{\mathbf{\Sigma}}$, est défini comme la solution de l'équation suivante

$$\widehat{\mathbf{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n} u(\mathbf{z}_i^H\widehat{\mathbf{\Sigma}}^{-1}\mathbf{z}_i)\mathbf{z}_i\mathbf{z}_i^H, \tag{F.7}$$

où u est une fonction réelle de pondération définie sur $[0, \infty)$ qui n'est pas nécessairement reliée à la p.d.f. de la distribution CES. La matrice de dispersion théorique de l'échantillon est définie comme la solution de

$$\mathbb{E}\left[u\left(\mathbf{z}^H\mathbf{\Sigma}_\sigma^{-1}\mathbf{z}\right)\mathbf{z}\mathbf{z}^H\right] = \mathbf{\Sigma}_\sigma = \sigma^{-1}\mathbf{\Sigma}, \tag{F.8}$$

où $\sigma > 0$ est solution de

$$\mathbb{E}\left[\Psi(\sigma\mathcal{Q})\right] = p \tag{F.9}$$

avec $\Psi(\sigma\mathcal{Q}) = u(\sigma\mathcal{Q})\sigma\mathcal{Q}$ et $\mathcal{Q} \overset{d}{=} \mathbf{z}^H\mathbf{\Sigma}^{-1}\mathbf{z}$.

Dans le cas des MLEs, la fonction de pondération dépend de la fonction génératrice $g(x)$ de densité comme $u(x) = -g'(x)/g(x)$.

Dans [117], les propriétés asymptotiques suivantes des M-estimateurs ont été prouvées.

Théorème F.2.2. *Propriétés asymptotiques*
Si $\widehat{\mathbf{\Sigma}}$ est un M-estimateur respectant les conditions de [119], alors

$$\sqrt{n}\mathrm{vec}\left(\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}_\sigma\right) \overset{d}{\to} \mathcal{GCN}\left(\mathbf{0}, \mathbf{C}, \mathbf{P}\right),$$

où les matrices de covariance et de pseudo-covariance asymptotiques sont

$$\begin{cases} \mathbf{C}_M = \vartheta_1\mathbf{\Sigma}_\sigma^T \otimes \mathbf{\Sigma}_\sigma + \vartheta_2\mathrm{vec}\left(\mathbf{\Sigma}_\sigma\right)\mathrm{vec}\left(\mathbf{\Sigma}_\sigma\right)^H, \\ \mathbf{P}_M = \vartheta_1\left(\mathbf{\Sigma}_\sigma^T \otimes \mathbf{\Sigma}_\sigma\right)\mathbf{K} + \vartheta_2\mathrm{vec}\left(\mathbf{\Sigma}_\sigma\right)\mathrm{vec}\left(\mathbf{\Sigma}_\sigma\right)^T. \end{cases} \tag{F.10}$$

Les constantes $\vartheta_1 > 0$ et $\vartheta_2 > -\vartheta_1/p$ sont données par [117]

$$\begin{aligned} \vartheta_1 &= c_M^{-2}a_M p(p+1), \\ \vartheta_2 &= (c_M - p^2)^{-2}(a_M - p^2) - c_M^{-2}a_M(p+1), \end{aligned} \tag{F.11}$$

où

$$\begin{aligned} a_M &= \mathbb{E}\left[\Psi^2(\sigma\mathcal{Q})\right], \\ c_M &= \mathbb{E}\left[\Psi'(\sigma\mathcal{Q})\sigma\mathcal{Q}\right] + p^2. \end{aligned} \tag{F.12}$$

Le tableau F.1 récapitule les fonctions de pondération $u(x)$ et les fonctions $\Psi(x)$ pour quelques exemples des M-estimateurs. Les paramètres λ et β du M-estimateur d'Huber sont réglables. Ils permettent de choisir quel sera le pourcentage de données atténuées (forme quadratique supérieure à λ) ainsi que le coefficient de proportionnalité entre la limite de l'estimateur et la matrice de dispersion de la distribution elliptique considérée (σ de l'Eq. (F.9)). Le M-estimateur de Student, correspond au M-estimater de Tyler pour $\nu = 0$ et à la SCM pour $\nu \to \infty$.

M-estimateur	$u(x)$	$\Psi(x)$
M-estimateur de Tyler	$\frac{p}{x}$	p
M-estimateur d'Huber	$\frac{1}{\beta}\min\left(1, \frac{\lambda}{x}\right)$	$\frac{1}{\beta}\min\left(x, \lambda\right)$
M-estimateur de Student	$\frac{2p+\nu}{2x+\nu}$	$\frac{2p+\nu}{2x+\nu}x$

TABLE F.1 – Exemples des M-estimateurs

F.2.2 Détection du signal

Dans cette partie, nous rappelons les problèmes de détection de « rang plein » et « rang faible » dans les perturbations gaussiennes homogènes ainsi que plusieurs détecteurs adaptés à ces situations.

Détection « rang plein »

On considère le problème consistant à détecter un signal complexe $\mathbf{p}$ dans les données reçues sous la forme $\mathbf{z} = \alpha\mathbf{p} + \mathbf{c}$, où $\mathbf{c}$ est un bruit complexe (fouillis), $\alpha \in \mathbb{C}$ une amplitude complexe inconnue et $\mathbf{p}$ un vecteur directionnel (*steering vector*) connu. Ce problème peut se formaliser à l'aide d'un test d'hypothèses binaires

$$\begin{cases} H_0 : \mathbf{z} = \mathbf{c} & \mathbf{z}_i = \mathbf{c}_i, \quad i = 1, \ldots, n, \\ H_1 : \mathbf{z} = \alpha\mathbf{p} + \mathbf{c} & \mathbf{z}_i = \mathbf{c}_i, \quad i = 1, \ldots, n, \end{cases} \tag{F.13}$$

où $\mathbf{c}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma})$ sont n observations indépendantes ne contenant pas de signal et utilisées pour estimer la matrice de covariance du bruit.

Nous nous intéressons à plusieurs détecteurs définis ci-dessous.

- En supposant que les données primaires et secondaires sont gaussiennes avec une matrice de covariance $\mathbf{C}$ inconnue, Kelly [90] a proposé le GLRT

$$\Lambda_{\text{Kelly}}\left(\widehat{\mathbf{C}}\right) = \frac{\left|\mathbf{p}^H\widehat{\mathbf{C}}^{-1}\mathbf{z}\right|^2}{\left(\mathbf{p}^H\widehat{\mathbf{C}}^{-1}\mathbf{p}\right)\left(n + \mathbf{z}^H\widehat{\mathbf{C}}^{-1}\mathbf{z}\right)} \underset{H_1}{\overset{H_0}{\lessgtr}} \lambda_{\text{Kelly}}, \tag{F.14}$$

  où $\widehat{\mathbf{C}}$ est donné par l'Eq. (F.3).

- Dans [150], Robey a proposé un autre GLRT dans la même configuration, mais en considérant que $\mathbf{C}$ est connu. La version adaptative du détecteur est alors obtenue

$$\Lambda_{\text{AMF}}\left(\widehat{\mathbf{C}}\right) = \frac{\left|\mathbf{p}^H\widehat{\mathbf{C}}^{-1}\mathbf{z}\right|^2}{n\left(\mathbf{p}^H\widehat{\mathbf{C}}^{-1}\mathbf{p}\right)} \underset{H_1}{\overset{H_0}{\lessgtr}} \lambda_{\text{AMF}}. \tag{F.15}$$

- Le détecteur adapté normalisé adaptatif [39, 97], a été dérivé pour le bruit gaussien partiellement homogène où $\mathbf{C}$ est différent entre les données primaires et secondaires, $\mathbf{c} \sim \mathcal{CN}(\mathbf{0}, \alpha\mathbf{C})$ et $\mathbf{c}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$

$$\Lambda_{\text{ANMF}}\left(\widehat{\mathbf{C}}\right) = \frac{\left|\mathbf{p}^H\widehat{\mathbf{C}}^{-1}\mathbf{z}\right|^2}{\left(\mathbf{p}^H\widehat{\mathbf{C}}^{-1}\mathbf{p}\right)\left(\mathbf{z}^H\widehat{\mathbf{C}}^{-1}\mathbf{z}\right)} \underset{H_1}{\overset{H_0}{\lessgtr}} \lambda_{\text{ANMF}}. \tag{F.16}$$

- En 2007, De Maio a proposé une nouvelle statistique de détection basée sur le test de Rao [118]

$$\Lambda_{\text{Rao}}\left(\widehat{\mathbf{C}}\right) = \frac{\left|\mathbf{p}^H\widehat{\mathbf{C}}^{-1}\mathbf{z}\right|^2 / \left(\mathbf{p}^H\widehat{\mathbf{C}}^{-1}\mathbf{p}\right)}{\left(n+\mathbf{z}^H\widehat{\mathbf{C}}^{-1}\mathbf{z}\right)\left[1+\frac{1}{n}\mathbf{z}^H\widehat{\mathbf{C}}^{-1}\mathbf{z}-\frac{1}{n}\frac{\left|\mathbf{p}^H\widehat{\mathbf{C}}^{-1}\mathbf{z}\right|^2}{\mathbf{p}^H\widehat{\mathbf{C}}^{-1}\mathbf{p}}\right]} \underset{H_1}{\overset{H_0}{\lessgtr}} \lambda_{\text{Rao}}. \tag{F.17}$$

Toutes ces statistiques assurent la propriété CFAR par rapport à la matrice de covariance et peuvent être caractérisés par la probabilité de détection et la probabilité de fausse alarme.

Détection « rang faible »

Dans le cas dit « rang faible » (LR), la perturbation est composée du fouillis gaussien de rang faible et du bruit additif blanc gaussien (AWGN)

$$\begin{cases} H_0 : \mathbf{z} = \mathbf{c} + \mathbf{n} & \mathbf{z}_i = \mathbf{c}_i + \mathbf{n}_i, \quad i = 1, \ldots, n \\ H_1 : \mathbf{z} = \alpha\mathbf{p} + \mathbf{c} + \mathbf{n} & \mathbf{z}_i = \mathbf{c}_i + \mathbf{n}_i, \quad i = 1, \ldots, n \end{cases} \tag{F.18}$$

où $\mathbf{c} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_r)$ est le fouillis gaussien du rang faible, i.e. rang $(\mathbf{\Sigma}_r) = r \ll p$ et $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \gamma^2\mathbf{I}_p)$. Par conséquent, la matrice de covariance des données secondaires peut être écrite sous la forme $\mathbf{\Sigma} = \mathbf{\Sigma}_r + \gamma^2\mathbf{I}_p$.

La décomposition en éléments propres (EVD) de la matrice $\mathbf{\Sigma}_r$ est définie comme

$$\mathbf{\Sigma}_r = \mathbf{U}_r\mathbf{\Lambda}_r\mathbf{U}_r^H \quad \text{avec} \quad \begin{array}{l} \mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_r] \in \mathcal{U}_r^p, \\ \mathbf{\Lambda}_r = \text{diag}(\boldsymbol{\lambda}_r), \\ \boldsymbol{\lambda}_r = [\lambda_1, \ldots, \lambda_r]\,. \end{array} \tag{F.19}$$

Le projecteur $\mathbf{\Pi}_r$ sur le sous-espace de fouillis et le projecteur $\mathbf{\Pi}_r^\perp$ orthogonal sur le fouillis sont définis comme

$$\begin{cases} \mathbf{\Pi}_r = \mathbf{U}_r\mathbf{U}_r^H \\ \mathbf{\Pi}_r^\perp = \mathbf{I} - \mathbf{\Pi}_r = \mathbf{U}_r^\perp\left(\mathbf{U}_r^\perp\right)^H \end{cases} \quad \text{avec} \quad \mathbf{U}_r^\perp = [\mathbf{u}_{r+1}, \ldots, \mathbf{u}_p]\,. \tag{F.20}$$

Afin de supprimer le fouillis, nous pouvons effectuer un blanchiment approximatif $\mathbf{\Sigma}^{-1} \sim \mathbf{\Pi}^\perp$, les traitements rang faible exploitent cette propriété en remplaçant $\widehat{\mathbf{C}}$ par $\widehat{\mathbf{\Pi}}^\perp$. Par exemple, le LR-ANMF prend la forme

$$\Lambda_{\text{LR}}\left(\widehat{\mathbf{\Pi}}_r^\perp\right) = \frac{\left|\mathbf{p}^H\widehat{\mathbf{\Pi}}_r^\perp\mathbf{z}\right|^2}{\left(\mathbf{p}^H\widehat{\mathbf{\Pi}}_r^\perp\mathbf{p}\right)\left(\mathbf{z}^H\widehat{\mathbf{\Pi}}_r^\perp\mathbf{z}\right)}, \tag{F.21}$$

où $\widehat{\mathbf{\Pi}}_r^\perp$ est un estimateur de $\mathbf{\Pi}_r^\perp$.

F.2.3 Images PolSAR

Dans cette partie, nous expliquons les principes de l'imagerie radar à synthèse d'ouverture. Les bases de polarimétrie SAR sont aussi récapitulées.

Images radar à synthèse d'ouverture (SAR)

Le radar est un système généralement utilisé pour détecter des objets et mesurer leur vitesse. Une autre application est l'observation et l'analyse de la surface du sol par imagerie SAR. Le grand avantage des systèmes SAR par rapport aux systèmes optiques est qu'ils fournissent des images haute résolution qui ne dépendent pas de la lumière du jour, ni des

conditions météorologiques. Comme un radar conventionnel, un capteur SAR transmet des signaux électromagnétiques radar et collecte les échos rétrodiffusés.

Lorsque le signal atteint le sol, il est diffusé dans une ou plusieurs directions en fonction des propriétés de la surface. Le signal résultant dans une cellule de résolution est obtenu en tant que somme cohérente de tous les signaux rétrodiffusés renvoyés à la cellule. Les fortes variations entre les amplitudes et les phases qui en résultent d'un pixel à l'autre provoquent un effet particulier observé dans les images SAR appelé speckle. Le speckle a un caractère multiplicatif et sa variance augmente avec l'intensité du signal. Une méthode largement utilisée pour la réduction du speckle est le filtrage *multi-look*, qui correspond au moyennage non cohérent de l'image d'intensité.

Polarimétrie

Les ondes SAR transmises sont polarisées et certains matériaux peuvent refléter des polarisations différentes avec des intensités différentes. Certains matériaux peuvent également convertir une polarisation en une autre. Les systèmes SAR peuvent alors transmettre un mélange de polarisations et utiliser des antennes de réception avec une polarisation spécifique, afin de collecter plusieurs signaux d'une même série d'impulsions.

Le principe de la polarimétrie SAR est expliqué par la matrice de diffusion, aussi appelée matrice de Sinclair, qui décrit la transformation du vecteur d'onde transmise en vecteur d'onde reçue effectuée par le diffuseur

$$\mathbf{S} = \begin{bmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{bmatrix}.$$

Les éléments de $\mathbf{S}$ sont les quatre amplitudes de diffusion complexes, où les indices horizontaux (H) ou verticaux (V) indiquent les polarisations associée reçues et transmises. Dans les images PolSAR, la phase absolue est dans la plupart des cas négligée et seules les phases relatives entre les éléments sont prises en compte. Dans la configuration dite monostatique, où l'antenne d'émission sert aussi d'antenne de réception, la matrice de Sinclair est symétrique, i.e. $S_{HV} = S_{VH}$. On peut alors réécrire la matrice de Sinclair sous forme vectorielle en la projetant sur une base orthogonale. Le vecteur complexe de taille 3×1, résultant de la projection lexicographique est donné par

$$\mathbf{k}_L = \begin{bmatrix} S_{HH} & \sqrt{2}S_{HV} & S_{VV} \end{bmatrix}^T$$

et il peut être alternativement remplacé par un vecteur obtenu avec la base de Pauli

$$\mathbf{k}_P = \frac{1}{\sqrt{2}} \begin{bmatrix} S_{HH} + S_{VV} & S_{HH} - S_{VV} & 2S_{HV} \end{bmatrix}^T. \tag{F.22}$$

Les moments du second ordre sont donnés par les matrices de covariance $\mathbf{C_s}$ et de cohérence $\mathbf{C_k}$ polarimétriques de la manière suivante :

$$\mathbf{C}_L = \mathbb{E}\left[\mathbf{k}_L \mathbf{k}_L^H\right] \tag{F.23}$$

et

$$\mathbf{C}_P = \mathbb{E}\left[\mathbf{k}_P \mathbf{k}_P^H\right]. \tag{F.24}$$

La décomposition de Cloude-Pottier [30] est basée sur l'idée qu'au sein de chaque pixel, il existe un mécanisme polarimétrique associé à chacun des vecteurs propres et dont l'intensité dépend de la valeur propre associée. On peut alors définir trois paramètres de $\mathbf{C}_L = \sum_{i=1}^{3} \lambda_i \mathbf{u}_i \mathbf{u}_i^H$, qui dépendent des valeurs propres et des vecteurs propres.

- L'entropie $H \in [0, 1]$ qui représente le degré de chaos à l'intérieur de la cellule de résolution

$$H = -\sum_{i=1}^{3} p_i \log_3 p_i \quad \text{où} \quad p_i = \frac{\lambda_i}{\sum_{i=1}^{3} \lambda_i}, \tag{F.25}$$

- L'anisotropie $A \in [0, 1]$ qui représente des mécanismes de rétrodiffusion secondaires

$$A = \frac{\lambda_2 - \lambda_3}{\lambda_2 + \lambda_3}, \tag{F.26}$$

- L'angle $\overline{\alpha} \in (-\pi, \pi]$ qui représente le type de mécanisme dominant

$$\overline{\alpha} = \sum_{i=1}^{3} p_i \alpha_i. \tag{F.27}$$

Modélisation des images PolSAR

Un modèle statistique précis est essentiel pour interpréter les informations obtenues sur une zone observée. Le modèle gaussien a été fréquemment utilisé lorsque la résolution spatiale des images PolSAR était modérée. Aujourd'hui, la plupart des systèmes SAR possèdent des résolutions beaucoup plus élevées, la distribution gaussienne ne peut donc plus décrire les fluctuations entre les cellules de résolution. Dans les scénarios à texture élevée, une alternative consiste à utiliser les lois gaussiennes composées (Definition F.2.5)

$$\mathbf{k} = \sqrt{\tau}\mathbf{n} \tag{F.28}$$

où τ est un paramètre de texture dont la distribution n'est pas spécifiée, d'espérance égale à 1, et $\mathbf{n}$ est le vecteur de speckle, indépendant de τ, qui suit une loi gaussienne.

Filtrage des images PolSAR

Les méthodes de filtrage PolSAR peuvent être réparties en plusieurs catégories en fonction de leur principe.

Le premier groupe comprend les méthodes de filtrage utilisant une fenêtre ou un voisinage locale, comme par exemple *multi-look* ou *boxcar* filtre, LLMMSE (*Local Linear Minimum Mean Square Error*) [108] filtre, IDAN (*Intensity-Driven Adaptive-Neighborhood*) [175], etc. Un autre groupe est composé des méthodes basées sur les équations aux dérivées partielles comme par exemple AD (*Anisotropic Diffusion*) filtre [143]. Ensuite, les méthodes variationnelles utilisent des informations de régularisation globales pour éliminer le bruit dans les images. Ces dernières années, un intérêt croissant s'est porté sur les méthodes de filtrage basées sur l'apprentissage automatique. Enfin, le dernier groupe est composé des méthodes non locales. Dans la section F.6, nous nous intéressons à ce dernier groupe et ces adaptations aux images PolSAR.

F.3 Nouvelles propriétés des M-estimateurs

Cette section propose une approche originale pour mieux comprendre le comportement des M-estimateurs robustes de la matrice de dispersion. À cette fin, une nouvelle représentation par « noyau gaussien » pour les distributions CES est proposée et l'équivalent Wishart (EW) d'un M-estimateur est introduit. La distribution asymptotique entre un M-estimateur et son EW est dérivée.

F.3.1 Modèle équivalent Wishart

Le modèle de « noyau gaussien », utilisé comme une alternative à la représentation stochastique classique (Eq. (F.5)) des vecteurs distribués CES est défini comme suit.

Définition F.3.1. *Représentation par noyaux gaussiens*
Un vecteur aléatoire $\mathbf{z} \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}})$ peut être représenté par

$$\mathbf{z} \overset{d}{=} \frac{\sqrt{\mathcal{Q}}}{\|\mathbf{g}\|}\mathbf{A}\mathbf{g} \tag{F.29}$$

où $\mathbf{\Sigma} = \mathbf{A}\mathbf{A}^H$ est une factorisation de $\mathbf{\Sigma}$ et $\mathbf{g} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. $\mathcal{Q}$ est une variable aléatoire réelle non négative, indépendante de $\mathbf{g}$ avec une p.d.f. dépendant uniquement de $g_{\mathbf{z}}$. Nous appelons $\mathbf{x} = \mathbf{A}\mathbf{g}$ le noyau gaussien de $\mathbf{z}$.

Nous pouvons maintenant introduire l'équivalent Wishart d'un M-estimateur.

Définition F.3.2. *Modèle équivalent Wishart*
Soit $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ un n-échantillon où $\mathbf{z}_i = \sqrt{\mathcal{Q}_i}/\|\mathbf{g}_i\|\mathbf{A}\mathbf{g}_i$ suivant $\mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}})$, $i = 1, \ldots, n$, et soient $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ leurs noyaux gaussiens sous la forme $\mathbf{x}_i = \mathbf{A}\mathbf{g}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma})$, $i = 1, \ldots, n$. Soit $\widehat{\mathbf{\Sigma}}$ est un M-estimateur construit avec $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ en utilisant l'Eq. (F.7). La SCM construit avec les noyaux gaussiens, i.e.,

$$\widehat{\mathbf{\Sigma}}_{\text{EW}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^H \tag{F.30}$$

représente l'Équivalent Wishart (EW) de $\widehat{\mathbf{\Sigma}}$. Notons que $\widehat{\mathbf{\Sigma}}_{\text{EW}}$ suit donc, par construction, une loi de Wishart. Cet estimateur est non observable et est utilisé à des fins théoriques.

F.3.2 Convergence vers l'EW

Nous considérons $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ un n-échantillon où $\mathbf{z}_i \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}})$, $i = 1, \ldots, n$, et nous estimons la matrice de dispersion de la population en utilisant un M-estimateur défini avec l'Eq. (F.7). Ensuite, nous obtenons les résultats suivants.

Théorème F.3.1. *Cas général*
Soit σ donné par l'Eq. (F.8). La distribution asymptotique de $\sigma\widehat{\mathbf{\Sigma}} - \widehat{\mathbf{\Sigma}}_{\text{EW}}$ est donnée par

$$\sqrt{n}\text{vec}\left(\sigma\widehat{\mathbf{\Sigma}} - \widehat{\mathbf{\Sigma}}_{\text{EW}}\right) \xrightarrow{d} \mathcal{GCN}(\mathbf{0}, \mathbf{\Gamma}, \mathbf{\Omega}) \tag{F.31}$$

où $\mathbf{\Gamma}$ et $\mathbf{\Omega}$ sont définis par

$$\begin{cases} \mathbf{\Gamma} = \sigma_1 \mathbf{\Sigma}^T \otimes \mathbf{\Sigma} + \sigma_2 \text{vec}(\mathbf{\Sigma})\text{vec}(\mathbf{\Sigma})^H, \\ \mathbf{\Omega} = \sigma_1 \left(\mathbf{\Sigma}^T \otimes \mathbf{\Sigma}\right)\mathbf{K} + \sigma_2 \text{vec}(\mathbf{\Sigma})\text{vec}(\mathbf{\Sigma})^T \end{cases} \tag{F.32}$$

avec σ_1 et σ_2 donnés par

$$\begin{aligned} \sigma_1 &= \left(a_M p(p+1) + c(c - 2b_M)\right)/c_M^2, \\ \sigma_2 &= \vartheta_2 - 2p(b_M - c_M)/c_M/(c_M - p^2), \end{aligned} \tag{F.33}$$

où a_M, c_M sont donnés par l'Eq. (F.12) et $b_M = \mathbb{E}[\Psi(\sigma\mathcal{Q})\|\mathbf{g}\|^2]$.

F.3.3 Cas particuliers

Dans la suite, nous présentons les résultats pour les M-estimateurs particuliers listés dans la section F.2 et nous discutons leurs valeurs.

- M-estimateur de Tyler $\implies \sigma_1 = \dfrac{1}{p}$ et $\sigma_2 = \dfrac{p-1}{p^2}$
- M-estimateur d'Huber $\implies b_M = (p(p+1)F_{2p+4}(2\lambda) + p\lambda(1 - F_{2p+2}(2\lambda)))/\beta$
- M-estimateur de Student $\implies \sigma_1 = (p+\nu/2)^{-1}$ et $\sigma_2 = 2/\nu\ (p+1+\nu/2)(p+\nu/2)^{-1}$

Discussion

Tout d'abord, il faut noter que résultats pour le M-estimateur de Tyler sont valides pour toutes les distributions CES. Les résultats pour le M-estimateur d'Huber ont été obtenus sous la configuration des données gaussiennes. Les résultats pour le M-estimateur de Student ont été obtenus dans le cas MLE.

Ensuite, une remarque importante est que les facteurs σ_1 et σ_2 sont beaucoup plus petits que ceux du régime asymptotique standard (ϑ_1 et ϑ_2 dans Théorème F.2.2).

Ce résultat montre que les M-estimateurs sont asymptotiquement plus proches de l'EW que de la vraie matrice de dispersion, ce qui signifie que le comportement d'un M-estimateur peut être mieux caractérisé par le comportement de la matrice de Wishart correspondante que par son régime asymptotique classique.

F.4 Estimation robuste des éléments propres de la matrice de dispersion

Cette section traite de l'estimation des paramètres de décomposition en éléments propres (EVD) des statistiques de second ordre. Trois cas différents, basés sur la structure de la matrice de dispersion, sont analysés :

- $\lambda_1 > \ldots > \lambda_p > 0$,
- $\lambda_1 = \ldots = \lambda_{j_1} > \lambda_{j_1+1} = \ldots = \lambda_{j_2} > \ldots > \lambda_{j_k} = \ldots = \lambda_p > 0$,
- $\lambda_1 > \ldots > \lambda_2 > \ldots \lambda_r > \lambda_{r+1} = \ldots = \lambda_p > 0$,

où $\lambda_1, \ldots, \lambda_p$ sont les valeurs propres de la dispersion.

F.4.1 Propriétés asymptotiques des éléments propres des M-estimateurs

L'EVD d'une matrice (de dispersion) $\mathbf{\Sigma}$ est définie comme

$$\mathbf{\Sigma} \overset{\text{EVD}}{=} \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H, \quad \text{with} \quad \begin{array}{l} \mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_p] \in \mathcal{U}_p^p, \\ \mathbf{\Lambda} = \text{diag}(\boldsymbol{\lambda}), \\ \boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_p]\,. \end{array} \tag{F.34}$$

Dans la suite, nous supposons que les valeurs propres sont ordonnées $\lambda_1 > \ldots > \lambda_p > 0$ et qu'un élément de chaque $\mathbf{u}_j$ est réel positif (garantissant l'unicité de cette définition).

En considérant à nouveau le modèle équivalent Wishart proposé dans la section F.3.2, nous définissons :

$$\begin{aligned} \widehat{\mathbf{\Sigma}} &\overset{\text{EVD}}{=} \widehat{\mathbf{U}}^M \widehat{\mathbf{\Lambda}}^M \left(\widehat{\mathbf{U}}^M\right)^H, \\ \widehat{\mathbf{\Sigma}}_{\text{EW}} &\overset{\text{EVD}}{=} \widehat{\mathbf{U}}^{\text{EW}} \widehat{\mathbf{\Lambda}}^{\text{EW}} \left(\widehat{\mathbf{U}}^{\text{EW}}\right)^H. \end{aligned} \tag{F.35}$$

Nous obtenons alors les résultats suivants.

Théorème F.4.1. *Régime standard*
La distribution asymptotique des valeurs propres et des vecteurs propres d'un M-estimateur (Eqs. (F.35)) est caractérisée par

$$\begin{cases} \sqrt{n}\left(\sigma\widehat{\boldsymbol{\lambda}}^M - \boldsymbol{\lambda}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \vartheta_1\mathbf{\Lambda}^2 + \vartheta_2\boldsymbol{\lambda}\boldsymbol{\lambda}^T\right), \\ \sqrt{n}\mathbf{\Pi}_j^{\perp}\widehat{\mathbf{u}}_j^M \xrightarrow{d} \mathcal{CN}\left(\mathbf{0}, \mathbf{\Xi}_j\right). \end{cases} \tag{F.36}$$

avec

$$\mathbf{\Xi}_j = \vartheta_1\lambda_j\mathbf{U}\mathbf{\Lambda}(\lambda_j\mathbf{I} - \mathbf{\Lambda})^{+^2}\mathbf{U}^H \tag{F.37}$$

où $\mathbf{\Pi}_j^{\perp} = \mathbf{I} - \mathbf{u}_j\mathbf{u}_j^H$ et ϑ_1, ϑ_2 données par l'Eq. (F.11).

Théorème F.4.2. *EW*
La distribution asymptotique de la différence entre les valeurs propres et les vecteurs propres d'un M-estimateur et de son EW est donnée par

$$\begin{cases} \sqrt{n}\left(\sigma\widehat{\boldsymbol{\lambda}}^{M} - \widehat{\boldsymbol{\lambda}}^{\mathrm{EW}}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \sigma_1\boldsymbol{\Lambda}^2 + \sigma_2\boldsymbol{\lambda}\boldsymbol{\lambda}^T\right), \\ \sqrt{n}\boldsymbol{\Pi}_j^{\perp}\left(\widehat{\mathbf{u}}_j^{M} - \widehat{\mathbf{u}}_j^{\mathrm{EW}}\right) \xrightarrow{d} \mathcal{CN}\left(\mathbf{0}, \sigma_1/\vartheta_1\boldsymbol{\Xi}_j\right). \end{cases} \tag{F.38}$$

avec $\boldsymbol{\Xi}_j$ *et* σ_1, σ_2 *donnés par les Eqs.* (F.37) *et* (F.33)*, respectivement.*

F.4.2 EFusion

Étant donné un n-échantillon $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ de vecteurs indépendants de dimension p, la SCM minimise de façon unique la fonction de log-vraisemblance

$$l(\boldsymbol{\Sigma}; \widehat{\boldsymbol{\Sigma}}_{\mathrm{SCM}}) = \mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\Sigma}}_{\mathrm{SCM}}) + \log\{\det(\boldsymbol{\Sigma})\} \tag{F.39}$$

sur $\boldsymbol{\Sigma} \in \mathcal{S}^{p\times p}$. Lorsque la taille de l'échantillon n n'est pas supérieur à p, les grandes (resp. petites) valeurs propres de la SCM ont tendance à surestimer (resp. sous-estimer) largement les valeurs propres de la vraie matrice de covariance. Pour palier ce problème, des estimateurs régularisés ou pénalisés de la matrice de covariance ont été introduits dans une série d'articles [102, 104, 26, 7, 1, 136, 163, 129, 11, 127]. Une SCM régularisée (RSCM) est alors obtenue en minimisant

$$L(\boldsymbol{\Sigma}; \widehat{\boldsymbol{\Sigma}}_{\mathrm{RSCM}}; \eta) = l(\boldsymbol{\Sigma}; \widehat{\boldsymbol{\Sigma}}_{\mathrm{RSCM}}) + \eta\Pi(\boldsymbol{\Sigma}), \tag{F.40}$$

où $\Pi(\boldsymbol{\Sigma})$ est une fonction de pénalité non négative et $\eta \geq 0$ un paramètre de régularisation. Dans cette section, nous supposons que la matrice de covariance possède seulement quelques valeurs propres, c'est-à-dire qu'il existe k groupes de valeurs propres distinctes. Nous proposons un estimateur RSCM qui regroupe les valeurs propres en pénalisant les grandes différences entre les valeurs propres successives.

Méthode proposée

Soit $d_1 \geq \cdots \geq d_p > 0$ et $\lambda_1 \geq \cdots \geq \lambda_p > 0$ les valeurs propres ordonnées de la SCM et de $\boldsymbol{\Sigma}$, respectivement. De plus, soient $r_j = \log(\lambda_j) - \log(\lambda_{j+1})$ les écarts entre les valeurs propres logarithmiques successives de $\boldsymbol{\Sigma}$ et soient $r_j^{[0]} = \log(d_j) - \log(d_{j+1})$ les écarts entre log-valeurs propres de $\widehat{\boldsymbol{\Sigma}}_{\mathrm{SCM}}$, $j = 1, \ldots, p-1$. Nous proposons de minimiser l'Eq. (F.40) avec la pondération non convexe suivante

$$\Pi(\boldsymbol{\Sigma}) = \sum_{j=1}^{p-1} \frac{1}{6} \cdot \min\left\{1, 1 - \left(1 - \frac{(r_j/s)^2}{c^2}\right)^3\right\}, \tag{F.41}$$

où c est un paramètre de réglage défini par l'utilisateur et s est l'écart type de $r_j^{[0]}$ pour $j = 1, \ldots, p-1$. Cette fonction est appelée fonction de Tukey [152]. Dans la minimisation de l'Eq. (F.40), la pénalité dans l'Eq. (F.41) attribue des poids relativement importants aux petits écarts r_j et très grands écarts à des poids plus petits. L'algorithme est détaillé dans Algorithme 3.

Nous comparons les performances de l'estimateur proposé avec elasso [171]. Dans elasso, $\Pi(\boldsymbol{\Sigma}) = \sum_{j=1}^{p} a_j \log(\lambda_j)$ est utilisé comme fonction de pénalité, où les poids a_j sont obtenus en centrant des quantiles décroissants de la loi Marčenko-Pastur. Nous générons un échantillon de taille $n = 3000$ suivant une distribution gaussienne $p = 100$. Semblable à [171], la matrice de covariance $\boldsymbol{\Sigma}$ possède 40 valeurs propres égales à 20, 30 égales à 10 et 30 égales à 2.

La figure 3.8 illustre le processus de regroupement des valeurs propres avec elasso (eh haut) et eFusion (en bas) en fonction de la valeur du paramètre de pénalité. Les résultats

Algorithm 3: Algorithme eFusion

Entrée : $\mathbf{d}$: Valeurs propres de la SCM $\widehat{\mathbf{\Sigma}}_{\text{SCM}}$;
η : Paramètre de pénalité ;
c : Paramètre de Tukey.
Sortie : $\widehat{\boldsymbol{\lambda}}$:Valeurs propres pénalisées

Initialisation: $k \leftarrow 0$; $\boldsymbol{\lambda}^{[0]} \leftarrow \mathbf{d}$
Calculer $s = \text{SD}(\mathbf{r}^{[0]})$,
Répéter
  Calculer les écarts :
  $r_j^{[k]} \leftarrow \log(\lambda_j^{[k]}) - \log(\lambda_{j+1}^{[k]}),\ \ j = 1, \ldots, p-1,$
  Calculer les poids :
  $w_j^{[k]} \leftarrow \rho_c'(r_j^{[k]}/s)/(r_j^{[k]}/s),\quad j = 1, \ldots, p-1,$
  Calculer les valeurs propres :
  $\log \lambda_j^{[k+1]} \leftarrow \frac{1}{w_j^{[k]} + w_{j-1}^{[k]}} \Big(\frac{s^2}{\eta}(d_j/\lambda_j^{[k]} - 1) + w_j^{[k]} \log \lambda_{j+1}^{[k]} + w_{j-1}^{[k]} \log \lambda_{j-1}^{[k+1]} \Big),$
  for $j = 1, \ldots, p$.
  $k \leftarrow k + 1$
jusqu'à *la convergence*
$\hat{\boldsymbol{\lambda}} \leftarrow \big(\exp(\log \lambda_1^{[k+1]}), \ldots, \exp(\log \lambda_p^{[k+1]}) \big)^T$

montrent l'amélioration significative que notre estimateur peut offrir. D'abord, il donne une estimation non biaisée car les trois groupes de valeurs propres sont bien séparés et proches de leur valeur réelle. De plus, il n'est pas nécessaire de rechercher une valeur optimale de paramètre η, contrairement à elasso.

F.4.3 Propriétés asymptotiques du sous-espace principal des M-estimateurs

Considérons le cas d'une matrice de dispersion à structure « rang faible plus identité », couramment utilisée en traitement du signal pour prendre en compte les signaux de faible dimension intégrés dans le bruit blanc

$$\mathbf{\Sigma} = \mathbf{\Sigma}_r + \gamma^2 \mathbf{I}_p \quad \overset{\text{EVD}}{=} \quad [\mathbf{U}_r | \mathbf{U}_r^\perp] \mathbf{\Lambda} [\mathbf{U}_r | \mathbf{U}_r^\perp]^H \tag{F.42}$$

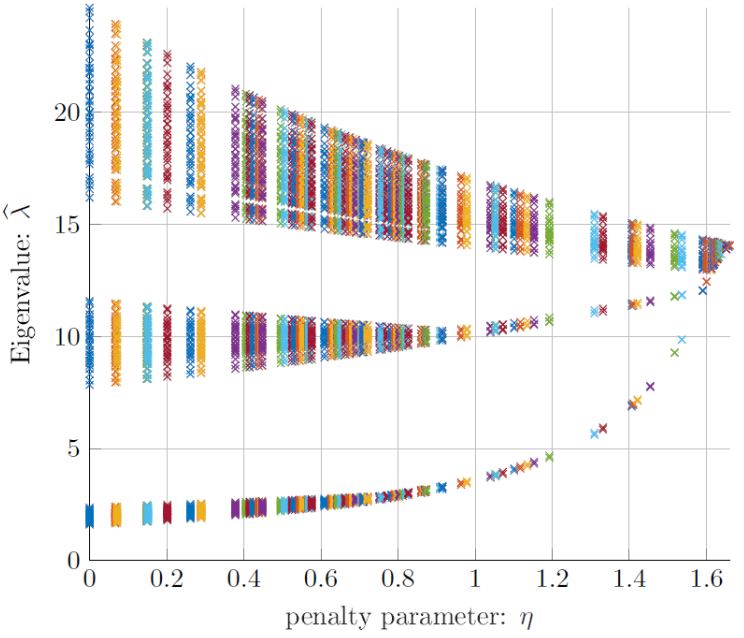
avec $\mathbf{\Sigma}_r = \mathbf{U}_r \mathbf{\Lambda}_r \mathbf{U}_r^H$, où $\mathbf{U}_r \in \mathcal{U}_r^p$ et $\mathbf{\Lambda}_r \in \mathbb{R}^{r \times r}$. Nous définissons $\mathcal{R}_r\{.\}$ l'opérateur qui extrait le sous-espace principal d'une matrice donnée comme

$$\begin{array}{rccc} \mathcal{R}_r : & \mathcal{H} & \longrightarrow & \mathcal{G}_r^p \\ & \mathbf{\Sigma} \overset{\text{EVD}}{=} [\mathbf{U}_r | \mathbf{U}_r^\perp] \mathbf{\Lambda} [\mathbf{U}_r | \mathbf{U}_r^\perp]^H & \longmapsto & \mathbf{U}_r \mathbf{U}_r^H \end{array} \tag{F.43}$$

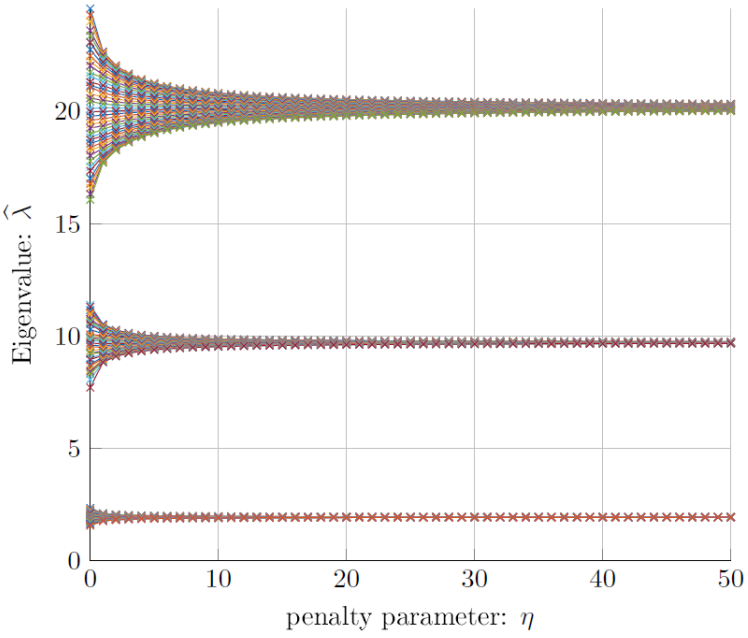
où $\mathcal{G}_r^p$ est l'ensemble des projecteurs orthogonaux de rang r de $\mathbb{C}^{p \times p}$. Soit $\widehat{\mathbf{\Sigma}}$ un M-estimateur construit avec un n-échantillon de vecteurs complexes indépendants de taille p avec $\mathbf{z}_i \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}})$ où $\mathbf{\Sigma}$ est défini par l'Eq. (F.42) et soit $\widehat{\mathbf{\Sigma}}_{\text{EW}}$ son EW défini par la Définition F.3.2. Nous avons les principaux sous-espaces correspondants

$$\begin{cases} \mathbf{\Pi}_r = \mathcal{R}_r\{\mathbf{\Sigma}\}, \\ \widehat{\mathbf{\Pi}}_r^M = \mathcal{R}_r\{\widehat{\mathbf{\Sigma}}\}, \\ \widehat{\mathbf{\Pi}}_r^{\text{EW}} = \mathcal{R}_r\{\widehat{\mathbf{\Sigma}}_{\text{EW}}\}. \end{cases} \tag{F.44}$$

Théorème F.4.3. *Projecteur robuste*
Soit $\widehat{\mathbf{\Pi}}_r^M$ *l'estimateur du sous-espace principal* $\mathbf{\Pi}_r$ *obtenu avec un* M*-estimateur*



(a) elasso



(b) eFusion avec $c = 0.794$

FIGURE F.1 – Exemple de groupement des valeurs propres avec elasso (en haut) et avec eFusion (en bas) avec $c = 0.794$.

(Eq. (F.44)). La distribution asymptotique de $\widehat{\mathbf{\Pi}}_r^M$ *est donnée par*

$$\textbf{\textit{(RS)}} \quad \sqrt{n}\mathrm{vec}\left(\widehat{\mathbf{\Pi}}_r^M - \mathbf{\Pi}_r\right) \xrightarrow{d} \mathcal{GCN}\left(\mathbf{0}, \vartheta_1 \mathbf{\Gamma}_{\mathbf{\Pi}}, \vartheta_1 \mathbf{\Gamma}_{\mathbf{\Pi}} \mathbf{K}\right), \tag{F.45}$$

$$\textbf{\textit{(EW)}} \quad \sqrt{n}\mathrm{vec}\left(\widehat{\mathbf{\Pi}}_r^M - \widehat{\mathbf{\Pi}}_r^{\mathrm{EW}}\right) \xrightarrow{d} \mathcal{GCN}\left(\mathbf{0}, \sigma_1 \mathbf{\Gamma}_{\mathbf{\Pi}}, \sigma_1 \mathbf{\Gamma}_{\mathbf{\Pi}} \mathbf{K}\right) \tag{F.46}$$

où

$$\mathbf{\Gamma_\Pi} = \mathbf{A}^T \otimes \mathbf{B} + \mathbf{B}^T \otimes \mathbf{A} \tag{F.47}$$

avec $\mathbf{A} = \mathbf{U}_r \left(\gamma^2 \mathbf{\Lambda}_r^{-2} + \mathbf{\Lambda}_r^{-1}\right) \mathbf{U}_r^H$, $\mathbf{B} = \gamma^2 \mathbf{\Pi}_r^\perp$, ϑ_1, ϑ_2 *et* σ_1, σ_2 *donnée par les Eqs.* (F.11) *et* (F.33), *respectivement.*

F.5 Détection robuste

Dans cette section, nous analysons les performances asymptotiques de la distance de Mahalanobis robuste et des détecteurs adaptatifs robustes dans le contexte d'observations non gaussiennes. Nous analysons les propriétés asymptotiques de différents détecteurs définis dans la section F.2.

F.5.1 Distance de Mahalanobis robuste

La distance de Mahalanobis [115, 113] est l'une des mesures les plus courantes en statistiques multivariées, en traitement du signal et en apprentissage automatique. La distance de Mahalanobis entre $\mathbf{z}$ et $\boldsymbol{\mu}$ est donnée par

$$\Delta^2(\boldsymbol{\mu}, \mathbf{\Sigma}) = (\mathbf{z} - \boldsymbol{\mu})^H \mathbf{\Sigma}^{-1} (\mathbf{z} - \boldsymbol{\mu}). \tag{F.48}$$

où $\boldsymbol{\mu}$ est la moyenne de la population et $\mathbf{\Sigma}$ la matrice de dispersion commune. Puisque nous travaillons avec des vecteurs de moyenne connue, nous pouvons simplement analyser $\Delta^2(\mathbf{\Sigma}) = \mathbf{z}^H \mathbf{\Sigma}^{-1} \mathbf{z}$. Si les données suivent la loi gaussienne, nous avons

$$\Delta^2 \left(\mathbf{\Sigma}\right) \sim (1/2)\, \chi^2_{2p}. \tag{F.49}$$

Si la SCM est utilisée au lieu de la vraie matrice de dispersion et sous l'hypothèse gaussienne

$$\Delta^2 \left(\widehat{\mathbf{\Sigma}}_{\text{SCM}}\right) \sim n\beta' \left(p, n - p + 1\right) \tag{F.50}$$

où $\beta'(a, b)$ désigne une distribution bêta de $1^{\text{ème}}$ espèce avec des paramètres de forme réels a et b.

Considérons un $(n+1)$-échantillon $(\mathbf{z}, \mathbf{z}_1, \ldots, \mathbf{z}_n)$ de vecteurs complexes indépendants de taille p avec $\mathbf{z}_i \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}})$ où $\mathbf{z}_i \sim \mathcal{CES}\left(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}}\right)$. Soit $\widehat{\mathbf{\Sigma}}$ un M-estimateur construit avec $(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ défini par l'Eq. (F.7) et soit $\widehat{\mathbf{\Sigma}}_{\text{EW}}$ son EW défini par la Définition F.3.2. Nous définissons alors la distance de Mahalanobis robuste Δ_M et sa distance de Mahalanobis équivalente (ME) comme

$$\begin{cases} \Delta = \Delta(\mathbf{\Sigma}), \\ \Delta_M = \Delta(\sigma\widehat{\mathbf{\Sigma}}), \\ \Delta_{\text{ME}} = \Delta(\widehat{\mathbf{\Sigma}}_{\text{EW}}). \end{cases} \tag{F.51}$$

On obtient alors les résultats suivants.

Théorème F.5.1. *Distance de Mahalanobis robuste*
Soit Δ_M *la distance de Mahalanobis robuste et* Δ_{ME} *sa GCEM, définies par les Eqs.* (F.51). *La distribution asymptotique conditionnelle de* Δ_M *est donnée par*

$$\textbf{\textit{(RS)}} \qquad \sqrt{n}\left(\Delta_M^2 - \Delta^2\right)_{\mathbf{z}} \xrightarrow{d} \mathcal{N}\left(0, (\vartheta_1 + \vartheta_2)\Delta^4\right), \tag{F.52}$$

$$\textbf{\textit{(ME)}} \qquad \sqrt{n}\left(\Delta_M^2 - \Delta_{\text{GCEM}}^2\right)_{\mathbf{z}} \xrightarrow{d} \mathcal{N}\left(0, (\sigma_1 + \sigma_2)\Delta^4\right), \tag{F.53}$$

où $(.)_{\mathbf{z}}$ *dénote la distribution conditionnelle à* $\mathbf{z}$, ϑ_1 *et* ϑ_2 *sont donnés par l'Eq.* (F.11) *et* σ_1 *et* σ_2 *sont donnés par l'Eq.* (F.33).

Remarque F.5.1. *La variance asymptotique de la distance Mahalanobis robuste donnée par l'Eq.* (F.53) *est plus petite que la variance donnée par l'Eq.* (F.52) *puisque* $\sigma_1 + \sigma_2 < \vartheta_1 + \vartheta_2$. *Ces résultats révèlent que la distribution de la distance de Mahalanobis robuste est mieux approximée avec la loi bêta de* $1^{ème}$ *espèce qu'avec la loi chi-carré.*

F.5.2 Détection « rang plein » robuste

Nous considérons le problème de détection de signal défini dans la section F.2 en supposant que les données secondaires sont non-gaussiennes.

$$\begin{cases} H_0 : \mathbf{z} = \mathbf{c} & \mathbf{z}_i = \mathbf{c}_i, \quad i = 1, \dots, n, \\ H_1 : \mathbf{z} = \alpha \mathbf{p} + \mathbf{c} & \mathbf{z}_i = \mathbf{c}_i, \quad i = 1, \dots, n, \end{cases} \tag{F.54}$$

où $\mathbf{c}_i \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}})$ sont n observations indépendantes ne contenant pas de signal et utilisées pour estimer la matrice de dispersion $\mathbf{\Sigma}$.

Considérons un $(n+1)$-échantillon $(\mathbf{z}, \mathbf{z}_1, \dots, \mathbf{z}_n)$ des données secondaires du problème défini par (F.54) avec $\mathbf{z}_i \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}})$. Soit $\widehat{\mathbf{\Sigma}}$ un M-estimateur construit avec $(\mathbf{z}_1, \dots, \mathbf{z}_n)$ défini par l'Eq. (F.7) et soit $\widehat{\mathbf{\Sigma}}_{\text{EW}}$ son EW défini par la Définition F.3.2. Soit $\Lambda(\cdot)$ une statistique de détection (soit Kelly l'Eq. (F.14), AMF l'Eq. (F.15), ANMF dans l'Eq. (F.16), ou Rao dans l'Eq. (F.17)). Nous définissons alors le détecteur robuste $\widehat{\Lambda}^M$ et son détecteur équivalent (DE) $\widehat{\Lambda}^{\text{DE}}$ comme

$$\begin{cases} \widehat{\Lambda}^M = \Lambda(\sigma\widehat{\mathbf{\Sigma}}), \\ \widehat{\Lambda}^{\text{DE}} = \Lambda(\widehat{\mathbf{\Sigma}}_{\text{EW}}). \end{cases} \tag{F.55}$$

On obtient alors les résultats suivants.

Théorème F.5.2. *AMF, Kelly et Rao robustes*
Soit $\widehat{\Lambda}^M$ le détecteur robuste avec $\widehat{\Lambda}^M \in \{\widehat{\Lambda}_{\text{Kelly}}, \widehat{\Lambda}_{\text{Rao}}, \widehat{\Lambda}_{\text{AMF}}\}$. Soit $\widehat{\Lambda}^{\text{DE}}$ le DE de $\widehat{\Lambda}^M$, définis par les Eqs. (F.55). La distribution asymptotique conditionnelle de $\widehat{\Lambda}^M$ est donnée par

$$\textbf{\textit{(RS)}} \qquad \sqrt{n}\left(\widehat{\Lambda}^M - \Lambda\right)_{\mathbf{z}} \xrightarrow{d} \mathcal{N}\left(0, \vartheta_1 \sigma_{\mathcal{X}} + \vartheta_2 \Lambda^2\right) \tag{F.56}$$

$$\textbf{\textit{(DE)}} \qquad \sqrt{n}\left(\widehat{\Lambda}^M - \widehat{\Lambda}^{\text{DE}}\right)_{\mathbf{z}} \xrightarrow{d} \mathcal{N}\left(0, \sigma_1 \sigma_{\mathcal{X}} + \sigma_2 \Lambda^2\right) \tag{F.57}$$

avec $\sigma_{\mathcal{X}} = \Lambda\left(2\mathbf{z}^H \mathbf{\Sigma}^{-1}\mathbf{z} - \Lambda\right)$, ϑ_1 et ϑ_2 définis par l'Eq. (F.11) et σ_1 et σ_2 définis par l'Eq. (F.33).

Théorème F.5.3. *ANMF robuste*
Soit $\widehat{\Lambda}^M_{\text{ANMF}}$ le détecteur robuste défini par l'Eq. (F.16). La distribution asymptotique conditionnelle de $\widehat{\Lambda}^M_{\text{ANMF}}$ est donnée par

$$\textbf{\textit{(RS)}} \qquad \sqrt{n}\left(\widehat{\Lambda}^M_{\text{ANMF}} - \Lambda_{\text{NMF}}\right)_{\mathbf{z}} \xrightarrow{d} \mathcal{N}\left(0, \vartheta_1 \sigma_{\mathcal{H}}\right) \tag{F.58}$$

$$\textbf{\textit{(DE)}} \qquad \sqrt{n}\left(\widehat{\Lambda}^M_{\text{ANMF}} - \widehat{\Lambda}^{\text{DE}}_{\text{ANMF}}\right)_{\mathbf{z}} \xrightarrow{d} \mathcal{N}\left(0, \sigma_1 \sigma_{\mathcal{H}}\right) \tag{F.59}$$

où $\sigma_{\mathcal{H}} = 2\Lambda_{\text{NMF}}(\Lambda_{\text{NMF}} - 1)^2$, ϑ_1 défini par l'Eq. (F.11) et σ_1 défini par l'Eq. (F.33).

Dans les résultats précédents, on peut noter que, comme $\sigma_1 < \vartheta_1$ et $\sigma_2 \leq \vartheta_2$, la variance asymptotique dans l'Eq. (F.57) (resp. l'Eq. (F.59)) est plus petite que celle de l'Eq. (F.56) (resp. l'Eq. (F.58)). Ce résultat justifie théoriquement que le comportement de $\widehat{\Lambda}^M$ est plus proche de $\widehat{\Lambda}^{\text{DE}}$ que de Λ. Une des conséquences importantes est la meilleure prédiction des performances de détection en utilisant des résultats déjà établis pour $\widehat{\Lambda}^{\text{DE}}$ au lieu de Λ.

Ceci est illustré sur la figure F.2, qui présente les courbes « PFA-seuil » pour les détecteurs t-ANMF, TyE-AMF, DE et SCM-ANMF, pour $n = 20$ et $n = 100$. Les relations théoriques pour le détecteur EW-ANMF et le NMF (théorique car $\mathbf{\Sigma}$ est supposé connu), sont aussi présentées. Notons que pour le M-estimateur de Student, on a $\nu = 2$ (cas MLE). On constate une très bonne adéquation entre les PFA obtenues pour t-ANMF, TyE-AMF, et la relation théorique, comme attendu par le théorème F.5.3 : le comportement des détecteurs robustes est mieux expliqué par celui du DE que par le NMF. Ceci est particulièrement flagrant pour n petit, puisque toutes les courbes se rapprochent lorsque n augmente.

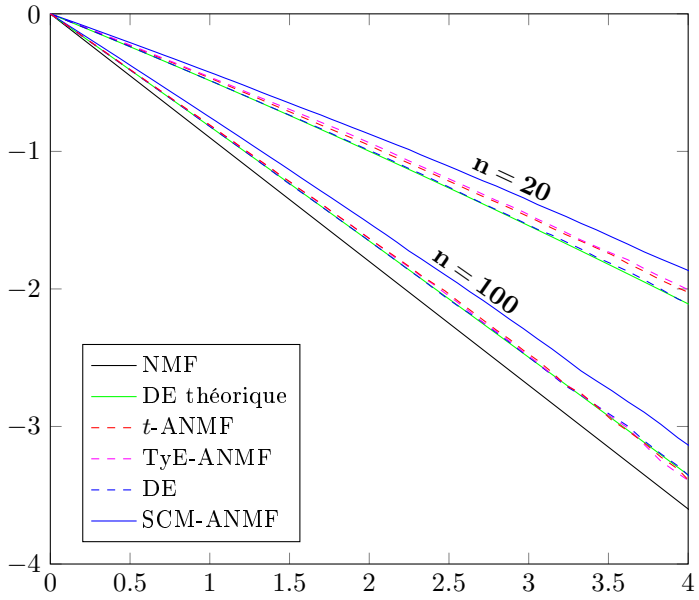


FIGURE F.2 – Comparaison entre les relations $P_{fa} - \lambda$ pour t-ANMF, TyE-ANMF and SCM-ANMF avec les résultats empiriques et théoriques pour le DE et le NMF ; données suivant une t-distribution avec $\nu = 2$, $p = 10$.

F.5.3 Détection « rang faible » robuste

Considérons maintenant le problème de détection de LR défini comme suit

$$\begin{cases} H_0 : \mathbf{z} = \widetilde{\mathbf{c}} & \mathbf{z}_i = \widetilde{\mathbf{c}}_i, \quad i = 1, \ldots, n, \\ H_1 : \mathbf{z} = \alpha \mathbf{p} + \widetilde{\mathbf{c}} & \mathbf{z}_i = \widetilde{\mathbf{c}}_i, \quad i = 1, \ldots, n, \end{cases} \tag{F.60}$$

où $\widetilde{\mathbf{c}}_i \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}})$ sont n observations indépendantes ne contenant pas de signal avec $\mathbf{\Sigma} = \mathbf{\Sigma}_r + \gamma^2 \mathbf{I}_p$ définie par l'Eq. (F.42).

Soit $\widehat{\mathbf{\Sigma}}$ un M-estimateur construit avec un n-échantillon de vecteurs complexes indépendants de taille p avec $\mathbf{z}_i \sim \mathcal{CES}(\mathbf{0}, \mathbf{\Sigma}, g_{\mathbf{z}})$ où $\mathbf{\Sigma}$ est défini par l'Eq. (F.42) et soit $\widehat{\mathbf{\Sigma}}_{\text{EW}}$ son EW défini par la Définition F.3.2. Nous avons les principaux sous-espaces et les détecteurs de LR correspondants

$$\begin{cases} \mathbf{\Pi}_r^{\perp} = \mathbf{I} - \mathcal{R}_r\{\mathbf{\Sigma}\}, \\ \widehat{\mathbf{\Pi}}_r^{\perp M} = \mathbf{I} - \mathcal{R}_r\{\widehat{\mathbf{\Sigma}}\}, \\ \widehat{\mathbf{\Pi}}_r^{\perp \text{EW}} = \mathbf{I} - \mathcal{R}_r\{\widehat{\mathbf{\Sigma}}_{\text{EW}}\}. \end{cases} \tag{F.61}$$

Ensuite, on peut définir les détecteurs de LR

$$\begin{cases} \Lambda_{\text{LR}} = \Lambda_{\text{LR}}\left(\mathbf{\Pi}_r^{\perp}\right), \\ \widehat{\Lambda}_{\text{LR}}^{M} = \Lambda_{\text{LR}}\left(\widehat{\mathbf{\Pi}}_r^{\perp M}\right), \\ \widehat{\Lambda}_{\text{LR}}^{\text{DE}} = \Lambda_{\text{LR}}\left(\widehat{\mathbf{\Pi}}_r^{\perp \text{EW}}\right). \end{cases} \tag{F.62}$$

Théorème F.5.4. *LR-ANMF robuste*
Considérons le test LR-ANMF robuste et son DE définis par l'Eq. (F.21)*. Ainsi, condition-*

nellement à la distribution de $\mathbf{z}$*, la distribution asymptotique de* $\widehat{\Lambda}_{\mathrm{LR}}^{M}$ *est donnée par*

$$\textbf{(RS)} \qquad \sqrt{n}\left(\widehat{\Lambda}_{\mathrm{LR}}^{M} - \Lambda_{\mathrm{LR}}\right)_{\mathbf{z}} \xrightarrow{d} \mathcal{N}\left(0, 2\vartheta_1\gamma^2\left(1-\Lambda_{\mathrm{LR}}\right)\left(b - a\Lambda_{\mathrm{LR}}\right)\right), \tag{F.63}$$

$$\textbf{(DE)} \qquad \sqrt{n}\left(\widehat{\Lambda}_{\mathrm{LR}}^{M} - \widehat{\Lambda}_{\mathrm{LR}}^{\mathrm{DE}}\right)_{\mathbf{z}} \xrightarrow{d} \mathcal{N}\left(0, 2\sigma_1\gamma^2\left(1-\Lambda_{\mathrm{LR}}\right)\left(b - a\Lambda_{\mathrm{LR}}\right)\right), \tag{F.64}$$

où

$$a = \frac{\mathbf{z}^H\mathbf{A}\mathbf{p}}{\mathbf{z}^H\mathbf{\Pi}_{\mathbf{c}}^{\perp}\mathbf{p}} + \frac{\mathbf{p}^H\mathbf{A}\mathbf{z}}{\mathbf{p}^H\mathbf{\Pi}_{\mathbf{c}}^{\perp}\mathbf{z}} \quad \textit{et} \quad b = \frac{\mathbf{z}^H\mathbf{A}\mathbf{z}}{\mathbf{z}^H\mathbf{\Pi}_{\mathbf{c}}^{\perp}\mathbf{z}} + \frac{\mathbf{p}^H\mathbf{A}\mathbf{p}}{\mathbf{p}^H\mathbf{\Pi}_{\mathbf{c}}^{\perp}\mathbf{p}},$$

avec $\mathbf{A}$ *défini dans le théorème F.4.3.*

F.6 Débruitage robuste des images polarimétriques

Dans cette section, nous proposons une nouvelle méthode de filtrage des images radar à synthèse d'ouverture polarimétrique (PolSAR), nommée M-NL. Plus précisément, nous développons une nouvelle approche statistique pour le calcul des poids dans les approches non locales (NL). L'objectif est de présenter un critère simple permettant de détecter des pixels similaires dans une image PolSAR, basé sur les nouvelles propriétés statistiques des M-estimateurs dérivés dans la section F.3.

F.6.1 Méthodes non locales

Contrairement aux filtres locaux, qui utilisent les valeurs d'un groupe de pixels entourant un pixel cible pour filtrer l'image, les méthodes non locales (NLM) utilisent tous les pixels de l'image au cours du traitement. De plus, au lieu d'une simple comparaison de pixels, la comparaison de *patches* (petit bloc de l'image) est proposée pour assurer une robustesse par rapport au bruit. La valeur finale dans le pixel central du patch est obtenue sous forme de moyenne pondérée basée sur les (dis)similarités calculées. Afin de réduire les coûts de calcul et d'améliorer les performances, la zone de recherche est limitée à une grande fenêtre [51]. Les méthodes NLM ont été principalement introduites pour les images numériques. Elles ont été adaptées avec succès au filtrage des données PolSAR. Dans la suite, nous présentons brièvement une des méthodes NLM avancées pour le filtrage PolSAR, NL-SAR [46].

NL-SAR

NL-SAR (SAR non local) est une méthode NLM avancée pour le débruitage d'images radar. Elle contient plusieurs étapes suivantes : (1) pré-estimation, (2) calcul des poids, (3) réduction du biais et (4) sélection de la meilleure estimation. La méthode fait d'abord une pré-estimation de matrice de dispersion dans chaque pixel. En comparant ces matrices pré-estimées, la méthode calcule les (dis)similarités entre pixels. Chaque pixel dans l'image est comparé aux pixels dans une fenêtre de recherche. Grâce aux (dis)similarités, la méthode calcule les poids associés aux pixels et calcule la matrice de dispersion dans le pixel central comme la moyenne pondérée. Pour le pixels où cette estimation introduit un grand biais, une étape supplémentaire est faite afin de réduire le biais. Finalement, cette procédure est répétée pour plusieurs ensembles de paramètres et l'estimation avec la variance minimale est choisie comme l'estimation finale.

F.6.2 M-NL

Dans cette section, nous présentons la méthode M-NL, proposée pour améliorer le calcul des poids de NL-SAR. La structure de l'algorithme est basée sur celle de NL-SAR. La méthode M-NL est récapitulée dans l'algorithme 4. La dernière partie est similaire à celle de NL-SAR (étapes 3 et 4) et ne sera pas détaillée en raison du manque d'espace.

Pré-estimation. La pré-estimation est faite pour chaque pixel en utilisant les pixels voisins provenant de petites fenêtres carrés de taille $S = (2s+1) \times (2s+1)$ où s est appelé *échelle*. L'estimateur utilisé est donnée par

$$\widehat{\mathbf{\Sigma}}_t = \frac{p+d}{S} \sum_{i=1}^{S} \frac{\mathbf{k}_i \mathbf{k}_i^H}{d + \mathbf{k}_i^H \widehat{\mathbf{\Sigma}}_t^{-1} \mathbf{k}_i} \tag{F.65}$$

où $\widehat{\mathbf{\Sigma}}_t$ représente un M-estimateur de Student.

Sélection des pixels. Ces valeurs pré-estimées permettent de sélectionner des échantillons voisins autour de chaque pixel dans une fenêtre de recherche circulaire. La comparaison entre pixels est ensuite faite en utilisant le M-test de Box

$$\mathcal{L}_{\text{Box}} = \left(\left| \widehat{\mathbf{\Sigma}}_1 \right|^{S/2} \left| \widehat{\mathbf{\Sigma}}_2 \right|^{S/2} \right) / \left| \widehat{\mathbf{\Sigma}} \right|^{S}$$

où $\widehat{\mathbf{\Sigma}}_1$ est obtenu avec $\mathbf{k}^{(1)} = (\mathbf{k}_1, \ldots, \mathbf{k}_S)$, $\widehat{\mathbf{\Sigma}}_2$ avec $\mathbf{k}^{(2)} = (\mathbf{k}_{S+1}, \ldots, \mathbf{k}_{2S})$ et $\widehat{\mathbf{\Sigma}}$ avec $\mathbf{k} = (\mathbf{k}^{(1)}, \mathbf{k}^{(2)})$. En modifiant la statistique $\mathcal{L}_{\text{Box}}$ [17], on obtient la distribution approchée suivante $u = -2(1 - \frac{13}{8S}) \ln(\mathcal{L}_{\text{Box}}) \sim \chi^2(6)$. La similarité entre deux patches centrés en pixels l et l' est alors calculée comme

$$\Delta_{\text{MNL}}(l, l') = \sum_{\tau} u\left[(l+\tau), (l'+\tau)\right], \tag{F.66}$$

où $\tau \in [-p, p]^2$ est un décalage 2D indiquant la position dans chaque patch de taille $P = (2p+1) \times (2p+1)$. Pour détecter les pixels similaires, les dissimilarités sont ensuite comparées à un seuil défini comme $\lambda = F^{-1}_{\chi^2(6P)}(1 - P_{fa})$ où $F^{-1}_{\chi^2(6P)}$ est l'inverse de fonction de répartition de $\chi^2(6P)$ et P_{fa} est la probabilité de fausse alarme choisie.

Calcul des poids. Une fois les pixels similaires choisis, les poids sont calculés en utilisant un noyau exponentiel

$$\omega_{\text{MNL}}(l, l') = \begin{cases} e^{-\frac{|\Delta_{\text{MNL}}(l,l') - c_{\text{MNL}}|}{\lambda_{\text{MNL}}}} & \text{if } l \neq l' \\ 1 & \text{if } l = l'. \end{cases} \tag{F.67}$$

où $c_{\text{MNL}} = \mathbb{E}\left[\Delta_{\text{MNL}}(l, l') | H_0\right]$. Enfin, l'estimateur pondéré du maximum de vraisemblance est donné par la moyenne pondérée

$$\widehat{\mathbf{\Sigma}}_{\text{NL}}(l) = \frac{\sum_{l'} \omega_{\text{MNL}}(l, l') \mathbf{k}' \mathbf{k}'^H}{\sum_{l'} \omega_{\text{MNL}}(l, l')}. \tag{F.68}$$

F.6.3 Résultats expérimentaux

Dans cette section, les résultats obtenus pour des données simulées et données réelles sont présentés. Les images simulées ont été générées à l'aide d'un champ de Markov suivant une distribution de Gibbs [56]. Ensuite, un comportement polarimétrique a été attribué aux différentes parties des images conçues. D'abord, un nombre aléatoire C de classes polarimétriques est choisi entre 3 et 5, les classes 1 à $C-1$ pour les diffuseurs distribués et la dernière classe pour les cibles (carrés de tailles variant entre 2×2 et 5×5 pixels). Après avoir attribué des signatures polarimétriques aux diffuseurs, un speckle gaussien est généré en fonction de celles-ci et les cibles sont ajoutées. Après le débruitage, les paramètres suivants ont été évalués : paramètres radiométriques σ (éléments diagonaux de la matrice de covariance (information de puissance)), paramètres de corrélation complexe ρ (dérivés des

Algorithm 4: Méthode M-NL

Initialisation : $\mathbb{W}, \mathbb{P}, \mathbb{S}, \lambda_{\text{MNL}}, c_{\text{MNL}}, \nu$
forall x, y **do**
  for s $\in \mathbb{S}$ *(taille de l'échelle)* **do**
    Pré-estimation avec l'Eq. (F.65)
forall x, y *(coordonnées du pixel l)* **do**
  for w $\in \mathbb{W}$ *(taille de la fenêtre de recherche)* **do**
    Calculer Δx et Δy
    $x' = x + \Delta x$
    $y' = y + \Delta y$; *coordonnées du pixel l'*
    for s $\in \mathbb{S}$ **do**
      for p $\in \mathbb{P}$ *(taille du patch)* **do**
        Calculer $\Delta_{\text{MNL}}(l, l')$ avec l'Eq. (F.66)
        if $\Delta_{\text{MNL}}(l, l') \leq \lambda_{\text{MNL}}[\text{p}]$ **then**
          Calculer $\omega(l, l')$ avec l'Eq. (F.67)
        else
          $\omega(l, l') \leftarrow 0$
  forall s, p, w **do**
    Calculer $\widehat{\mathbf{\Sigma}}_{\text{NL}}$ avec l'Eq. (F.68)
    Réduction du biais $\rightarrow \widehat{\mathbf{\Sigma}}_{\text{NLRB}}$
  retourner La meilleure estimation

trois termes complexes non diagonaux (corrélation de canaux)), paramètres de décomposition incohérente (Entropie (H), Anisotropie (A) et l'angle alpha moyen ($\overline{\alpha}$)-mécanisme de diffusion, signatures de polarisation (SP) et préservation des bords (EP)).

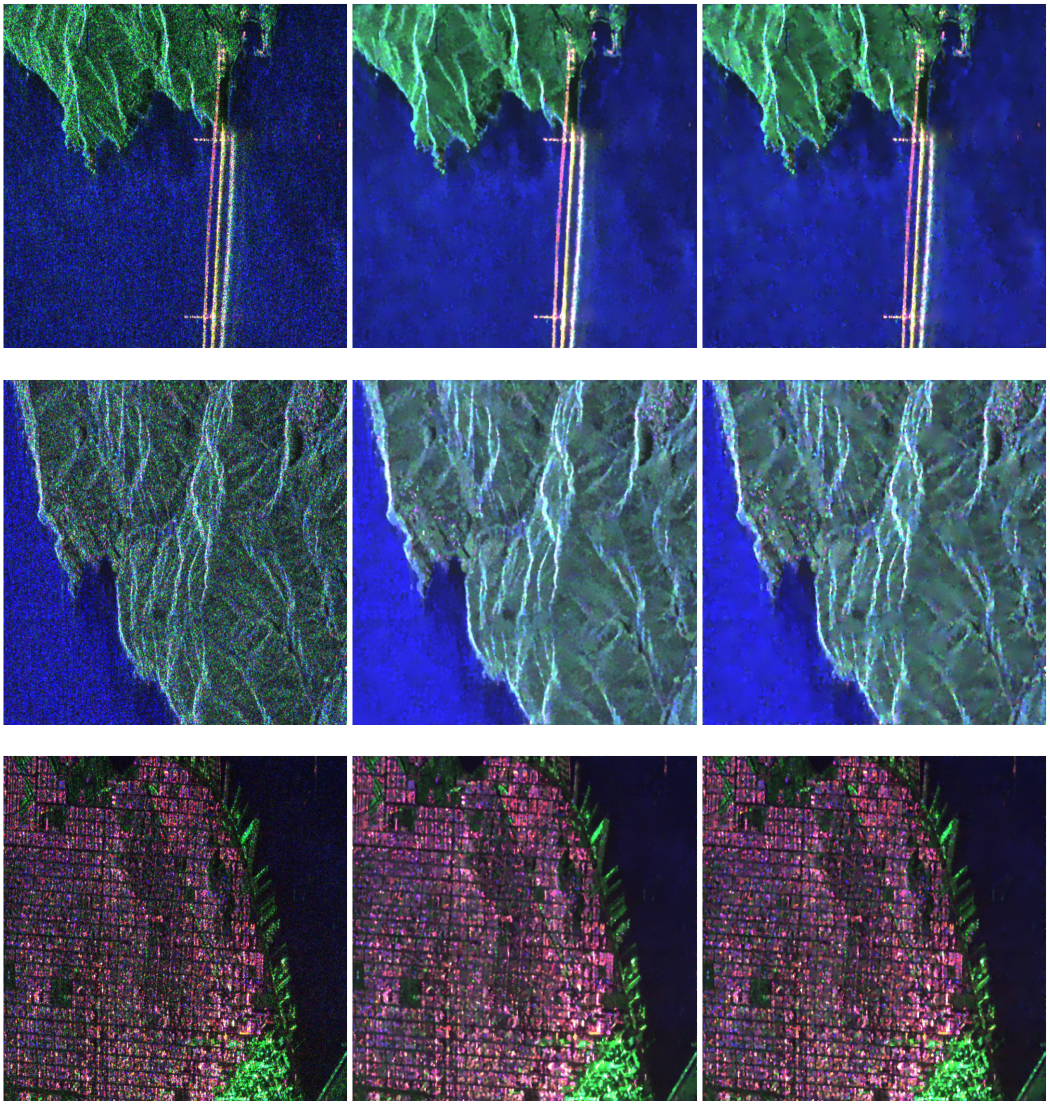
Pour effectuer la comparaison, nous avons simulé cent images PolSAR artificielles de 128× 128 pixels comme décrit précédemment. L'ensemble des paramètres utilisés dans les deux méthodes est le suivant : taille de la fenêtre $\{3^2, 5^2, \ldots, 25^2\}$, taille du patch $\{3^2, 5^2, \ldots, 11^2\}$ et échelle $\{0, 1, 2\}$. Comme le speckle est gaussien, nous avons choisi un ν suffisamment grand ($\nu = 100$) afin de conserver les informations sur la texture et d'assurer la convergence de la solution lors de l'étape de pré-estimation. Les valeurs de λ_{MNL} ont été calculées à l'aide de la formule correspondante de la section F.6.2.

Le tableau F.2 présente les résultats obtenus pour les paramètres d'évaluation définis ci-dessus. Les résultats ont été calculés sur l'ensemble des images PolSAR simulées et les valeurs finales sont comparées. On peut noter que M-NL offre meilleurs résultats que NL-SAR dans presque toutes les mesures sauf ρ et $\overline{\alpha}$.

Filters	σ	$\lvert\rho\rvert$	$\angle\rho$	H	A	$\overline{\alpha}$	PS	EP
NL-SAR	2.21	**7.47**	**11.96**	14.51	35.84	**10.51**	1.15	0.45
M-NL	**1.56**	9.10	14.47	**14.49**	**33.96**	10.97	**1.05**	**0.56**

TABLE F.2 – Résultats de filtrage pour les données simulées : toutes les mesures sauf EP (EP $\in [0, 1]$) sont des erreurs relatives absolues en %.

Les résultats pour les données réelles sont montrés sur la figure F.3. Trois parties différentes de la baie de San Francisco sont présentées de haut en bas, représentant différents scénarios d'images PolSAR telles que l'eau, la végétation et les zones urbaines. Dans ce cas, nous ne disposons d'aucune information sur la vérité terrain, nous ne pouvons donc analyser



Speckle NL-SAR M-NL

FIGURE F.3 – Données réelles : Baie de San Francisco - 512×512 images PolSAR. De gauche à droite : speckle, résultats obtenus avec NL-SAR et résultats obtenus avec M-NL.

les résultats que de manière visuelle. Premièrement, on peut noter que M -NL lisse mieux les zones homogènes, tout en préservant bien les bords dans les scénarios texturés. On peut également noter que, comme dans le cas des données simulées, M -NL donne des résultats avec un contraste plus élevé par rapport à NL-SAR.

F.7 Conclusion

Dans cette thèse, nous avons montré que les propriétés statistiques d'un M-estimateur peuvent être bien décrites avec une distribution de Wishart. De plus, il a été révélé que les paramètres EVD (valeurs propres / décomposition de vecteurs propres) des M-estimateurs se comportent de manière similaire aux paramètres EVD correspondants d'une matrice de Wishart. Les mêmes conclusions ont été tirées pour les principaux sous-espaces obtenus avec

un M-estimateur. De plus, les résultats ont été étendus aux détecteurs adaptatifs. Il a été démontré que les propriétés d'un détecteur robuste dans l'environnement CES, obtenues avec un M-estimateur au lieu de la SCM, se comportent comme les statistiques de détection gaussiennes correspondantes. Enfin, l'application des nouvelles propriétés statistiques des M-estimateurs au filtrage des image radar polarimétriques (PolSAR) a été présentée.

En reprenant les éléments précédents, on peut tirer une conclusion générale pour cette thèse : Il faut toujours privilégier les techniques robustes, car elles présentent de bonnes performances lorsqu'elles sont appliquées à des jeux de données réelles et peuvent être caractérisées avec précision grâce à des modèles très simples.

Suite à la conclusion précédente, les résultats de cette thèse peuvent être étendus à différentes problématiques et peuvent être appliqués à de nombreuses applications de traitement du signal et d'apprentissage automatique.

Bibliography

[1] Y. I. Abramovich and O. Besson. Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach—part 1: The over-sampled case. *IEEE Transactions on Signal Processing*, 61(23):5807–5818, 2013.

[2] Y. I. Abramovich and O. Besson. Fluctuating target detection in fluctuating k-distributed clutter. *IEEE Signal Processing Letters*, 22(10):1791–1795, 2015.

[3] M. Abramowitz and I. A. Stegun. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, volume 55. Courier Corporation, 1964.

[4] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, December 1974.

[5] A. Balleri, A. Nehorai, and J. Wang. Maximum likelihood estimation for Compound-Gaussian clutter with inverse gamma texture. *Aerospace and Electronic Systems, IEEE Transactions on*, 43(2):775–779, April 2007.

[6] O. Besson and Y. Abramovich. Adaptive detection in elliptically distributed noise and under-sampled scenario. *IEEE Signal Processing Letters*, 21(12):1531–1535, 2014.

[7] O. Besson and Y. I. Abramovich. Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach—part 2: The under-sampled case. *IEEE Transactions on Signal Processing*, 61(23):5819–5829, 2013.

[8] O. Besson, A. Coluccia, E. Chaumette, G. Ricci, and F. Vincent. Generalized likelihood ratio test for detection of gaussian rank-one signals in gaussian noise with unknown statistics. *IEEE Transactions on Signal Processing*, 65(4):1082–1092, 2016.

[9] R. Bhatia and C. Davis. A better bound on the variance. *The American Mathematical Monthly*, 107(4):353–357, 2000.

[10] S. Bidon, O. Besson, and J.-Y. Tourneret. Knowledge-aided stap in heterogeneous clutter using a hierarchical bayesian algorithm. *IEEE Transactions on Aerospace and Electronic Systems*, 47(3):1863–1879, 2011.

[11] J. Bien and R. J. Tibshirani. Sparse estimation of a covariance matrix. *Biometrika*, 98(4):807–820, 2011.

[12] J. Billingsley, A. Farina, F. Gini, M. Greco, and L. Verrazzani. Statistical analyses of measured radar ground clutter data. *IEEE Trans. on Aero. and Elec. Syst.*, 35(2):579 – 593, 1999.

[13] M. Bilodeau and D. Brenner. *Theory of Multivariate Statistics*. New York, NY. USA:Springer-Verlag, 1999.

[14] G. Boente. Asymptotic theory for robust principal components. *Journal of Multivariate Analysis*, 21(1):67 – 78, 1987.

[15] L. Bombrun, S. N. Anfinsen, and O. Harant. A complete coverage of log-cumulant space in terms of distributions for polarimetric sar data. In *5th International Workshop on Science and Applications of SAR Polarimetry and Polarimetric Interferometry (POLinSAR 2011)*, pages 1–8, 2011.

[16] C. Bouveyron and C. Brunet-Saumard. Model-based clustering of high-dimensional data: A review. *Computational Statistics & Data Analysis*, 71:52–78, 2014.

[17] G. Box. A general distribution theory for a class of likelihood criteria. *Biometrika*, 36, 1949.

[18] A. Breloy, G. Ginolhac, F. Pascal, and P. Forster. Clutter subspace estimation in low rank heterogeneous noise context. *IEEE Transactions on Signal Processing*, 63(9):2173–2182, 2015.

[19] A. Breloy, G. Ginolhac, A. Renaux, and F. Bouchard. Intrinsic Cramér-Rao Bounds for Scatter and Shape Matrices Estimation in CES Distributions. *submitted to IEEE Signal Processing Letters*, 2018.

[20] L. E. Brennan and F. Staudaher. Subclutter visibility demonstration. Technical report, 1992.

[21] A. Buades, B. Coll, and J. M. Morel. A Non-Local Algorithm for Image Denoising. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2 - Volume 02*, CVPR '05, pages 60–65, Washington, DC, USA, 2005. IEEE Computer Society.

[22] A. Buades, B. Coll, and J. M. Morel. A Review of Image Denoising Algorithms, with a New One. *Multiscale Modeling & Simulation*, 4(2):490–530, 2005.

[23] A. Buades, B. Coll, and J.-M. Morel. Nonlocal image and movie denoising. *International Journal of Computer Vision*, 76(2):123–139, Feb 2008.

[24] C.-I. Chang and S.-S. Chiang. Anomaly detection and classification for hyperspectral imagery. *IEEE transactions on geoscience and remote sensing*, 40(6):1314–1325, 2002.

[25] J. Chen, Y. Chen, W. An, Y. Cui, and J. Yang. Nonlocal Filtering for Polarimetric SAR Data: A Pretest Approach. *IEEE Transactions on Geoscience and Remote Sensing*, 49(5):1744–1754, 2011.

[26] Y. Chen, A. Wiesel, Y. C. Eldar, and A. O. Hero. Shrinkage algorithms for mmse covariance estimation. *IEEE Transactions on Signal Processing*, 58(10):5016–5029, 2010.

[27] Y. Chen, A. Wiesel, and A. O. Hero. Robust shrinkage estimation of high-dimensional covariance matrices. *Signal Processing, IEEE Transactions on*, 59(9):4097–4107, 2011.

[28] G. Chierchia, D. Cozzolino, G. Poggi, and L. Verdoliva. Sar image despeckling through convolutional neural networks. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 5438–5441. IEEE, 2017.

[29] D. Ciuonzo, V. Carotenuto, and A. De Maio. On multiple covariance equality testing with application to SAR change detection. *IEEE Transactions on Signal Processing*, 65(19):5078–5091, 2017.

[30] S. R. Cloude and E. Pottier. A review of target decomposition theorems in radar polarimetry. *IEEE transactions on geoscience and remote sensing*, 34(2):498–518, 1996.

[31] A. Combernoux, F. Pascal, G. Ginolhac, and M. Lesturgie. Theoretical performance of low rank adaptive filters in the large dimensional regime. *IEEE Transactions on Aerospace and Electronic Systems*, 2019.

[32] E. Conte and A. De Maio. Adaptive radar detection of distributed targets in non-Gaussian noise. In *RADAR*, Edinburgh, UK, 2002. IET.

[33] E. Conte, A. De Maio, and G. Ricci. Covariance matrix estimation for adaptive cfar detection in compound-gaussian clutter. *IEEE Transactions on Aerospace and Electronic Systems*, 38(2):415–426, 2002.

[34] E. Conte, A. De Maio, and G. Ricci. Recursive estimation of the covariance matrix of a compound-gaussian process and its application to adaptive cfar detection. *IEEE Transactions on signal processing*, 50(8):1908–1915, 2002.

[35] E. Conte and M. Longo. Characterisation of radar clutter as a spherically invariant random process. In *IEE Proceedings F (Communications, Radar and Signal Processing)*, volume 134, pages 191–197. IET, 1987.

[36] E. Conte and M. Longo. Characterization of radar clutter as a spherically invariant random process. *IEE Proceeding, Part. F*, 134(2):191–197, April 1987.

[37] E. Conte, M. Longo, and M. Lops. Modelling and simulation of non-rayleigh radar clutter. In *IEE Proceedings F (Radar and Signal Processing)*, volume 138, pages 121–130. IET, 1991.

[38] E. Conte, M. Lops, and G. Ricci. Asymptotically optimum radar detection in compound-gaussian clutter. *IEEE Transactions on Aerospace and Electronic Systems*, 31(2):617–625, 1995.

[39] E. Conte, M. Lops, and G. Ricci. Asymptotically optimum radar detection in Compound-Gaussian clutter. *IEEE Transactions on Aerospace and Electronic Systems*, 31(2):617–625, April 1995.

[40] R. Couillet, A. Kammoun, and F. Pascal. Second order statistics of robust estimators of scatter. application to glrt detection for elliptical signals. *Journal of Multivariate Analysis*, 143:249–274, 2016.

[41] R. Couillet, F. Pascal, and J. W. Silverstein. The Random Matrix Regime of Maronna's M-estimator with elliptically distributed samples. *Journal of Multivariate Analysis*, 139:56–78, July 2015.

[42] C. Croux and G. Haesbroeck. Principal component analysis based on robust estimators of the covariance or correlation matrix: Influence functions and efficiencies. *Biometrika*, 87:603–618, 2000.

[43] A. De Maio, A. Farina, and G. Foglia. Adaptive radar detection: A bayesian approach. In *2007 IEEE Radar Conference*, pages 624–629. IEEE, 2007.

[44] A. De Maio, A. Farina, and G. Foglia. Knowledge-aided bayesian radar detectors & their application to live data. *IEEE Transactions on Aerospace and Electronic Systems*, 46(1):170–183, 2010.

[45] A. De Maio and S. Greco. *Modern radar detection theory*. The Institution of Engineering and Technology, 2016.

[46] C. A. Deledalle, L. Denis, F. Tupin, A. Reigber, and M. Jäger. NL-SAR: A Unified Nonlocal Framework for Resolution-Preserving (Pol)(In)SAR denoising. *IEEE Transactions on Geoscience and Remote Sensing*, 53(4):2021–2038, 2015.

[47] C. A. Deledalle, F. Tupin, and L. Denis. Polarimetric SAR estimation based on nonlocal means. In *2010 IEEE International Geoscience and Remote Sensing Symposium*, pages 2515–2518, July 2010.

[48] X. Deng, C. López-Martínez, J. Chen, and P. Han. Statistical modeling of polarimetric sar data: A survey and challenges. *Remote Sensing*, 9(4):348, 2017.

[49] O. D'Hondt, S. Guillaso, and O. Hellwich. Iterative bilateral filtering of polarimetric sar data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(3):1628–1639, 2013.

[50] A. Doulgeris, S. N. Anfinsen, and T. Eltoft. Analysis of non-gaussian polsar data. In *2007 IEEE International Geoscience and Remote Sensing Symposium*, pages 160–163. IEEE, 2007.

[51] V. Duval, J.-F. Aujol, and Y. Gousseau. A bias-variance approach for the nonlocal means. *SIAM Journal on Imaging Sciences*, 4(2):760–788, 2011.

[52] E. Eberlein, U. Keller, et al. Hyperbolic distributions in finance. *Bernoulli*, 1(3):281–299, 1995.

[53] T. Eltoft, S. N. Anfinsen, and A. P. Doulgeris. A multitexture model for multilook polarimetric synthetic aperture radar data. *IEEE Transactions on Geoscience and Remote Sensing*, 52(5):2910–2919, 2013.

[54] K. W. Fang. *Symmetric Multivariate and Related Distributions: 0.* Chapman and Hall/CRC, 2017.

[55] S. Foucher, G. Farage, and G. Benie. Speckle filtering of polsar and polinsar images using trace-based partial differential equations. In *2006 IEEE International Symposium on Geoscience and Remote Sensing*, pages 2545–2548. IEEE, 2006.

[56] S. Foucher and C. López-Martínez. Analysis, Evaluation, and Comparison of Polarimetric SAR Speckle Filtering Techniques. *IEEE Transactions on Image Processing*, 23(4):1751–1764, 2014.

[57] G. Frahm. *Generalized elliptical distributions: theory and applications.* PhD thesis, Universität zu Köln, 2004.

[58] G. Frahm, M. Junker, and R. Schmidt. Estimating the tail-dependence coefficient: properties and pitfalls. *Insurance: mathematics and Economics*, 37(1):80–100, 2005.

[59] A. C. Frery, H.-J. Muller, C. d. C. F. Yanasse, and S. J. S. Sant'Anna. A model for extremely heterogeneous clutter. *IEEE transactions on geoscience and remote sensing*, 35(3):648–659, 1997.

[60] J. Frontera-Pons, M. A. Veganzones, S. Velasco-Forero, F. Pascal, J. P. Ovarlez, and J. Chanussot. Robust anomaly detection in hyperspectral imaging. In *2014 IEEE Geoscience and Remote Sensing Symposium*, pages 4604–4607, July 2014.

[61] V. S. Frost, J. A. Stiles, K. S. Shanmugan, and J. C. Holtzman. A model for radar images and its application to adaptive digital filtering of multiplicative noise. *IEEE Transactions on pattern analysis and machine intelligence*, (2):157–166, 1982.

[62] F. Gini. Sub-optimum coherent radar detection in a mixture of K-distributed and Gaussian clutter. *IEEE Proceedings - Radar, Sonar and Navigation*, 144(1):39–48, February 1997.

[63] F. Gini and M. Greco. Suboptimum approach to adaptive coherent radar detection in compound-Gaussian clutter. *IEEE Transactions on Aerospace and Electronic Systems*, 35(3):1095–1104, 1999.

[64] F. Gini and M. Greco. Covariance matrix estimation for cfar detection in correlated heavy tailed clutter. *Signal Processing*, 82(12):1847–1859, 2002.

[65] F. Gini, M. Greco, M. Diani, and L. Verrazzani. Performance analysis of two adaptive radar detectors against non-gaussian real sea clutter data. *IEEE Transactions on Aerospace and Electronic Systems*, 36(4):1429–1439, 2000.

[66] F. Gini and M. S. Greco. Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter. *Signal Processing, special section on SP with Heavy Tailed Distributions*, 82(12):1847–1859, December 2002.

[67] F. Gini, M. S. Greco, M. Diani, and L. Verrazzani. Performance analysis of two adaptive radar detectors against non-Gaussian real sea clutter data. *Aerospace and Electronic Systems, IEEE Transactions on*, 36(4):1429–1439, October 2000.

[68] F. Gini, M. V. Greco, A. Farina, and P. Lombardo. Optimum and mismatched detection against k-distributed plus gaussian clutter. *IEEE Transactions on Aerospace and Electronic Systems*, 34(3):860–876, 1998.

[69] 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*, 52(4), 2016.

[70] J. Goldman. Detection in the presence of spherically symmetric random vectors. *IEEE Transactions on Information Theory*, 22(1):52–59, 1976.

[71] N. R. Goodman. Statistical analysis based on a certain multivariate complex gaussian distribution (an introduction). *The Annals of mathematical statistics*, 34(1):152–177, 1963.

[72] N. R. Goodman. Statistical analysis based on a certain multivariate complex gaussian distribution (an introduction). *Ann. Math. Statist.*, 34(1):152–177, 03 1963.

[73] B. Goossens, H. Q. Luong, A. Pizurica, and W. Philips. An improved non-local denoising algorithm. 2008.

[74] S. Goze and A. Lopes. A mmse speckle filter for full resolution sar polarimetric data. *Journal of electromagnetic waves and applications*, 7(5):717–737, 1993.

[75] P. J. Green. Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *Journal of the Royal Statistical Society. Series B (Methodological)*, 46(2):149–192, 1984.

[76] J. Gu, J. Yang, H. Zhang, Y. Peng, C. Wang, and H. Zhang. Speckle filtering in polarimetric sar data based on the subspace decomposition. *IEEE transactions on geoscience and remote sensing*, 42(8):1635–1641, 2004.

[77] A. K. Gupta and D. K. Nagar. *Matrix Variate Distributions*. Chapman & Hall/CRC, 2000.

[78] A. S. Hadi. Identifying multiple outliers in multivariate data. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 761–771, 1992.

[79] A. Haimovich. Asymptotic distribution of the conditional signal-to-noise ratio in an eigenanalysis-based adaptive array. *IEEE Transactions on Aerospace and Electronic Systems*, 33(3):988–997, 1997.

[80] F. R. Hampel. A general qualitative definition of robustness. *The Annals of Mathematical Statistics*, pages 1887–1896, 1971.

[81] F. R. Hampel. The influence curve and its role in robust estimation. *Journal of the american statistical association*, 69(346):383–393, 1974.

[82] P. J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, January 1964.

[83] P. J. Huber et al. The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 221–233. University of California Press, 1967.

[84] E. Jakeman and P. Pusey. Significance of k distributions in scattering experiments. *Physical Review Letters*, 40(9):546, 1978.

[85] A. Kammoun, R. Couillet, F. Pascal, and M. Alouini. Optimal Design of the Adaptive Normalized Matched Filter Detector Using Regularized Tyler Estimators. *IEEE Transactions on Aerospace and Electronic Systems*, 54(2):755–769, April 2018.

[86] A. Kammoun, R. Couillet, F. Pascal, and M.-S. Alouini. Optimal design of the adaptive normalized matched filter detector using regularized tyler estimators. *IEEE Transactions on Aerospace and Electronic Systems*, 54(2):755–769, 2017.

[87] D. Kelker. Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā: The Indian Journal of Statistics, Series A*, 32(4):419–430, December 1970.

[88] E. Kelly. Performance of an adaptive detection algorithm; rejection of unwanted signals. *IEEE Transactions on Aerospace and Electronic Systems*, 25(2):122–133, 1989.

[89] E. J. Kelly. Adaptive detection in non-stationary interference. part 1 and part 2. Technical report, MASSACHUSETTS INST OF TECH LEXINGTON LINCOLN LAB, 1985.

[90] E. J. Kelly. An adaptive detection algorithm. *IEEE Transactions on Aerospace and Electronic Systems*, 22(2):115–127, March 1986.

[91] E. J. Kelly. Adaptive detection in non-stationary interference, part iii. *Massachusetts Institute of Technology, Lincoln Laboratory, Lexington, MA, Tech. Rep*, 761, 1987.

[92] E. J. Kelly and K. M. Forsythe. Adaptive detection and parameter estimation for multidimensional signal models. Technical report, MASSACHUSETTS INST OF TECH LEXINGTON LINCOLN LAB, 1989.

[93] J. T. Kent. Data analysis for shapes and images. *Journal of statistical planning and inference*, 57(2):181–193, 1997.

[94] J. T. Kent and D. E. Tyler. Maximum likelihood estimation for the wrapped Cauchy distribution. *Journal of Applied Statistics*, 15(2):247–254, 1988.

[95] I. P. Kirsteins and D. W. Tufts. Adaptive detection using low rank approximation to a data matrix. *IEEE Transactions on Aerospace and Electronic Systems*, 30(1):55–67, 1994.

[96] T. Kollo and H. Neudecker. Asymptotics of eigenvalues and unit-length eigenvectors of sample variance and correlation matrices. *Journal of Multivariate Analysis*, 47(2):283 – 300, 1993.

[97] S. Kraut and L. L. Scharf. The CFAR adaptive subspace detector is a scale-invariant GLRT. *IEEE Transactions on Signal Processing*, 47(9):2538–2541, September 1999.

[98] S. Kraut, L. L. Scharf, and R. W. Butler. The adaptive coherence estimator: a uniformly most-powerful-invariant adaptive detection statistic. *IEEE Transactions on Signal Processing*, 53(2):427–438, 2005.

[99] S. Kraut, L. L. Scharf, and L. T. McWhorter. Adaptive subspace detectors. *IEEE Transactions on Signal Processing*, 49(1):1–16, January 2001.

[100] H. Krim and P. Forster. Projections on unstructured subspaces. *IEEE Transactions on Signal Processing*, 44(10):2634–2637, Oct 1996.

[101] H. Krim, P. Forster, and J. G. Proakis. Operator approach to performance analysis of root-music and root-min-norm. *IEEE Transactions on Signal Processing*, 40(7):1687–1696, Jul 1992.

[102] O. Ledoit and M. Wolf. Honey, I shrunk the sample covariance matrix. 2003.

[103] O. Ledoit and M. Wolf. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10(5):603–621, 2003.

[104] O. Ledoit and M. Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivar. Anal.*, 88(2):365–411, Feb. 2004.

[105] J. Lee and E. Pottier. *Polarimetric Radar Imaging, From Basics to Applications*. Boca Raton: CRC Press, 2009.

[106] J.-S. Lee. Speckle analysis and smoothing of synthetic aperture radar images. *Computer graphics and image processing*, 17(1):24–32, 1981.

[107] J.-S. Lee. A simple speckle smoothing algorithm for synthetic aperture radar images. *IEEE Transactions on Systems, Man, and Cybernetics*, (1):85–89, 1983.

[108] J.-S. Lee, M. R. Grunes, and G. De Grandi. Polarimetric sar speckle filtering and its implication for classification. *IEEE Transactions on Geoscience and remote sensing*, 37(5):2363–2373, 1999.

[109] J.-S. Lee, M. R. Grunes, and R. Kwok. Classification of multi-look polarimetric sar imagery based on complex wishart distribution. *International Journal of Remote Sensing*, 15(11):2299–2311, 1994.

[110] J.-S. Lee, M. R. Grunes, and S. A. Mango. Speckle reduction in multipolarization, multifrequency sar imagery. *IEEE Transactions on Geoscience and remote sensing*, 29(4):535–544, 1991.

[111] J.-S. Lee, J.-H. Wen, T. L. Ainsworth, K.-S. Chen, and A. J. Chen. Improved sigma filter for speckle filtering of sar imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 47(1):202–213, 2008.

[112] H. Lim, A. Swartz, H. Yueh, J. A. Kong, R. Shin, and J. Van Zyl. Classification of earth terrain using polarimetric synthetic aperture radar images. *Journal of Geophysical Research: Solid Earth*, 94(B6):7049–7057, 1989.

[113] R. D. Maesschalck, D. Jouan-Rimbaud, and D. Massart. The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, 50(1):1 – 18, 2000.

[114] J. R. Magnus. On differentiating eigenvalues and eigenvectors. *Econometric Theory*, 1(2):179–191, 1985.

[115] P. C. Mahalanobis. On the generalized distance in statistics. *Proceedings of the National Institute of Sciences (Calcutta)*, 2:49–55, 1936.

[116] M. Mahot. *Estimation robuste de la matrice de covariance en traitement du signal*. PhD thesis, École normale supérieure de Cachan-ENS Cachan, 2012.

[117] M. Mahot, F. Pascal, P. Forster, and J.-P. Ovarlez. Asymptotic properties of robust complex covariance matrix estimates. *IEEE Transactions on Signal Processing*, 61(13):3348–3356, July 2013.

[118] A. D. Maio. Rao test for adaptive detection in gaussian interference with unknown covariance matrix. *IEEE Transactions on Signal Processing*, 55(7):3577–3584, July 2007.

[119] R. A. Maronna. Robust M-estimators of multivariate location and scatter. *Annals of Statistics*, 4(1):51–67, January 1976.

[120] R. A. Maronna, R. D. Martin, V. J. Yohai, and M. Salibián-Barrera. *Robust statistics: theory and methods (with R)*. Wiley, 2018.

[121] G. Mercier, L. Bouchemakh, and Y. Smara. The use of multidimensional copulas to describe amplitude distribution of polarimetric sar data. In *2007 IEEE International Geoscience and Remote Sensing Symposium*, pages 2236–2239. IEEE, 2007.

[122] G. Moser, J. Zerubia, and S. B. Serpico. Dictionary-based stochastic expectation-maximization for sar amplitude probability density function estimation. *IEEE Transactions on Geoscience and Remote Sensing*, 44(1):188–200, 2005.

[123] R. J. Muirhead. Aspects of multivariate statistical analysis. *JOHN WILEY & SONS, INC., 605 THIRD AVE., NEW YORK, NY 10158, USA, 1982, 656*, 1982.

[124] R. J. Muirhead. *Aspects of multivariate statistical theory*, volume 197. John Wiley & Sons, 2009.

[125] X. Nie, H. Qiao, and B. Zhang. A variational model for polsar data speckle reduction based on the wishart distribution. *IEEE Transactions on Image Processing*, 24(4):1209–1222, 2015.

[126] X. Nie, H. Qiao, B. Zhang, and X. Huang. A nonlocal tv-based variational method for polsar data speckle reduction. *IEEE Transactions on Image Processing*, 25(6):2620–2634, 2016.

[127] E. Ollila and E. Raninen. Optimal shrinkage covariance matrix estimation under random sampling from elliptical distributions. *arXiv preprint arXiv:1808.10188*, 2018.

[128] E. Ollila and D. E. Tyler. Distribution-free detection under complex elliptically symmetric clutter distribution. In *2012 IEEE 7th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pages 413–416. IEEE, 2012.

[129] E. Ollila and D. E. Tyler. Regularized m-estimators of scatter matrix. *Signal Processing, IEEE Transactions on*, 62(22):6059–6070, Nov. 2014.

[130] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor. Complex elliptically symmetric distributions: Survey, new results and applications. *Signal Processing, IEEE Transactions on*, 60(11):5597–5625, November 2012.

[131] J.-P. Ovarlez, F. Pascal, and A. Breloy. Asymptotic detection performance analysis of the robust adaptive normalized matched filter. In *2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 137–140. IEEE, 2015.

[132] J.-P. Ovarlez, F. Pascal, and P. Forster. Covariance matrix estimation in SIRV and elliptical processes and their applications in radar detection, 2015.

[133] P. Parker and A. Swindlehurst. Space-time autoregressive filtering for matched subspace stap. *IEEE Transactions on Aerospace and Electronic Systems*, 39(2):510–520, 2003.

[134] F. Pascal, L. Bombrun, J.-Y. Tourneret, and Y. Berthoumieu. Parameter estimation for multivariate generalized Gaussian distributions. *Signal Processing, IEEE Transactions on*, 61(23):5960–5971, December 2013.

[135] 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. *Signal Processing, IEEE Transactions on*, 56(1):34–48, January 2008.

[136] F. Pascal, Y. Chitour, and Y. Quek. Generalized Robust Shrinkage Estimator and Its Application to STAP Detection Problem. *IEEE Transactions on Signal Processing*, 62(21):5640–5651, Nov 2014.

[137] F. Pascal, P. Forster, J.-P. Ovarlez, and P. Larzabal. Performance analysis of covariance matrix estimates in impulsive noise. *IEEE Transactions on Signal Processing*, 56(6):2206–2217, June 2008.

[138] F. Pascal and J.-P. Ovarlez. Asymptotic detection performance of the robust ANMF. In *2015 23rd European Signal Processing Conference (EUSIPCO)*, pages 524–528. IEEE, 2015.

[139] F. Pascal and J.-P. Ovarlez. Asymptotic Properties of the Robust ANMF. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-15*, pages 2594–2598, Brisbane, Australia, April 2015.

[140] F. Pascal and J.-P. Ovarlez. Asymptotic Properties of the Robust ANMF. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-15*, Brisbane, Australia, April 2015.

[141] F. Pascal, J. P. Ovarlez, P. Forster, and P. Larzabal. Constant false alarm rate detection in spherically invariant random processes. pages 2143–2146, Sept 2004.

[142] F. Pascal, J.-P. Ovarlez, P. Forster, and P. Larzabal. On a sirv-cfar detector with radar experimentations in impulsive clutter. In *2006 14th European Signal Processing Conference (EUSIPCO)*, pages 1–5. IEEE, 2006.

[143] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on pattern analysis and machine intelligence*, 12(7):629–639, 1990.

[144] K. B. Petersen, M. S. Pedersen, et al. The matrix cookbook. *Technical University of Denmark*, 7:15, 2008.

[145] P. Pudil, J. Novovičová, and J. Kittler. Floating search methods in feature selection. *Pattern recognition letters*, 15(11):1119–1125, 1994.

[146] R. Raghavan. Statistical interpretation of a data adaptive clutter subspace estimation algorithm. *IEEE Transactions on Aerospace and Electronic Systems*, 48(2):1370–1384, 2012.

[147] M. Rangaswamy. Statistical analysis of the nonhomogeneity detector for non-Gaussian interference backgrounds. *Signal Processing, IEEE Transactions on*, 53(6):2101–2111, June 2005.

[148] M. Rangaswamy, F. C. Lin, and K. R. Gerlach. Robust adaptive signal processing methods for heterogeneous radar clutter scenarios. *Signal Processing*, 84(9):1653–1665, 2004.

[149] I. S. Reed, J. D. Mallett, and L. E. Brennan. Rapid convergence rate in adaptive arrays. *IEEE Transactions on Aerospace and Electronic Systems*, (6):853–863, 1974.

[150] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg. A cfar adaptive matched filter detector. *IEEE Transactions on Aerospace and Electronic Systems*, 28(1):208–216, Jan 1992.

[151] J. R. Roman, M. Rangaswamy, D. W. Davis, Q. Zhang, B. Himed, and J. H. Michels. Parametric adaptive matched filter for airborne radar applications. *IEEE Transactions on Aerospace and Electronic Systems*, 36(2):677–692, 2000.

[152] P. J. Rousseeuw and A. M. Leroy. *Robust regression and outlier detection*. John Wiley, New York, 1987.

[153] P. J. Rousseeuw and B. C. Van Zomeren. Unmasking multivariate outliers and leverage points. *Journal of the American Statistical association*, 85(411):633–639, 1990.

[154] J. Salmon and Y. Strozecki. From patches to pixels in non-local methods: Weighted-average reprojection. In *2010 IEEE International Conference on Image Processing*, pages 1929–1932. IEEE, 2010.

[155] K. J. Sangston, F. Gini, and M. S. Greco. Coherent radar target detection in heavy-tailed compound-Gaussian clutter. *IEEE Transactions on Aerospace and Electronic Systems*, 48(1):64–77, 2012.

[156] K. J. Sangston, F. Gini, M. V. Greco, and A. Farina. Structures for radar detection in compound Gaussian clutter. *IEEE Transactions on Aerospace and Electronic Systems*, 35(2):445–458, 1999.

[157] L. L. Scharf. *Statistical signal processing*, volume 98. Addison-Wesley Reading, MA, 1991.

[158] J. Schou and H. Skriver. Restoration of polarimetric sar images using simulated annealing. *IEEE Transactions on Geoscience and Remote Sensing*, 39(9):2005–2016, 2001.

[159] P. J. Schreier and L. L. Scharf. *Statistical signal processing of complex-valued data: the theory of improper and noncircular signals*. Cambridge university press, 2010.

[160] S. T. Smith. Covariance, subspace, and intrinsic Cramér-Rao bounds. *IEEE Transactions on Signal Processing*, 53(5):1610–1630, 2005.

[161] P. Stoica and Y. Selen. Model-order selection: a review of information criterion rules. *IEEE Signal Processing Magazine*, 21(4):36–47, July 2004.

[162] Y. Sun, P. Babu, and D. P. Palomar. Regularized Tyler's Scatter Estimator: Existence, Uniqueness and Algorithms. *Signal Processing, IEEE Transactions on*, 62(19):5143–5156, Oct. 2014.

[163] Y. Sun, P. Babu, and D. P. Palomar. Regularized tyler's scatter estimator: Existence, uniqueness, and algorithms. *IEEE Transactions on Signal Processing*, 62(19):5143–5156, 2014.

[164] E. Terreaux, J.-P. Ovarlez, and F. Pascal. Robust model order selection in large dimensional elliptically symmetric noise. *arXiv:1710.06735*, 2018.

[165] C. Tison, J.-M. Nicolas, F. Tupin, and H. Maître. A new statistical model for markovian classification of urban areas in high-resolution sar images. *IEEE transactions on geoscience and remote sensing*, 42(10):2046–2057, 2004.

[166] J. W. Tukey. A survey of sampling from contaminated distributions. *Contributions to probability and statistics*, pages 448–485, 1960.

[167] J. W. Tukey. The future of data analysis. *The annals of mathematical statistics*, 33(1):1–67, 1962.

[168] D. Tyler. Asymptotic inference for eigenvectors. *The Annals of Statistics*, pages 725–736, 1981.

[169] D. E. Tyler. Radial estimates and the test for sphericity. *Biometrika*, 69(2):429, 1982.

[170] D. E. Tyler. A distribution-free M-estimator of multivariate scatter. *The Annals of Statistics*, 15(1):234–251, 1987.

[171] D. E. Tyler and M. Yi. Lassoing Eigenvalues. *arXiv:1805.08300v1*, 2018.

[172] P. Vallet, G. Ginolhac, F. Pascal, and P. Forster. An improved low rank detector in the high dimensional regime. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5336–5340. IEEE, 2019.

[173] H. L. Van Trees. *Detection, estimation, and modulation theory, part I: detection, estimation, and linear modulation theory*. John Wiley & Sons, 2004.

[174] 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*, 48(4):1809–1826, 2009.

[175] G. Vasile, E. Trouvé, J.-S. Lee, and V. Buzuloiu. Intensity-driven adaptive-neighborhood technique for polarimetric and interferometric sar parameters estimation. *IEEE Transactions on Geoscience and Remote Sensing*, 44(6):1609–1621, 2006.

[176] J. Vinogradova, R. Couillet, and W. Hachem. Estimation of toeplitz covariance matrices in large dimensional regime with application to source detection. *IEEE Transactions on Signal Processing*, 63(18):4903–4913, 2015.

[177] J. Ward. Space-Time Adaptive Processing for Airborne Radar MIT Lincoln Lab., Lexington. Technical report, MA, Tech. Rep. 1015, 1994.

[178] K. Ward. Compound representation of high resolution sea clutter. *Electronics letters*, 17(16):561–563, 1981.

[179] S. Watts. Radar detection prediction in sea clutter using the compound k-distribution model. *IEE Proceeding, Part. F*, 132(7):613–620, December 1985.

[180] M. Wax and I. Ziskind. Detection of the number of coherent signals by the mdl principle. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(8):1190–1196, Aug 1989.

[181] K. Q. Weinberger, J. Blitzer, and L. K. Saul. Distance metric learning for large margin nearest neighbor classification. In *Advances in neural information processing systems*, pages 1473–1480, 2006.

[182] A. Wiesel. Unified framework to regularized covariance estimation in scaled Gaussian models. *Signal Processing, IEEE Transactions on*, 60(1):29–38, 2012.

[183] R. Wolke and H. Schwetlick. Iteratively reweighted least squares: Algorithms, convergence analysis, and numerical comparisons. *SIAM Journal on Scientific and Statistical Computing*, 9(5):907–921, 1988.

[184] S. Xiang, F. Nie, and C. Zhang. Learning a Mahalanobis distance metric for data clustering and classification. *Pattern Recognition*, 41(12):3600 – 3612, 2008.

[185] B. Xu, Y. Cui, B. Zuo, J. Yang, and J. Song. Polarimetric sar image filtering based on patch ordering and simultaneous sparse coding. *IEEE Transactions on Geoscience and Remote Sensing*, 54(7):4079–4093, 2016.

[186] X. Yang, L. Denis, F. Tupin, and W. Yang. Sar image despeckling through convolutional neural networks. In *JURSE*, 2019.

[187] K. Yao. A representation theorem and its applications to spherically invariant random processes. *Information Theory, IEEE Transactions on*, 19(5):600–608, September 1973.

[188] S. Yueh, J. A. Kong, J. Jao, R. Shin, and L. Novak. K-distribution and polarimetric terrain radar clutter. *Journal of Electromagnetic Waves and Applications*, 3(8):747–768, 1989.

[189] A. Zanella, M. Chiani, and M. Z. Win. On the marginal distribution of the eigenvalues of Wishart matrices. *IEEE Transactions on Communications*, 57(4):1050–1060, April 2009.

[190] A. Zanella, M. Chiani, and M. Z. Win. On the marginal distribution of the eigenvalues of wishart matrices. *IEEE Transactions on Communications*, 57(4):1050–1060, April 2009.

[191] J. Zhao and Q. Jiang. Probabilistic PCA for t distributions. *Neurocomputing*, 69(16-18):2217–2226, 2006.

[192] H. Zhong, J. Zhang, and G. Liu. Robust polarimetric sar despeckling based on non-local means and distributed lee filter. *IEEE Transactions on Geoscience and Remote Sensing*, 52(7):4198–4210, 2013.

[193] J. J. Zyl, H. A. Zebker, and C. Elachi. Imaging radar polarization signatures: Theory and observation. *Radio Science*, 22(4):529–543, 1987.

Publications

International journals

[J1] **G. Drašković** and F. Pascal, "New Insights Into the Statistical Properties of M-Estimators," *IEEE Transactions on Signal Processing*, vol. 66, no. 16, pp. 4253-4263, 2018.

[J2] **G. Drašković**, A. Breloy, and F. Pascal, "On the asymptotics of Maronna's robust PCA." *IEEE Transactions on Signal Processing*, vol. 67, no. 19, pp. 4964-4975, 2019.

[J3] **G. Drašković**, A. Breloy, and F. Pascal, "On the performance of robust plug-in detectors using M-estimators," *Signal Processing*, vol. 167, pp. 107282, 2020.

[J4] **G. Drašković**, F. Pascal, and F.Tupin, "M-NL: Robust NL-means approach for PolSAR image denoising," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 6, pp. 997–1001, 2019.

International conferences

[C1] **G. Drašković** and F. Pascal, "New properties for Tyler's covariance matrix estimator," *2016 50th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, 2016, pp. 820-824.

[C2] **G. Drašković**, F. Pascal, A. Breloy, and J.-Y. Tourneret, "New asymptotic properties for the robust ANMF," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-17*, New Orleans, USA, March 2017, pp. 3429-3433.

[C3] **G. Drašković** and F. Pascal, "New insights into the statistical properties of M-estimators and application to signal detection," *International Conference on Robust Statistics (ICORS)*, Leuven, Belgium, July 2018.

[C4] T. Diskin, **G. Drašković**, F. Pascal, and A. Wiesel,"Deep robust regression," *IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Curaçao, Dutch Antilles, December 2017, pp. 1-5.

[C5] S. Basiri, E. Ollila, **G. Drašković**, and F. Pascal, "Fusing eigenvalues," *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, United Kingdom, 2019, pp. 4968-4972.

[C6] V. Roizman, **G. Drašković**, and F. Pascal, "A new clustering algorithm for PolSAR image segmentation," accepted in *IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2019.*

National conferences

[C7] **G. Drašković**, F. Pascal, A. Breloy et J.-Y. Tourneret, "Nouvelles propriétés asymptotiques de détecteurs robustes," *colloque GRETSI sur le traitement du signal et des images*, Juan-les-Pins, France, Septembre 2017.

[C8] **G. Drašković**, F. Pascal et F. Tupin, "Débruitage des images polarimétriques avec le M-NL robuste," *colloque GRETSI sur le traitement du signal et des images*, Lille, France, Août 2019.

[C9] **G. Draškovi**ć, A. Breloy et F. Pascal, "Caractérisations asymptotiques pour les composantes principales des M-estimateurs," *colloque GRETSI sur le traitement du signal et des images*, Lille, Août 2019.

Seminars

"New insights into the statistical properties of M-estimators with application to signal detection and PolSAR image denoising"

[S1] Facultad de Ciencias, Montevideo, Uruguay, December 2018

[S2] Inria/CEA Parietal Team, France, February 2019

[S3] Aalto University, Helsinki, Finland, August 2019

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Titre : Statistiques des estimateurs robustes pour le traitement du signal et des images
Mots-clés : Estimation robuste, distributions CES, Loi de Wishart, détection du signal, images PolSAR.

Résumé : Un des défis majeurs en traitement radar consiste à identifier une cible cachée dans un environnement bruité. Pour ce faire, il est nécessaire de caractériser finement les propriétés statistiques du bruit, en particulier sa matrice de covariance. Sous l'hypothèse gaussienne, cette dernière est estimée par la matrice de covariance empirique (SCM) dont le comportement est parfaitement connu. Cependant, dans de nombreuses applications actuelles, tels les systèmes radar modernes à haute résolution par exemple, les données collectées sont de nature hétérogène, et ne peuvent être proprement décrites par un processus gaussien. Pour pallier ce problème, les distributions symétriques elliptiques complexes, caractérisant mieux ces phénomènes physiques complexes, ont été proposées. Dans ce cas, les performances de la SCM sont très médiocres et les M-estimateurs apparaissent comme une bonne alternative, principalement en raison de leur flexibilité par rapport au modèle statistique et de leur robustesse aux données aberrantes et/ou aux données manquantes. Cependant, le comportement de tels estimateurs reste encore mal compris. Dans ce contexte, les contributions de cette thèse sont multiples.

D'abord, une approche originale pour analyser les propriétés statistiques des M-estimateurs est proposée, révélant que les propriétés statistiques des M-estimateurs peuvent être bien approximées par une distribution de Wishart. Grâce à ces résultats, nous analysons la décomposition de la matrice de covariance en éléments propres. Selon l'application, la matrice de covariance peut posséder une structure particulière impliquant valeurs propres multiples contenant les informations d'intérêt. Nous abordons ainsi divers scénarios rencontrés dans la pratique et proposons des procédures robustes basées sur des M-estimateurs. De plus, nous étudions le problème de la détection robuste du signal. Les propriétés statistiques de diverses statistiques de détection adaptative construites avec des M-estimateurs sont analysées. Enfin, la dernière partie de ces travaux est consacrée au traitement des images radar à synthèse d'ouverture polarimétriques (PolSAR). En imagerie PolSAR, un effet particulier appelé speckle dégrade considérablement la qualité de l'image. Dans cette thèse, nous montrons comment les nouvelles propriétés statistiques des M-estimateurs peuvent être exploitées afin de construire de nouvelles techniques pour la réduction du speckle.

Title: Robust estimation analysis for signal and image processing
Keywords: Robust estimation, CES distributions, Wishart distribution, signal detection, PolSAR imaging.

Abstract: One of the main challenges in radar processing is to identify a target hidden in a disturbance environment. To this end, the noise statistical properties, especially the ones of the disturbance covariance matrix, need to be determined. Under the Gaussian assumption, the latter is estimated by the sample covariance matrix (SCM) whose behavior is perfectly known. However, in many applications, such as, for instance, the modern high resolution radar systems, collected data exhibit a heterogeneous nature that cannot be adequately described by a Gaussian process. To overcome this problem, Complex Elliptically Symmetric distributions have been proposed since they can correctly model these data behavior. In this case, the SCM performs very poorly and M-estimators appear as a good alternative, mainly due to their flexibility to the statistical model and their robustness to outliers and/or missing data. However, the behavior of such estimators still remains unclear and not well understood. In this context, the contributions of this thesis are multiple.

First, an original approach to analyze the statistical properties of M-estimators is proposed, revealing that the statistical properties of M-estimators can be approximately well-described by a Wishart distribution. Thanks to these results, we go further and analyze the eigendecomposition of the covariance matrix. Depending on the application, the covariance matrix can exhibit a particular structure involving multiple eigenvalues containing the information of interest. We thus address various scenarios met in practice and propose robust procedures based on M-estimators. Furthermore, we study the robust signal detection problem. The statistical properties of various adaptive detection statistics built with M-estimators are analyzed. Finally, the last part deals with polarimetric synthetic aperture radar (PolSAR) image processing. In PolSAR imaging, a particular effect called speckle significantly degrades the image quality. In this thesis, we demonstrate how the new statistical properties of M-estimators can be exploited in order to build new despeckling techniques.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France