



# Robust target detection for Hyperspectral Imaging.

Joana Maria Frontera Pons

## ► To cite this version:

Joana Maria Frontera Pons. Robust target detection for Hyperspectral Imaging.. Other. Supélec, 2014. English. <NNT : 2014SUPL0024>. <tel-01165283>

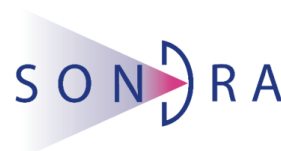
**HAL Id: tel-01165283**

**<https://tel.archives-ouvertes.fr/tel-01165283>**

Submitted on 18 Jun 2015

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

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



N° d'ordre : 2014-24-TH

**SUPELEC**

**ECOLE DOCTORALE STITS**

*« Sciences et Technologies de l'Information des Télécommunications et des Systèmes »*

**THÈSE DE DOCTORAT**

**DOMAINE : STIC**

**Spécialité : Traitement du signal**

**Soutenue le 10 décembre 2014**

**par :**

**Joana FRONTERA-PONS**

# **Robust target detection for Hyperspectral Imaging** **(Détection robuste de cibles en imagerie Hyperspectrale)**

**Directeur de thèse :** Jean-Philippe OVARLEZ

Maître de Recherche, ONERA

**Composition du jury :**

**Président du jury :** Jocelyn CHANUSSOT

**Rapporteurs :** Grégoire MERCIER

Cédric RICHARD

**Examineurs :** Jesús ANGULO

José M. BIOUCAS-DIAS

Sylvie MARCOS

Frédéric PASCAL (Encadrant)

Professeur des Universités, INP Grenoble

Professeur des Universités, Institut Telecom-Bretagne

Professeur des Universités, Université de Nice

Maître de Recherche, MINES ParisTech

Associate Professor, Instituto de Telecomunicações

Directeur de recherche CNRS, Gif-sur-Yvette

Professeur, Supélec, Gif-sur-Yvette



# Acknowledgements

This thesis was conducted at SONDRA laboratory in Supélec. I would like to thank all the people who have helped me through the completion of this dissertation.

First and foremost, I would like to begin by thanking my supervisors. Thanks to my thesis director Jean-Philippe Ovarlez for his guidance, support and enthusiasm. Your experience and scientific opening have been invaluable for me to find my own path in research. I would also like to thank my supervisor Frédéric Pascal who has taught me so much and has been endlessly patient with my countless questions. Your support and help have accompanied me during this thesis making the progress of research possible. It has been a great pleasure and privilege for me to have worked with both of you over the past three years.

I would also like to extend my appreciation to all of the members of my Ph.D. evaluation committee. Professors Grégoire Mercier and Cédric Richard for accepting this task and for their generous and professional commitment to the review. I also wish to say thank you to Professors Jocelyn Chanussot, José M. Bioucas-Dias and Sylvie Marcos for having agreed to be in my thesis jury. A special thank you goes to Professor Jesús Angulo for having introduced me into the world of research.

I would like to thank all my office colleagues at Supélec for making this a wonderful workplace. Notably, I want to thank Pierre Formont, Mélanie Mahot and Alice Combernoux for all the moments we enjoyed over these past three years, in the office and everywhere else. I want to express my gratitude to Marc Lesturgie for welcoming me in the laboratory and for all the opportunities SONDRA has offered me. Many thanks to Virginie Bouvier and to Anne-Hélène Picot for their indispensable help in solving administrative formalities. I also want to mention all other people in SONDRA and in Supélec with whom I had the pleasure to share many great times in the office Régis, Laetitia, Arnaud, Azza, Jérôme, Maxime, Takis, Miguel, and many others. I would also like to express my deepest gratitude to Alexandre Renaux. Many thanks for your support, your kindness and all the good advice you have given me. I also want to thank Eric Chaumette for his good cheer and for sharing his particular insight over science.

Part of this work has been possible thanks to the collaboration of some good friends and colleagues. I would like to acknowledge Miguel Veganzones for keeping his mind open and for joining me in the exploration of new topics. Thanks to Santiago Velasco for all the help he has given me throughout the years, specially during the time we spent together in Singapore.

I would like to say thank you to my colleagues during my undergraduate studies at UPC for the good times we had over the past years: Eva Lagunas, Marc Maceira,

Ricard Sauch, Laia Nadal and Lledó Esquerra. Even when we are apart I can count on you being on the other side of the screen. I also take this opportunity to greet my friends in Paris: Alba, Xisca, Julien, Claude, Julie, Quique, Irene, Miao, Will, Teresa, Julien and Javi. And of course my friends overseas, Arne and Isabel. Your encouraging words all through my studies have meant a world to me.

I am grateful to my family for giving me the opportunity to continue my studies all these years. My parents, Miquel and Francisca, and my brother Miquel deserve special mention for their unconditional support. I would like to say thank you to my grandmother Antonia and my grandfather Antoni for their love and kindness.

Last but not the least, special thanks to Armand for his support and understanding. Thank you for having been there for me through every difficulty and every success. These pages would have been completely different without you, and not only the french part. This manuscript is dedicated to you.

JOANA MARIA FRONTERA PONS





# Résumé

## 0.1 Introduction

L'imagerie hyperspectrale (HSI) repose sur le fait que, pour un matériau donné, la quantité de rayonnement émis varie avec la longueur d'onde. Les capteurs HSI mesurent le rayonnement des matériaux au sein de chaque pixel pour un très grand nombre de bandes spectrales contiguës et fournissent des images contenant des informations à la fois spatiales et spectrales. Les méthodes classiques de détection adaptative prennent généralement pour hypothèse un fond gaussien de vecteur moyenne nul ou connu. Cependant, quand le vecteur moyenne est inconnu, comme c'est le cas pour l'image hyperspectrale, il doit être inclus dans le processus de détection. Nous proposons dans ce travail d'étendre les méthodes classiques de détection pour lesquelles la matrice de covariance et le vecteur moyenne sont tous deux inconnus. Cependant, la distribution statistique multivariée des pixels de l'environnement peut s'éloigner de l'hypothèse gaussienne classiquement utilisée. La classe de distributions elliptiques a déjà été popularisée pour la caractérisation de fond pour l'HSI. Bien que ces modèles non gaussiens aient déjà été exploités dans la modélisation du fond et dans la conception de détecteurs, l'estimation des paramètres (matrice de covariance, vecteur moyenne) est encore généralement effectuée en utilisant des estimateurs conventionnels gaussiens. Dans ce contexte, nous analysons des méthodes d'estimation robuste plus appropriées à ces distributions non-gaussiennes : les  $M$ -estimateurs. Ces méthodes de détection couplées à ces nouveaux estimateurs permettent, d'une part d'améliorer les performances de détection dans un environnement non-gaussien, mais d'autre part de garder les mêmes performances que celles des détecteurs conventionnels dans un environnement gaussien. Elles fournissent ainsi un cadre unifié pour la détection de cibles et la détection d'anomalies pour la HSI.

## 0.2 État de l'art

Les méthodes de détection de cibles hyperspectrales sont couramment utilisées pour détecter des cibles intégrées dans le fond et qui, généralement, ne peuvent être résolues par la résolution spatiale [Matteoli et al. \(2010a\)](#). Il existe deux méthodes différentes pour la détection de cibles dans la littérature de HSI [Manolakis et al. \(2003\)](#): détection des anomalies et détection des cibles

Lorsque la signature spectrale de la cible souhaitée est connue, elle peut être utilisée comme vecteur de direction dans les techniques de détection des cibles. Le problème



de la détection de la cible peut être formulée comme un test d'hypothèses qui décide entre la présence ou l'absence d'une cible. Le cadre typique proposé en théorie de détection est basé sur l'approche de Neyman-Pearson [Van Trees \(2004\)](#). Les critères de conception consistent à maximiser la probabilité de détection pour une probabilité de fausse alarme donnée.

Dans de nombreuses situations pratiques, il n'y a pas suffisamment d'informations sur la cible à détecter, ainsi les méthodes de détection des anomalies sont largement utilisées [Eismann et al. \(2009\)](#). Les détecteurs d'anomalies recherchent des pixels dans l'image avec des caractéristiques spectrales qui diffèrent du fond. Ils peuvent être interprétés comme cas particulier de détection de cibles dans lequel aucune information a priori sur les éventails des cibles d'intérêt n'est disponible. Notez que, puisque les détecteurs d'anomalies n'emploient aucune connaissance à priori, ils ne peuvent pas distinguer les cibles et les détections vraies des pixels lumineux du fond ou des cibles qui ne sont pas d'intérêt. Ensuite, nous plaçons le cadre et décrivons les hypothèses classiques dans le traitement des signaux.

### 0.2.1 Notions préliminaires

Nous définissons les termes dont la connaissance est essentielle pour lire cette thèse et qui seront utilisés dans le présent document.

**Definition 0.2.1.** *Vecteur moyenne*

*Le vecteur moyenne d'un vecteur aléatoire complexe  $\mathbf{x} = \mathbf{u} + j\mathbf{v}$  est défini comme:*

$$\mathbb{E}[\mathbf{x}] = \mathbb{E}[\mathbf{u}] + j\mathbb{E}[\mathbf{v}]. \quad (1)$$

**Definition 0.2.2.** *Matrice de Covariance*

*La matrice de covariance d'un vecteur aléatoire complexe  $\mathbf{x} = \mathbf{u} + j\mathbf{v}$  de dimension  $m$  est définie comme:*

$$\mathbf{M} = \text{cov}(\mathbf{x}) = \mathbb{E}[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])^H]$$

*et  $\mathbf{M}$  appartient à la classe des matrices hermitiques semi-définies positives de taille  $m \times m$ .*

Il est souvent supposé que des signaux, les interférences, le bruit et le fond sont modélisés en tant que procédés stochastiques gaussiens. En effet, cette hypothèse semble raisonnable dans beaucoup d'applications [Kay \(1998\)](#). Nous rappelons d'abord, la définition d'un vecteur gaussien.

**Definition 0.2.3.** *Distribution gaussienne complexe*

*Un vecteur de dimension  $m$   $\mathbf{x} = \mathbf{u} + j\mathbf{v}$  a une distribution gaussienne complexe de moyenne  $\boldsymbol{\mu}$  et matrice de covariance  $\boldsymbol{\Sigma}$  si  $\mathbf{x} = (\mathbf{u}^T, \mathbf{v}^T)^T \in \mathbb{R}^{2m}$  a une distribution normale [van den Bos \(1995\)](#). Si  $\text{rank}(\boldsymbol{\Sigma}) = m$ , la fonction de densité de probabilité (DDP) existe et est de la forme:*

$$f_{\mathbf{x}}(\mathbf{x}) = \pi^{-m} |\boldsymbol{\Sigma}|^{-1} \exp\{-(\mathbf{x} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\}. \quad (2)$$

Dans la suite, elle sera notée comme  $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . Les estimateurs optimaux au sens du maximum de vraisemblance (MV) sont le Sample Mean Vector (SMV) et la Sample Covariance Matrix (SCM). Considérons  $\mathbf{x}_1, \dots, \mathbf{x}_N$  un  $N$ -échantillons de vecteurs indépendants gaussiens complexes, de dimension  $m$  et moyenne  $\boldsymbol{\mu}$ . Le SMV et la SCM s'écrivent alors:

$$\hat{\boldsymbol{\mu}}_{SMV} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i, \quad \hat{\boldsymbol{\Sigma}}_{SCM} = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^H. \quad (3)$$

En outre, nous noterons la Centered SCM (CSCM) comme:

$$\hat{\boldsymbol{\Sigma}}_{CSCM} = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^H. \quad (4)$$

**Definition 0.2.4.** *Distribution de Wishart complexe*

Soit  $\mathbf{x}_1, \dots, \mathbf{x}_N$  un  $N$ -échantillons de vecteurs indépendants et identiquement distribués (IID) où  $\mathbf{x}_i \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . Définissons  $\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}}_{SMV}$  et  $\hat{\mathbf{W}} = N \hat{\boldsymbol{\Sigma}}_{SCM}$  dénommée la matrice de Wishart. Ainsi, on a (voir [Gupta and Nagar \(2000\)](#) pour le cas réel):

- $\hat{\boldsymbol{\mu}}$  et  $\hat{\mathbf{W}}$  sont distribués indépendamment;
- $\hat{\boldsymbol{\mu}} \sim \mathcal{CN}(\boldsymbol{\mu}, \frac{1}{N} \boldsymbol{\Sigma})$ ;
- $\hat{\mathbf{W}} \sim \mathcal{CW}(N - 1, \boldsymbol{\Sigma})$  suit une distribution de Wishart avec  $N - 1$  degrés de liberté.

La distribution asymptotique de la matrice de Wishart est [Bilodeau and Brenner \(1999\)](#):

$$\sqrt{N} \text{vec}(\hat{\mathbf{W}} - \boldsymbol{\Sigma}) \xrightarrow{d} \mathcal{CN}(\mathbf{0}_{m^2, 1}, (\boldsymbol{\Sigma}^T \otimes \boldsymbol{\Sigma}), (\boldsymbol{\Sigma}^T \otimes \boldsymbol{\Sigma}) \mathbf{K}). \quad (5)$$

où  $\mathbf{K}$  est la matrice de commutation. Par conséquent, les performances obtenues avec les estimateurs SMV-SCM sont parfaitement connues. Ces estimateurs sont sans biais, convergents et asymptotiquement gaussiens.

## 0.2.2 Techniques classiques de détection de cibles

En traitement du signal, le problème de la détection de cibles est généralement énoncé comme un test d'hypothèses binaire qui détermine la présence ou l'absence d'une cible dans la cellule analysée. Par conséquent, le signal reçu est un vecteur  $\mathbf{x}$  de dimension  $m$  dénommé vecteur d'observation. Le signal complexe connu caractérisant la cible visée est désigné par  $\mathbf{s}$ , et il est corrompu par un bruit additif  $\mathbf{b}$  dénommé fond. Le problème de détection peut s'écrire:

$$\begin{cases} \mathcal{H}_0 : \mathbf{x} = \mathbf{b}, & \mathbf{x}_i = \mathbf{b}_i, i = 1, \dots, N \\ \mathcal{H}_1 : \mathbf{x} = \mathbf{s} + \mathbf{b}, & \mathbf{x}_i = \mathbf{b}_i, i = 1, \dots, N, \end{cases} \quad (6)$$

où  $\mathbf{x}_i$  sont les données dites secondaires utilisées pour estimer les paramètres du bruit, et le signal  $\mathbf{s}$  peut être écrit sous la forme  $\alpha \mathbf{p}$ , où  $\alpha$  est une amplitude complexe inconnue, et  $\mathbf{p}$  décrit le signal qu'on cherche. Les détecteurs sont conçus suivant le critère de Neyman-Pearson. Il vise à maximiser la probabilité de détection (PD) pour une probabilité de fausse alarme (PFA) fixe, qui est la probabilité de décider que la cible est présente, alors qu'en fait elle ne l'est pas. Le test optimal en fonction de ce critère est le rapport de vraisemblance (RV) donné par :

$$\Lambda(\mathbf{x}) = \frac{p(\mathbf{x}|\mathcal{H}_1)}{p(\mathbf{x}|\mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta. \quad (7)$$

Le seuil de détection  $\eta$  est déterminé selon une valeur fixe pour la probabilité de fausse alarme  $PFA = k$  et calculé comme :

$$PFA = \mathbb{P}(\Lambda(\mathbf{x}; \mathcal{H}_0) > \eta) = k, \quad (8)$$

Le but de ce chapitre est de fournir des expressions analytiques pour cette équation. Puisque les paramètres statistiques du fond, sont supposés être inconnus, ils doivent être estimés à partir de  $\mathbf{x}_1, \dots, \mathbf{x}_N \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  une série de  $N$  IID données secondaires sans cible. Ensuite, le détecteur adaptatif est obtenu en remplaçant dans Eq. (7) les paramètres inconnus par leurs estimations menant au test de rapport de vraisemblance généralisé (RVG).

### 0.2.2.1 Adaptive Matched Filter

Le filtre adapté (MF) est le détecteur optimal pour maximiser le rapport du signal au bruit en présence d'un bruit additif gaussien avec des paramètres connus [Kay \(1998\)](#). Et le détecteur MF prend la forme:

$$\Lambda_{MF} = \frac{|\mathbf{p}^H \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})|^2}{(\mathbf{p}^H \boldsymbol{\Sigma}^{-1} \mathbf{p})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda. \quad (9)$$

La courbe "PFA-seuil" correspond au comportement statistique d'un détecteur sous hypothèse nulle. Par ailleurs, la relation "PFA-seuil" est donnée par [Kay \(1998\)](#):

$$PFA_{MF} = \exp(-\lambda). \quad (10)$$

Le filtre adapté adaptatif (AMF), noté  $\Lambda_{AMF\hat{\boldsymbol{\Sigma}}}^{(N)}$  pour souligner la dépendance avec  $N$ , est habituellement construit en remplaçant la matrice de covariance  $\boldsymbol{\Sigma}$  par son estimateur  $\hat{\boldsymbol{\Sigma}}$  obtenu à partir des  $N$  données secondaires. Le vecteur moyenne est généralement connu. Ainsi, la version adaptative devient:

$$\Lambda_{AMF\hat{\boldsymbol{\Sigma}}}^{(N)} = \frac{|\mathbf{p}^H \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{x} - \boldsymbol{\mu})|^2}{(\mathbf{p}^H \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{p})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda. \quad (11)$$

Alors, la relation théorique "PFA-seuil" est donnée par [Robey et al. \(1992\)](#) pour un estimateur suivant une distribution Wishart obtenu avec  $\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Sigma}}_{SCM}$ :

$$PFA_{AMF\hat{\boldsymbol{\Sigma}}} = {}_2F_1 \left( N - m + 1, N - m + 2; N + 1; -\frac{\lambda}{N} \right), \quad (12)$$

où  ${}_2F_1(\cdot)$  est la fonction hypergéométrique [Abramowitz et al. \(1964\)](#) définie comme:

$${}_2F_1(a, b; c; z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 \frac{t^{b-1}(1-t)^{c-b-1}}{(1-tz)^a} dt. \quad (13)$$

Ce détecteur possède les propriétés Taux de Fausse Alarme Constant (TFAC) dans le sens où l'expression de fausse alarme ne dépend que de la dimension du vecteur  $m$  et le nombre de données secondaires utilisées pour l'estimation  $N$ . Notez qu'il est également indépendant de la matrice de covariance de bruit  $\Sigma$ . Cependant, sa performance s'appuie fortement sur le bon ajustement du modèle gaussien et le taux de fausses alarmes est fortement augmenté lorsque l'hypothèse normale n'est pas vérifiée.

### 0.2.2.2 Détecteur de Kelly

Le détecteur de Kelly a été établi dans [Kelly \(1986\)](#). Dans ce cas, on suppose que la matrice de covariance  $\Sigma$  est inconnue, le vecteur moyenne est censé être connu. Par conséquent, le détecteur de Kelly adaptatif prend la forme suivante:

$$\Lambda_{Kelly\hat{\Sigma}}^{(N)} = \frac{|\mathbf{p}^H \hat{\Sigma}_{CSCM}^{-1} (\mathbf{x} - \boldsymbol{\mu})|^2}{\left(\mathbf{p}^H \hat{\Sigma}_{CSCM}^{-1} \mathbf{p}\right) \left(N + (\mathbf{x} - \boldsymbol{\mu})^H \hat{\Sigma}_{CSCM}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda, \quad (14)$$

où  $\lambda = 1 - \eta^{-\frac{1}{N+1}}$ . Comme montré dans [Kelly \(1986\)](#), la PFA pour le test de Kelly est donnée par:

$$PFA_{Kelly} = (1 - \lambda)^{N-m+1}. \quad (15)$$

Le détecteur de Kelly est un test TFAC, dans lequel la PFA est indépendante de la véritable matrice de covariance. Cependant, il n'a pas de propriété d'optimalité connue dans le sens de maximiser la probabilité de détection pour une probabilité de fausse alarme donnée comme expliqué dans [Kelly \(1986\)](#).

### 0.2.2.3 Adaptive Normalized Matched Filter

Le filtre adapté normalisé (NMF) est obtenu lorsque l'on considère que la matrice de covariance est différente dans les deux hypothèses; c'est à dire que le fond a la même structure de covariance mais une variance différente. Le détecteur NMF peut être écrit selon [Scharf and Friedlander \(1994\)](#):

$$\Lambda_{NMF} = \frac{|\mathbf{p}^H \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})|^2}{(\mathbf{p}^H \Sigma^{-1} \mathbf{p}) ((\mathbf{x} - \boldsymbol{\mu})^H \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}))} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda, \quad (16)$$

où  $\lambda = 1 - \eta^{-\frac{1}{m}}$  et pour lequel on a [Scharf and Friedlander \(1994\)](#):

$$PFA_{NMF} = (1 - \lambda)^{m-1}. \quad (17)$$

Le filtre adapté normalisé adaptatif (ANMF) est généralement obtenu lorsque la matrice de covariance du bruit inconnue est remplacée par une estimation menant à [Kraut et al. \(2001\)](#):

$$\Lambda_{ANMF\hat{\Sigma}}^{(N)} = \frac{|\mathbf{p}^H \hat{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})|^2}{\left(\mathbf{p}^H \hat{\Sigma}^{-1} \mathbf{p}\right) \left((\mathbf{x} - \boldsymbol{\mu})^H \hat{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda. \quad (18)$$

Et la PFA pour un estimateur suivant une distribution de Wishart obtenue avec  $\hat{\Sigma} = \hat{\Sigma}_{SCM}$  peut s'écrire selon [Kraut et al. \(2001\)](#):

$$PFA_{ANMF\hat{\Sigma}} = (1 - \lambda)^{a-1} {}_2F_1(a, a-1; b-1; \lambda), \quad (19)$$

où  $a = N - m + 2$  and  $b = N + 2$ .

Ce test de détection détient des propriétés TFAC importantes comme que sa distribution est indépendante de la matrice de covariance de bruit.

## 0.3 Détection de cibles dans un environnement gaussien

Le but de ce chapitre est de généraliser les méthodes classiques de détection de cibles dans le cas où le vecteur moyenne est inconnu et doit être estimé.

### 0.3.1 Détection Gaussienne de moyenne différent de zéro

Supposons maintenant que le vecteur moyenne  $\mu$  est inconnu comme c'est le cas par exemple en HSI et dérivons les nouveaux schémas de détection. Puis, en utilisant le calcul standard sur les distributions de Wishart, récapitulé dans la Section 0.2.1, les distributions de chaque test de détection sont fournies.

#### 0.3.1.1 Adaptive Matched Filter

Quand la matrice de covariance et le vecteur moyenne sont à la fois inconnus, ils sont remplacés par leurs estimations à partir des données secondaires dans Eq. (9) et le détecteur AMF s'écrit:

$$\Lambda_{AMF\hat{\Sigma},\hat{\boldsymbol{\mu}}}^{(N)} = \frac{|\mathbf{p}^H \hat{\Sigma}^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}})|^2}{\left(\mathbf{p}^H \hat{\Sigma}^{-1} \mathbf{p}\right)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda, \quad (20)$$

où la notation  $\Lambda_{AMF\hat{\Sigma},\hat{\boldsymbol{\mu}}}^{(N)}$  est utilisée pour souligner la dépendance sur le vecteur moyenne estimé  $\hat{\boldsymbol{\mu}}$ . La distribution de ce test de détection est donnée dans la proposition suivante, à travers sa PFA.

**Proposition 0.3.1.** *Sous des hypothèses gaussiennes, la relation théorique entre la PFA et le seuil est donnée par*

$$PFA_{AMF_{\hat{\Sigma}, \hat{\mu}}} = {}_2F_1 \left( N - m, N - m + 1; N; -\frac{\lambda'}{N - 1} \right), \quad (21)$$

où  $\lambda' = \frac{(N-1)}{(N+1)}\lambda$ ,  $\hat{\Sigma} = \hat{\Sigma}_{SCM}$  et  $\hat{\mu} = \hat{\mu}_{SMV}$ , rappelé dans la Section 0.2.1.

### 0.3.1.2 Détecteur de Kelly

Le détecteur de Kelly pour le vecteur moyenne inconnu et la matrice de covariance inconnue doit être dérivé car il n'est pas le Kelly précédent dans lequel une estimation du vecteur moyenne est branchée. Suivant les mêmes lignes que dans Kelly (1986), nous supposons maintenant que tant le vecteur moyenne et la matrice de covariance sont inconnus. Ainsi, le RVG peut maintenant s'écrire selon la définition suivante.

**Definition 0.3.1** (Le détecteur de Kelly généralisé). *Sous des hypothèses gaussiennes, l'extension du test de Kelly prend la forme suivante lorsque le vecteur moyenne et la matrice de covariance du fond sont tous les deux inconnus:*

$$\Lambda = \frac{\beta(N) \left| \mathbf{p}^H \hat{\mathbf{S}}_0^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_0) \right|^2}{(\mathbf{p}^H \hat{\mathbf{S}}_0^{-1} \mathbf{p}) \left( 1 + (\mathbf{x} - \hat{\boldsymbol{\mu}}_0)^H \hat{\mathbf{S}}_0^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_0) \right)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda, \quad (22)$$

où  $\beta(N) = \frac{N+1}{N}$ ,  $\lambda = \frac{\eta-1}{\eta}$  et

- $\hat{\mathbf{S}}_0 = \sum_{i=1}^N (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_0)(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_0)^H,$
- $\hat{\boldsymbol{\mu}}_0 = \frac{1}{N+1} \left( \mathbf{x} + \sum_{i=1}^N \mathbf{x}_i \right).$

On peut remarquer que tous les deux l'estimateur de la matrice de covariance  $\hat{\mathbf{S}}_0$  et l'estimateur la moyenne  $\hat{\boldsymbol{\mu}}_0$  dépendent de la donnée sous test  $\mathbf{x}$ , ce qui n'est pas le cas dans d'autres détecteurs classiques où les paramètres inconnus sont estimés à partir de données secondaires sans cibles. Par conséquent,  $\hat{\mathbf{S}}_0$  et  $\mathbf{x} - \hat{\boldsymbol{\mu}}_0$  ne sont pas indépendants. De plus, l'estimateur de la matrice de covariance  $\hat{\mathbf{S}}_0$  ne suit pas une distribution de Wishart à cause de l'estimateur de la moyenne  $\hat{\boldsymbol{\mu}}_0$ . En conséquence, la dérivation de la distribution de ce test est très difficile.

Egalement, on peut utiliser le schéma classique rappelé dans Eq. (14) et remplacer l'estimateur classique de la moyenne basé uniquement sur des données secondaires.

Cela mène au détecteur de Kelly “plug-in”:

$$\Lambda_{Kelly\hat{\Sigma},\hat{\mu}}^{(N)} = \frac{|\mathbf{p}^H \hat{\Sigma}_{SCM}^{-1} (\mathbf{x} - \hat{\mu}_{SMV})|^2}{\left(\mathbf{p}^H \hat{\Sigma}_{SCM}^{-1} \mathbf{p}\right) \left(N + (\mathbf{x} - \hat{\mu}_{SMV})^H \hat{\Sigma}_{SCM}^{-1} (\mathbf{x} - \hat{\mu}_{SMV})\right)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda. \quad (23)$$

Dans ce cas, la distribution peut être obtenue. Tel est l’objet de la proposition suivante.

**Proposition 0.3.2.** *La relation théorique entre la PFA et le seuil est donnée par*

$$PFA_{Kelly\hat{\Sigma},\hat{\mu}} = \frac{\Gamma(N)}{\Gamma(N-m+1)\Gamma(m-1)} \int_0^1 \left[ 1 + \frac{\lambda}{1-\lambda} \left( 1 - \frac{u}{N+1} \right) \right]^{m-N} u^{N-m} (1-u)^{m-2} du, \quad (24)$$

où  $\hat{\Sigma} = \hat{\Sigma}_{SCM}$  and  $\hat{\mu} = \hat{\mu}_{SMV}$ .

### 0.3.1.3 Adaptive Normalized Matched Filter

De même, l’ANMF dans le cas où le vecteur moyenne et la matrice de covariance sont tous les deux estimés devient:

$$\Lambda_{ANMF\hat{\Sigma},\hat{\mu}} = \frac{|\mathbf{p}^H \hat{\Sigma}^{-1} (\mathbf{x} - \hat{\mu})|^2}{(\mathbf{p}^H \hat{\Sigma}^{-1} \mathbf{p}) \left( (\mathbf{x} - \hat{\mu})^H \hat{\Sigma}^{-1} (\mathbf{x} - \hat{\mu}) \right)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda. \quad (25)$$

**Proposition 0.3.3.** *La relation théorique entre la PFA et le seuil est donnée par*

$$PFA_{ANMF\hat{\Sigma},\hat{\mu}} = (1-\lambda)^{a-1} {}_2F_1(a, a-1; b-1; \lambda), \quad (26)$$

où  $a = (N-1) - m + 2$  and  $b = (N-1) + 2$ ,  $\hat{\Sigma} = \hat{\Sigma}_{SCM}$  et  $\hat{\mu} = \hat{\mu}_{SMV}$ .

### 0.3.1.4 Simulations

Dans cette section, nous validons l’analyse théorique sur des données simulées. Les expériences ont été réalisées sur des vecteurs gaussiens de dimension  $m = 5$ , pour différentes valeurs de  $N$ , le nombre de données secondaires et les calculs ont été faits pour  $10^6$  essais de Monte-Carlo. La vraie covariance est choisie comme une matrice de Toeplitz dont les entrées sont  $\Sigma_{i,j} = \rho^{|i-j|}$  et  $\rho = 0.4$ . Le vecteur moyenne est réglé pour avoir toutes les entrées égales à  $(3 + 4j)$ .

Fig. 1 montre la régulation de la fausse alarme pour le MF, l’AMF lorsque seule la matrice de covariance est inconnue et l’ANMF où à la fois la matrice de covariance et le vecteur moyenne sont inconnus pour différentes valeurs de  $N$ . L’ajustement parfait entre les courbes vertes et jaunes illustre les résultats de la Proposition 0.3.1.

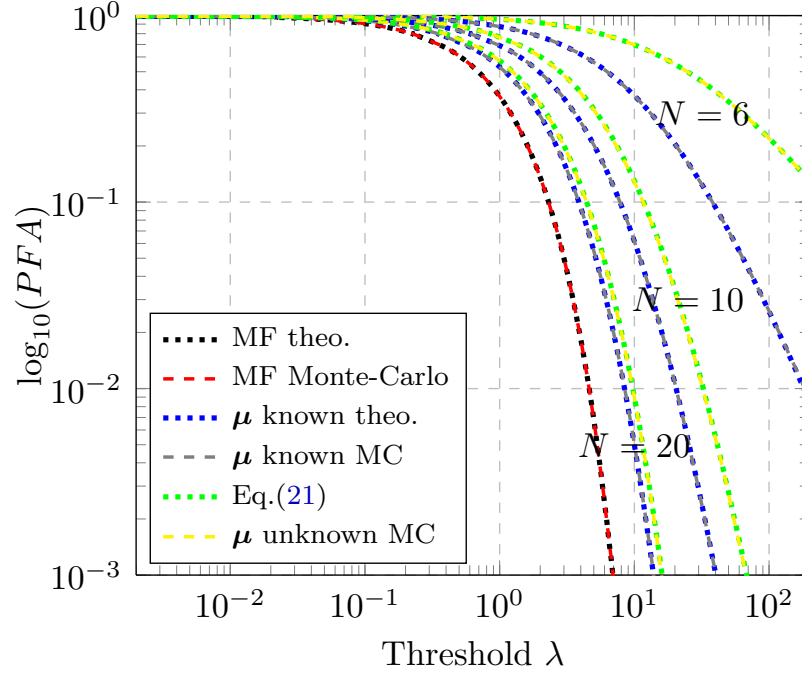


Figure 1: PFA par rapport au seuil de l'AMF lorsque (1)  $\mu$  et  $\Sigma$  sont connus (MF) (courbes rouges et noires) (2) seulement  $\mu$  est connue (courbes grises et bleues) (3) Proposition 0.3.1: les deux  $\mu$  et  $\Sigma$  sont inconnus (courbes jaunes et vertes).

Fig. 2 et Fig. 3 présentent la régulation de la FA pour le détecteur de Kelly et l'ANMF respectivement, sous hypothèse gaussienne. Pour plus de clarté, les résultats sont affichés en termes de seuil  $\eta$  respectivement d'après Eq. (14),  $\eta = (1 - \lambda)^{-(N+1)}$ , et Eq. (16),  $\eta = (1 - \lambda)^{-m}$ , et une échelle logarithmique est utilisée. Cela valide les résultats des Propositions 0.3.2 et 0.3.3 pour les SCM-SMV.

### 0.3.2 Détection des anomalies

Récapitulons maintenant les détecteurs d'anomalies les plus populaires (voir e.g. Nasrabadi (2014) pour une étude complète sur les méthodes de détection d'anomalies).

#### 0.3.2.1 Détecteur de Reed-Xiaoli

Le détecteur original de Reed-Xiaoli (RXD) proposé dans Reed and Yu (1990) est communément considéré comme le détecteur d'anomalies de référence pour les données hyperspectrales. Le schéma de détection prend la forme:

$$\Lambda(\mathbf{X}) = \frac{(\mathbf{X}\boldsymbol{\alpha}^T)^T(\mathbf{X}\mathbf{X}^T)^{-1}(\mathbf{X}\boldsymbol{\alpha}^T)}{\boldsymbol{\alpha}\boldsymbol{\alpha}^T} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda.$$

Comme les données hyperspectrales ne sont pas de moyenne nulle, nous considérons maintenant que le fond  $\mathbf{b}_i$  est distribué selon  $\mathcal{N}(\mu, \Sigma)$  et le vecteur moyenne  $\mu$  est



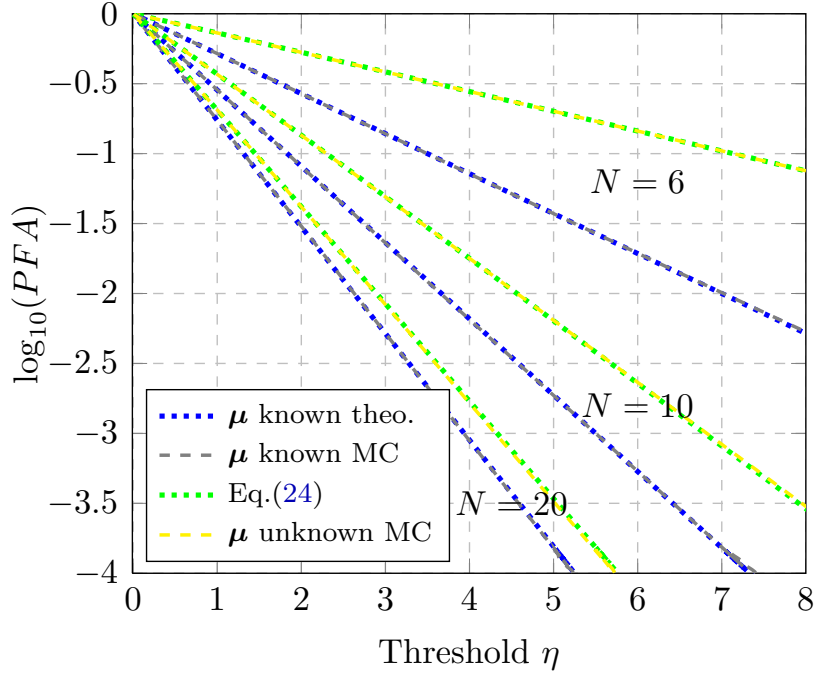


Figure 2: PFA par rapport au seuil de le détecteur "plug-in" de Kelly lorsque (1) seulement  $\mu$  est connue (courbes grises et bleues) (2) Proposition 0.3.2: les deux  $\mu$  et  $\Sigma$  sont inconnus (courbes jaunes et vertes).

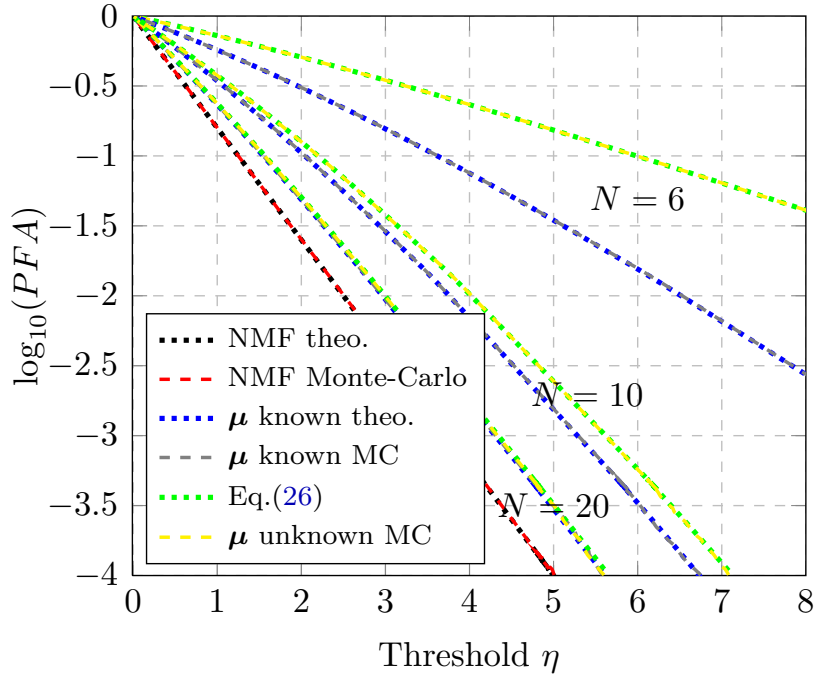


Figure 3: PFA par au rapport seuil de l'ANMF lorsque (1)  $\mu$  et  $\Sigma$  sont connus (NMF) (courbes rouges et noires) (2) seulement  $\mu$  est connue (courbes grises et bleues) (3) Proposition 0.3.3: les deux  $\mu$  et  $\Sigma$  sont inconnus (courbes jaunes et vertes).

censé être connu. Dans le cas où juste une anomalie dans les données sous test doit être détectée, le vecteur d'amplitude correspondant peut être écrit comme  $\boldsymbol{\alpha}_i = [0 \dots 0 \mathbf{1} 0 \dots 0]^T$  où 1 est à la position  $i^{ieme}$  et le détecteur précédent s'écrit:

$$\Lambda_{RXD} = (\mathbf{x}_i - \boldsymbol{\mu})^T \hat{\boldsymbol{\Sigma}}_{SCM}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda.$$

Enfin, étant donné que le vecteur moyenne est généralement inconnu, il peut être remplacé sur le détecteur dans son estimation par  $\hat{\boldsymbol{\mu}}_{SMV}$ . Le détecteur résultant, communément appelé RVG en deux étapes, s'écrit ainsi:

$$\Lambda_{ARXD} = (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{SMV})^T \hat{\boldsymbol{\Sigma}}_{SCM}^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{SMV}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda. \quad (27)$$

L'estimation de la matrice de covariance  $\hat{\boldsymbol{\Sigma}}_{SCM}$  dans Eq. (27), est effectuée sur tout l'ensemble de données, c'est à dire y compris le vecteur  $\mathbf{x}_i$  sous test. Dans ce qui suit, le test dans Eq. (27) sera désigné comme RXD Adaptatif (ARXD), pour souligner le fait que le vecteur moyenne inconnu est remplacé par son estimation.

### 0.3.2.2 Détecteur d'anomalies de Kelly

Après le développement proposé dans Kelly (1986), nous détaillons maintenant un détecteur d'anomalies classique souvent désigné par erreur comme le RXD. Comme dans le détecteur de Kelly classique, la matrice de covariance  $\boldsymbol{\Sigma}$  est inconnue et le vecteur moyenne  $\boldsymbol{\mu}$  est censé être connu. Toutefois, pour la dérivation du détecteur d'anomalies, l'amplitude du signal  $\alpha$  est supposée être connue et le paramètre inconnu est maintenant le vecteur de direction  $\mathbf{p}$ . Maximisant le RV sur la signature spectrale inconnue  $\mathbf{p}$ , il est facile de montrer que le test de RVG résultant est équivalent à:

$$\Lambda_{KellyAD \hat{\boldsymbol{\Sigma}}} = (\mathbf{x} - \boldsymbol{\mu})^T \hat{\boldsymbol{\Sigma}}_{SCM}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda. \quad (28)$$

Comme discuté ci-dessus, lorsque le vecteur moyenne est inconnu, il peut être remplacé sur le détecteur (RVG en deux étapes) par son MV ayant:

$$\Lambda_{KellyAD \hat{\boldsymbol{\Sigma}}, \hat{\boldsymbol{\mu}}}^{(N)} = (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})^T \hat{\boldsymbol{\Sigma}}_{SCM}^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda. \quad (29)$$

La distribution de ce test de détection est donnée dans la proposition suivante.

**Proposition 0.3.4.** *La distribution du détecteur sous l'hypothèse gaussienne est donnée par*

$$\frac{N - m}{m(N + 1)} \Lambda_{KellyAD \hat{\boldsymbol{\Sigma}}, \hat{\boldsymbol{\mu}}}^{(N)} \sim F_{m, N-m}, \quad (30)$$

où  $F_{m, N-m}$  est la distribution  $F$  décentrée avec  $m$  et  $N - m$  degrés de liberté et  $\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Sigma}}_{SCM}$  et  $\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}}_{SMV}$ .

### 0.3.2.3 RXD Normalisé et détecteur de cibles uniforme

Nous rappelons maintenant deux variantes populaires de la distance de Mahalanobis décrit dans [Chang and Chiang \(2002\)](#): RXD Normalisé (N-RXD) et le détecteur de cibles uniforme (UTD).

Le N-RXD prend la forme:

$$\Lambda_{N-RXD} = \frac{(\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})^T}{\|\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}\|} \hat{\boldsymbol{\Sigma}}_{SCM}^{-1} \frac{(\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})^T}{\|\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}\|} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda, \quad (31)$$

où  $\|\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}\|^2 = (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})^T (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})$  représente la norme euclidienne du vecteur.

L'UTD est un autre test de détection d'anomalie répandu. Il a d'abord été introduit dans [Harsanyi \(1993\)](#) et est défini comme:

$$\Lambda_{UTD} = (\mathbf{1} - \hat{\boldsymbol{\mu}}_{SMV})^T \hat{\boldsymbol{\Sigma}}_{SCM}^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda. \quad (32)$$

avec  $\mathbf{1} = [1, \dots, 1]^T$  est le vecteur unité de taille  $m$ .

### 0.3.2.4 Détecteur d'anomalies de Kelly generalisé

Dans le cas où la matrice de covariance et le vecteur moyenne sont à la fois inconnus, nous devons dériver un nouveau détecteur. Cette stratégie est similaire à celle proposée dans la section 0.3.1.2 pour le test de détection de Kelly généralisé. Ensuite, le RV doit être maximisé par rapport à  $\mathbf{p}$ . Le maxima est obtenu en prenant:

$$\hat{\mathbf{p}} = \frac{N+1}{N} \frac{(\mathbf{x} - \boldsymbol{\mu}_0)}{\alpha}. \quad (33)$$

Par conséquent, le détecteur résultant prend la forme:

$$\Lambda_{G-KellyAD} = (\mathbf{x} - \hat{\boldsymbol{\mu}}_0)^H \hat{\mathbf{S}}_0^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_0) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda, \quad (34)$$

où  $\hat{\mathbf{S}}_0 = \sum_{i=1}^N (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_0)(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_0)^H$ , et  $\hat{\boldsymbol{\mu}}_0 = \frac{1}{N+1} \left( \mathbf{x} + \sum_{i=1}^N \mathbf{x}_i \right)$ . A nouveau, l'estimateur du vecteur moyenne  $\boldsymbol{\mu}_0$  et la matrice de covariance  $\mathbf{S}_0$  dépendent de la donnée sous test  $\mathbf{x}$ . Ainsi,  $\mathbf{x} - \hat{\boldsymbol{\mu}}_0$  et  $\hat{\mathbf{S}}_0$  ne sont pas indépendants.

## 0.4 Détection de cibles dans un environnement non-gaussien

Ce chapitre examine les caractéristiques principales de la famille des distributions elliptiques et les correspondants estimateurs robustes,  $M$ -estimateurs. Nous décrivons les différentes techniques de détection de cibles analysées et présentons les détecteurs d'anomalies dans un environnement non-gaussien.

### 0.4.1 Distributions elliptiques

Dans cette section, nous présentons la classe des distributions elliptiques complexes [Krishnaiah and Lin \(1986\)](#), [Ollila et al. \(2012\)](#). Elle décrit une famille de distributions multivariées qui servent principalement comme des alternatives à longue queue au modèle normal multivarié.

#### 0.4.1.1 Définition

**Definition 0.4.1.** *Un vecteur aléatoire complexe de dimension  $m$   $\mathbf{z}$  a une distribution elliptique complexe (CE) si sa fonction caractéristique est de la forme:*

$$\Phi_{\mathbf{z}}(\mathbf{c}) = \exp(j \Re(\mathbf{c}^H \boldsymbol{\mu})) \phi(\mathbf{c}^H \boldsymbol{\Sigma} \mathbf{c}), \quad (35)$$

pour une fonction  $\phi : \mathbb{R}^+ \rightarrow \mathbb{R}$ , appelée fonction génératrice, une matrice semi-définie positive  $\boldsymbol{\Sigma}$ , appelée matrice de dispersion et  $\boldsymbol{\mu} \in \mathbb{C}^m$  le vecteur de position. Nous écrirons  $\mathbf{z} \sim \mathcal{CE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ .

**Theorem 0.4.1.** (Théorème de représentation stochastique)

*Un vecteur aléatoire complexe de dimension  $m$   $\mathbf{z} \sim \mathcal{CE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$  avec un  $\text{rang}(\boldsymbol{\Sigma}) = k \leq m$  si et seulement si il admet la représentation suivante:*

$$\mathbf{z} \stackrel{d}{=} \boldsymbol{\mu} + \mathcal{R} \mathbf{A} \mathcal{U}^{(k)}, \quad (36)$$

où  $\mathcal{U}^{(k)}$  est un vecteur aléatoire de dimension  $k$  distribué uniformément sur la  $k$ -sphère complexe unitaire  $\mathbb{CS}^k$ ;  $\mathcal{R}$  est une variable aléatoire non négative appelée génératrice de la forme, étant stochastiquement indépendante de  $\mathcal{U}^{(k)}$ ;  $\boldsymbol{\mu} \in \mathbb{C}^m$  et  $\boldsymbol{\Sigma} = \mathbf{A} \mathbf{A}^H$  est une factorisation de  $\boldsymbol{\Sigma}$  où  $\mathbf{A} \in \mathbb{C}^{m \times k}$  avec  $\text{rank}(\mathbf{A}) = k$ .

La génératrice de la forme  $\mathcal{R}$  détermine la forme de la distribution, en particulier la queue de la distribution. D'après  $\mathbf{z} \sim \mathcal{CE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ , il n'entraîne pas que  $\mathbf{z}$  a une DDP  $f_{\mathbf{z}}(\cdot)$ . Si elle existe, elle peut être liée à la fonction de densité de la génératrice de la forme  $\mathcal{R}$ , à condition que  $\mathcal{R}$  soit absolument continue. Alors, la DDP de  $\mathbf{z}$  a la forme:

$$f_{\mathbf{z}}(\mathbf{z}) = c_{m,h} |\boldsymbol{\Sigma}|^{-1} h((\mathbf{z} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \boldsymbol{\mu})) \quad (37)$$

où  $h$  est une fonction telle que (37) définit une DDP dans  $\mathbb{C}^m$ . La fonction  $h$  est généralement appelée *générateur de densité* et elle est supposé être connue seulement approximativement. La classe de distributions elliptiques comprend un grand nombre de distributions largement répandues, comme par exemple la gaussienne multivariée [Goodman \(1963\)](#), la distribution  $K$  [Conte et al. \(1991\)](#) ou la distribution de Student multivariée [Krishnaiah and Lin \(1986\)](#).

### 0.4.2 Estimation robuste

Afin d'améliorer l'estimation des paramètres, l'objectif est de trouver un modèle approprié et utiliser les MLE correspondantes. Cette méthode conduit à des estimateurs

asymptotiquement efficaces mais pas nécessairement robustes. En effet, l'estimateur robuste est plutôt celui qui est encore assez fiable, quel que soit le départ de données, à défaut d'être optimale dans certains scénarios. Dans des applications réelles, bien que les distributions elliptiques offrent beaucoup de distributions possibles, le risque que les données ne suivent pas le modèle considéré demeure.

#### 0.4.2.1 $M$ -Estimators

Lorsque le générateur de densité  $h_m$  est inconnu, les  $M$ -estimateurs permettent une approche alternative pour l'estimation robuste des paramètres des populations elliptiques. Supposons  $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N$  échantillons IID à partir d'une  $\mathcal{CE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h_m)$  avec  $N > m$ . Les  $M$ -estimateurs complexes de position et de dispersion sont définis comme les solutions conjointes de:

$$\hat{\boldsymbol{\mu}}_N = \frac{\sum_{i=1}^N u_1(t_i) \mathbf{z}_i}{\sum_{i=1}^N u_1(t_i)}, \quad \hat{\boldsymbol{\Sigma}}_N = \frac{1}{N} \sum_{i=1}^N u_2(t_i^2) (\mathbf{z}_i - \hat{\boldsymbol{\mu}}) (\mathbf{z}_i - \hat{\boldsymbol{\mu}})^H, \quad (38)$$

où  $t_i = ((\mathbf{z}_i - \hat{\boldsymbol{\mu}})^H \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{z}_i - \hat{\boldsymbol{\mu}}))^{1/2}$  et  $u_1(\cdot), u_2(\cdot)$  dénotent toute *fonctions de pondération* valeur réelle sur la forme quadratique  $t_i$ . L'objectif principal de  $u_1(\cdot)$  et  $u_2(\cdot)$  est d'atténuer les contributions des valeurs aberrantes. Le choix de  $u_1(\cdot)$  et  $u_2(\cdot)$  n'a pas besoin d'être lié à une distribution elliptique particulière et, par conséquent, les  $M$ -estimateurs constituent une large classe d'estimateurs qui inclut les MLE pour le cas particulier  $u_1(t) = -h'_m(t^2)/h_m(t^2)$  et  $u_2(t^2) = u_1(t)$ .

L'existence et l'unicité ont été prouvées dans le cas réel, à condition que les fonctions  $u_1(\cdot), u_2(\cdot)$  vérifient un ensemble d'hypothèses générales formulées par Maronna [Maronna \(1976\)](#). Olilla a montré dans [Olilla and Koivunen \(2003a\)](#) que ces conditions sont également vérifiées dans le cas complexe.

**Remark 0.4.1.** *En général, une fonction de pondération consistante devrait tendre vers zéro à l'infini. Cela signifie que des poids petits sont donnés à ces observations qui sont très éloignées en termes de la forme quadratique  $t_i$ . Les observations qui restent lointaines du modèle de base auront une moindre contribution à l'estimation des paramètres. Par exemple, la SCM donne un poids unitaire ( $u_2(t) = u_1(t) = 1$ ) à toutes les observations, et donc est naturellement non robuste. Lorsqu'il s'agit de modèles à queue lourde, l'utilisation d'estimations robustes diminue l'impact des échantillons très impulsifs et les valeurs aberrantes possibles dans les cellules de référence.*

**0.4.2.1.1 Exemples de  $M$ -estimateurs** Nous présentons ici quelques exemples des  $M$ -estimateurs.

**0.4.2.1.1.1 Huber type  $M$ -estimateurs** Les  $M$ -estimateurs d'*Huber* peuvent être obtenus comme les solutions des équations implicites suivantes:

$$\hat{\boldsymbol{\mu}}_{Hub} = \frac{1}{N} \sum_{i=1}^N \left[ \mathbf{z}_i \mathbb{1}_{t_i \leq k} + k \frac{\mathbf{z}_i}{\left( (\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{Hub})^H \hat{\mathbf{M}}_{Hub}^{-1} (\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{Hub}) \right)^{1/2}} \mathbb{1}_{t_i > k} \right], \quad (39)$$

$$\begin{aligned} \hat{\mathbf{M}}_{Hub} = \\ \frac{1}{N\beta} \sum_{i=1}^N \left[ (\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{Hub}) (\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{Hub})^H \mathbb{1}_{t_i^2 \leq k^2} + k^2 \frac{(\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{Hub})(\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{Hub})^H}{(\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{Hub})^H \hat{\mathbf{M}}_{Hub}^{-1} (\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{Hub})} \mathbb{1}_{t_i^2 > k^2} \right], \end{aligned} \quad (40)$$

où  $t_i^2$  est la forme quadratique  $(\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{Hub})^H \hat{\boldsymbol{\Sigma}}_{Hub}^{-1} (\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{Hub})$  et  $\mathbb{1}(\cdot)$  est la fonction indicateur définie comme  $\mathbb{1}_A = 1$  si  $A$  et  $\mathbb{1}_A = 0$  autrement.

Les constantes  $k$  et  $\beta$  sont des paramètres réglables qui permettent de choisir le pourcentage de données atténuées. Le choix  $q = 1$  donne  $u_1 = 1$  et  $u_2 = 1$  et les estimateurs d'*Huber* correspondent respectivement à la SMV et à la SCM, alors que le choix  $q = 0$  mène à  $u_1(t) = t^{-1}$  et  $u_2(t^2) = mt^{-2}$  et les estimateurs d'*Huber* correspondent à des estimateurs du Point Fixe définis ci-dessous.. Pour une valeur intermédiaire de  $q$ , les estimateurs d'*Huber* pourraient être interprétés comme un mélange entre le Point Fixe et la SCM classique. Les valeurs de la forme quadratique inférieures à  $k^2$  sont conservées et traitées comme avec les estimateurs SMV et SCM (correspondant à la première sommation); et les valeurs de  $t_i^2$  supérieures à  $k^2$  sont atténués par la fonction de pondération de manière similaire à l'approche Point Fixe. Dans un contexte gaussien complexe, il peut être établi que lorsque  $N$  tend vers l'infini, la proportion de données traitées avec le SCM est égale au paramètre  $q$ .

**0.4.2.1.1.2 Estimateurs du Point Fixe** Les estimateurs du Point Fixe (FPE), selon la définition proposée par Tyler [Tyler \(1987\)](#), satisfont les équations suivantes:

$$\hat{\boldsymbol{\mu}}_{FP} = \frac{\sum_{i=1}^N \frac{\mathbf{x}_i}{\left( (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP})^T \hat{\boldsymbol{\Sigma}}_{FP}^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP}) \right)^{1/2}}}{\sum_{i=1}^N \frac{1}{\left( (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP})^T \hat{\boldsymbol{\Sigma}}_{FP}^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP}) \right)^{1/2}}} \quad (41)$$

$$\hat{\boldsymbol{\Sigma}}_{FP} = \frac{m}{N} \sum_{i=1}^N \frac{(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP})(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP})^T}{\left( (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP})^T \hat{\boldsymbol{\Sigma}}_{FP}^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP}) \right)} \quad (42)$$

qui sont des cas particuliers de (38) pour  $u_1(t) = t^{-1}$  et  $u_2(t^2) = mt^{-2}$ . Ils sont définis par les équations de point fixe et peuvent être facilement calculés en utilisant

un algorithme récursif. Les FPE ont été largement étudiés en statistiques et dans la littérature de traitement du signal. Nous nous référons à [Pascal et al. \(2008b\)](#) pour une analyse détaillée de la performance.

Il est intéressant de souligner que  $\hat{\Sigma}_{SCM}$  et  $\hat{\Sigma}_{FP}$  ont la même distribution gaussienne asymptotique qui diffère sur leur second moment par un facteur  $\frac{m+1}{m}$ , i.e. pour  $N$  suffisamment grand,  $\hat{\Sigma}_{FP}$  se comporte comme une matrice de Wishart avec  $\frac{m}{m+1}N$  degrés de liberté. Remarquons que la distribution de  $\hat{\mathbf{M}}_{FP}$  ne dépend pas de la distribution elliptique spécifique. Afin d'assurer la consistance et la gaussianité asymptotique, la répartition de la population ne peut pas être trop fortement concentrée autour du centre. La consistance et la distribution asymptotique de  $\hat{\mathbf{M}}_{FP}$  sont démontrées pour l'estimation conjointe du vecteur de position et la matrice de dispersion dans le cas réel dans [Tyler \(1987\)](#).

### 0.4.3 Détection non-gaussienne adaptative

L'objectif de cette section est d'étendre les détecteurs gaussiens classiques pour le cadre non-gaussien. Cette section présente la contribution la plus importante de ce travail.

#### 0.4.3.1 Adaptive Normalized Matched Filter

Si le fond ne respecte pas l'hypothèse gaussienne, les performances du détecteur peuvent être détériorées, augmentant le taux de fausses alarmes. Pour tenir compte de l'hétérogénéité et de la non-gaussianité pour la modélisation du fond, une solution possible est d'utiliser le test ANMF construit avec des estimateurs robustes.

Nous remplaçons la matrice de covariance et le vecteur moyenne par des MLE ou des  $M$ -estimateurs robustes de dispersion et de position car ils sont des estimateurs consistants de la matrice de covariance et le vecteur moyenne dans la classe des distributions elliptiques (RVG en deux étapes). Ainsi, le ANMF prend la forme (voir Eq. (25)):

$$\Lambda_{ANMF, \hat{\Sigma}, \hat{\mu}} = \frac{|\mathbf{p}^H \hat{\Sigma}_N^{-1} (\mathbf{x} - \hat{\mu}_N)|^2}{(\mathbf{p}^H \hat{\Sigma}_N^{-1} \mathbf{p}) \left( (\mathbf{x} - \hat{\mu}_N)^H \hat{\Sigma}_N^{-1} (\mathbf{x} - \hat{\mu}_N) \right)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda. \quad (43)$$

où  $\hat{\mu}_N$  et  $\hat{\Sigma}_N$  dénotent tout couple de  $M$ -estimateurs et où  $N$  souligne la dépendance avec le nombre de données secondaires. Lorsque les  $M$ -estimateurs sont utilisés conjointement avec le ANMF, la fausse-alarme peut être réglée selon la proposition suivante.

**Proposition 0.4.1.** *La relation théorique entre la PFA et le seuil pour l'ANMF, construite avec des  $M$ -estimateurs  $\hat{\mu}_N$  et  $\hat{\Sigma}$ , est donnée par:*

$$PFA_{ANMF, \hat{\Sigma}, \hat{\mu}} = (1 - \lambda)^{a-1} {}_2F_1(a, a-1; b-1; \lambda), \quad (44)$$

avec  $a = \sigma_1(N-1) - m + 2$  et  $b = \sigma_1(N-1) + 2$ , où  $N$  est le nombre de données secondaires et  $m$  la dimension des vecteurs.  $\sigma_1$  est liée au choix particulier des  $M$ -estimateurs et obtenue selon:

$$\sigma_1 = \frac{\mathbb{E}[\psi_2^2(\sigma t^2)]}{m(m+1)(1 + [m(m+1)]^{-1} \mathbb{E}[\sigma t^2 \psi_2'(\sigma t^2)])^2}. \quad (45)$$

Bien que les FPE n'appartiennent pas à la classe des  $M$ -estimateurs (car ils ne satisfont pas les conditions de Maronna), ces résultats peuvent également être étendues à la FPE. La "PFA-seuil" approximée est obtenue en remplaçant dans Eq. (26)  $N-1$  par  $\frac{m}{m+1}(N-1)$  puisque  $\sigma_1 = \frac{m+1}{m}$  qui est une extension de [Pascal et al. \(2006\)](#) pour le vecteur moyenne inconnu.

Comme le fond est non-gaussien et/ou hétérogène, la distribution statistique de l'ANMF construit avec la SCM ne peut pas être prédite mais elle va sûrement varier avec le fond. L'ANMF construit avec les  $M$ -estimateurs (et en particulier les FPE) et en particulier il permet de surmonter la non-gaussianité et/ou de l'hétérogénéité des données. Cela implique, grâce aux propriétés de la CE, que le détecteur se comporte selon la même distribution quelle que soit la vraie CE, c'est à dire qu'elle est indépendante de la distribution (voir [Ollila and Tyler \(2012\)](#)).

#### 0.4.3.2 Simulations

La régulation de la FA est présentée ici pour le test ANMF.

##### 0.4.3.2.0.3 Analyse gaussienne

Les expériences ont été menées sur des vecteurs gaussiens de dimension  $m = 10$  pour  $N = 50$  données secondaires et les calculs ont été faits par  $10^6$  essais de Monte-Carlo. La vraie covariance est choisie comme une matrice de Toeplitz dont les entrées sont  $\Sigma_{i,j} = \rho^{|i-j|}$  et  $\rho = 0.4$ . Le vecteur moyenne est réglé pour avoir toutes les entrées égales à  $(3 + 4j)$ .

La Fig. 4 présente la régulation FA pour la ANMF sous l'hypothèse gaussienne, tant pour les FPE comme pour les SMC-SMV. La superposition entre relation théorique (courbe verte) et des données simulées (courbe jaune) valide les résultats de la Proposition 0.4.1 et il montre que le facteur de correction  $\sigma_1$  (Eq. (45)) permet de réguler parfaitement la FA, même dans le contexte gaussien.

##### 0.4.3.2.0.4 Analyse non-gaussienne



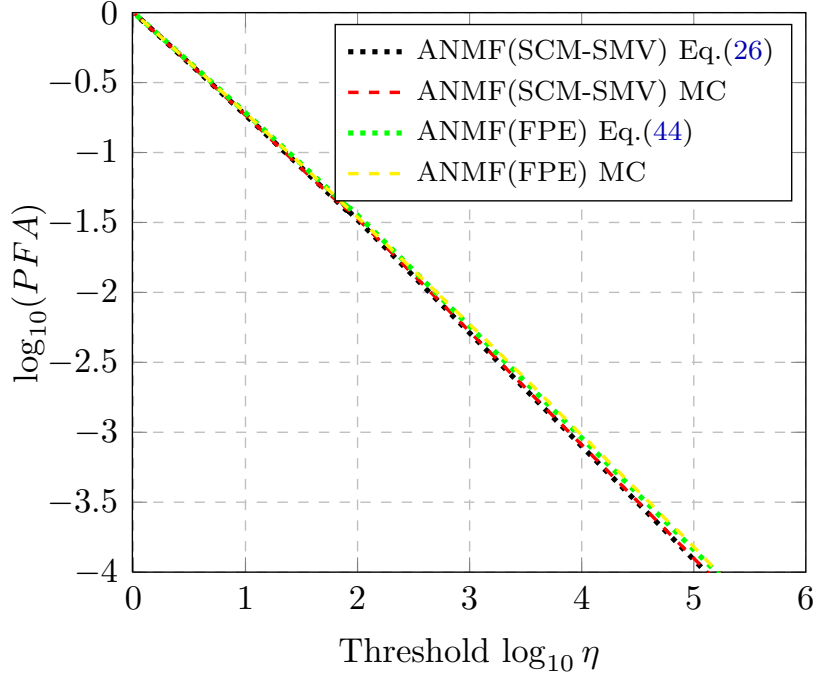


Figure 4: PFA par rapport au seuil pour l’ANMF sous une distribution gaussienne pour  $m = 10$  et  $N = 50$  lorsque (1) les SCM-SMV sont utilisés (courbes rouges et noires) (2) Proposition 0.4.1: les FPE sont utilisés (courbes jaunes et vertes).

Dans cette section, les expériences ont été réalisées sur une  $K$ -distribution avec le paramètre de forme  $\nu = 0.5$  pour des vecteurs gaussiens de dimension  $m = 10$ ,  $N = 50$  données secondaires et les calculs ont été faits par  $10^6$  essais de Monte-Carlo. Sous une  $K$ -distribution, comme représenté sur la Fig. 5, la relation théorique “PFA-seuil” dans l’Eq. (44) est superposée avec les simulations de Monte-Carlo pour les FPE, tandis que pour les SCM-SMV, la relation théorique “PFA-seuil” obtenue dans la Section 0.3.1.3 n’est plus valide (puisque l’hypothèse gaussienne n’est plus respectée). Nous avons laissé la relation théorique “PFA-seuil” pour les estimateurs gaussiens (courbe noire) pour information. Notez que sur les deux contextes gaussien et  $K$ -distribution, la régulation de la FA pour les FPE conduit aux mêmes résultats. Ainsi, la courbe ne dépend que de la taille du vecteur  $m$  et du nombre de données secondaire  $N$ . Ce fait met l’accent sur l’invariance maximale obtenue avec le ANMF construit avec les FPE, c’est à dire la distribution du détecteur sous l’hypothèse de cibles absentes reste la même pour toutes les différentes distributions impulsives dans la classe de distributions CE.

#### 0.4.4 Détection des anomalies

Le détecteur d’anomalies de Kelly présente l’avantage suivant: le vecteur moyenne, la matrice de covariance et le vecteur d’observation sont indépendants les uns des autres. Ce n’est pas le cas pour le Kelly généralisé ou le RXD classique. Cela

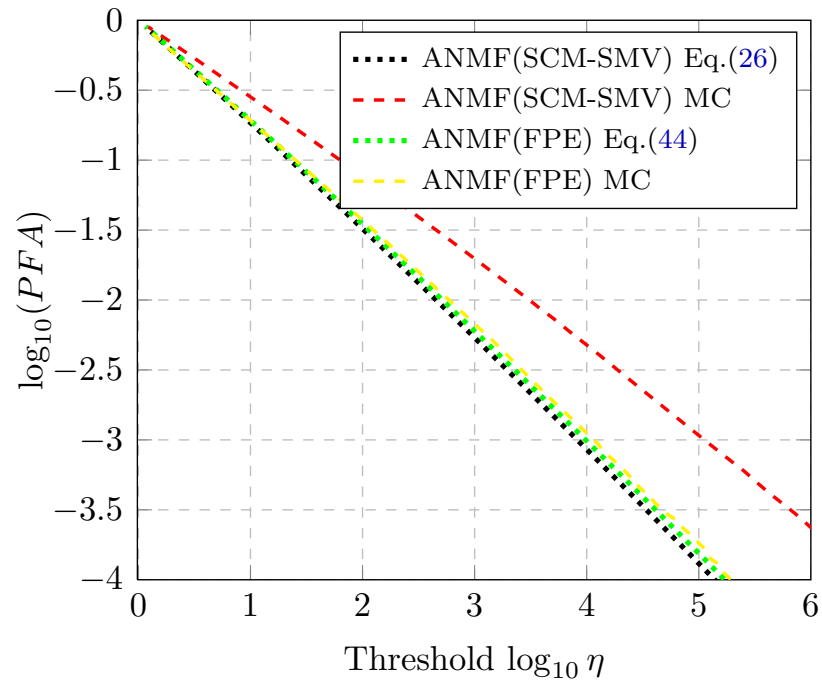


Figure 5: PFA par rapport au seuil pour l'ANMF sous une  $K$ -distribution avec le paramètre de forme  $\nu = 0.3$  pour  $m = 10$  and  $N = 50$  lorsque (1) les SCM-SMV sont utilisés (courbes rouges et noires) (2) Proposition 0.4.1: les FPE sont utilisés (courbes jaunes et vertes).

permet de remplacer les paramètres inconnus par un MLE où des  $M$ -estimateurs et le détecteur peut être écrit ainsi:

$$\Lambda_{\text{KellyAD}, \hat{\Sigma}, \hat{\mu}} = (\mathbf{x} - \hat{\mu}_N)^T \hat{\Sigma}_N^{-1} (\mathbf{x} - \hat{\mu}_N) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda, \quad (46)$$

et  $\hat{\mu}_N$  et  $\hat{\Sigma}_N^{-1}$  sont les mêmes que dans Eq. (38). Il est important de souligner que la distribution de ce détecteur est encore une question ouverte. En fait, il va sûrement dépendre de la distribution CE sous-jacente, à savoir la distribution va changer avec le choix de  $h_m(\cdot)$ .

## 0.5 Application à l'imagerie hyperspectrale

Les propositions théoriques ont été analysées dans les chapitres précédents à l'aide de simulations. Ici, nous discutons les résultats obtenus dans de vraies images hyperspectrales. Les schémas de détection sont évalués pour l'image Blind Test HYMAP.

### 0.5.1 Image Blind Test HYMAP

Nous illustrons ici la non-gaussianité d'une image hyperspectrale. La scène analysée est le "blind test" fournie dans [Snyder et al. \(2008\)](#). L'image a été collectée en Juillet 2006, dans les alentours de la petite ville de Cooke City, Montana, USA. L'image hyperspectrale a été recueillie par le capteur HYMAP exploité par HyVista. L'image HYMAP a une résolution au sol d'environ 3 mètres. Les véhicules civils et des petits panneaux de tissu ont été utilisés comme cibles. L'image est constituée de  $280 \times 800$  pixels et 126 bandes spectrales. La composition de la couleur réelle de la scène est représentée sur Fig. 6.



Figure 6: Composition de la vraie couleur de la scène HYMAP.

Les diagrammes de dépassement sont essentiellement des histogrammes cumulatifs des valeurs de distance de Mahalanobis, et ils fournissent un moyen utile pour visualiser les queues lourdes de distributions multivariées. La Fig. 7 présente les

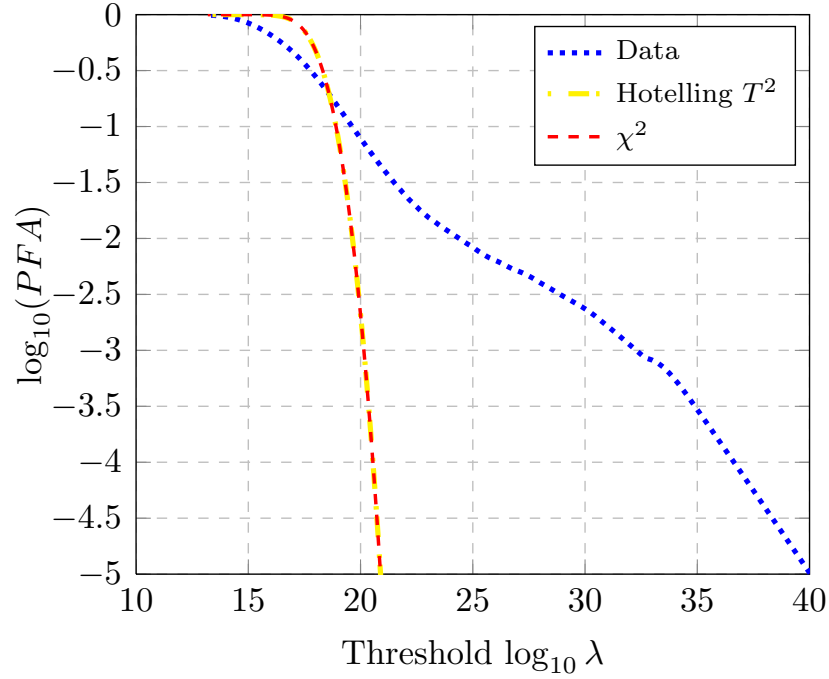


Figure 7: Fonction de répartition complémentaire de la distance de Mahalanobis  $\mathbb{P}(T^2 > \lambda)$  pour l'image HYMAP (bleu) et la relation théorique pour la distribution normale multivariée (jaune et rouge).

résultats théoriques lorsque la distribution gaussienne est présumée et la distribution des données hyperspectrales expérimentales qui dévie du comportement attendu. Cela suggère la nécessité des distributions non-gaussiennes pour modéliser le fond pour l'imagerie hyperspectrale.

#### 0.5.1.1 Régulation de la Fausse Alarme

Puisque les données hyperspectrales sont réelles et positives, nous proposons d'utiliser un filtre de Hilbert pour les rendre complexes et faire correspondre la distribution des détecteurs de cibles. Pour éviter le problème bien connu de la haute dimension, nous avons choisi séquentiellement huit bandes dans la représentation complexe. Dans cette approche, la matrice de covariance et le vecteur moyenne sont tous les deux estimés en utilisant une fenêtre glissante de taille  $11 \times 9$ , ayant  $N = 98$  des données secondaires.

Nous montrons dans la Fig. 8 le résultat de la méthode de détection ANMF classique construit avec les SMV-SCM (courbe rouge) et les FPE robustes (courbe jaune). En outre, les relations théoriques “PFA-seuil” sont représentées pour les estimateurs SMV-SCM (courbe noire) et les FPE (courbe verte). Les résultats obtenus sur des données réelles HSI sur une région non-gaussienne sont en accord avec la relation théorique calculée pour les FPE, tandis que la distribution du détecteur construit avec les SMV-SCM ne suit plus la relation théorique “PFA-seuil” pour un fond

non-gaussien. Remarquons que la distribution du détecteur sous l'hypothèse nulle ne dépend pas du matériau sous-jacent, pour l'ANMF-FPE. En outre, la distribution du détecteur peut dévier du comportement prévu dès lors que les estimateurs gaussiens sont utilisés. Cela souligne l'intérêt des estimateurs robustes dans la détection de cibles pour HSI et suggère son utilisation dans d'autres applications HSI où la matrice de covariance et le vecteur moyenne sont inconnus et doivent être estimés à partir du fond (e.g. classification, unmixing, etc.).

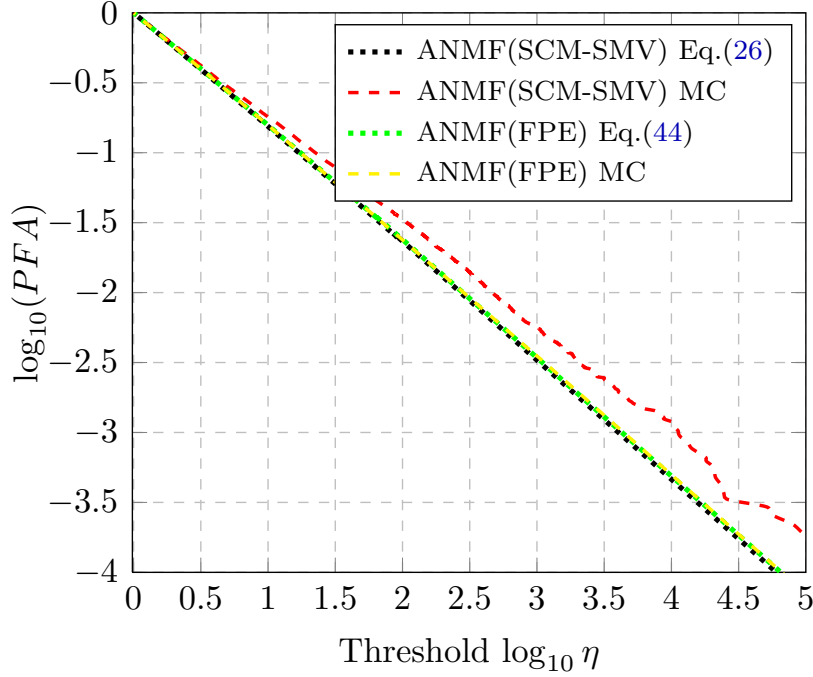


Figure 8: PFA par rapport au seuil pour l'ANMF pour une scène HSI réelle et  $m = 8$  et  $N = 98$  (1) les SCM-SMV sont utilisés (courbes rouges et noires) et (2) les FPE sont utilisés (courbes jaunes et vertes).

### 0.5.1.2 Détection des anomalies

Nous présentons des résultats pour la détection d'anomalies sur une image hyperspectrale réelle dans laquelle le fond ne peut pas être caractérisé avec une distribution gaussienne et des cibles artificielles ont été introduites comme des anomalies.

#### 0.5.1.2.1 Détecteur d'anomalies de Kelly robuste

Le jeu de données d'origine consiste à  $50 \times 50$  pixels avec 126 bandes. Ce jeu correspond à la région en haut à gauche de la scène dans Fig. 6 à partir de laquelle nous avons choisi séquentiellement neuf bandes. Pour les cibles artificielles avec signature spectrale connue introduites comme des anomalies dans le fond, nous avons utilisé le spectre de la Fig. 10 (a) comme endmember. Il caractérise le matériel de tissu de la Fig. 10 (b) mesurée dans un laboratoire et disponible dans le projet de blind test.

La Fig. 9 montre la position et la forme des cibles considérées. Pour cet exemple, la matrice de covariance et le vecteur moyenne sont estimés en utilisant une fenêtre glissante de taille  $9 \times 9$  ayant  $N = 80$  de données secondaires.

Les résultats pour le  $\Lambda_{Kelly\ AD}$  construit avec les SMV-SCM, les FPE et les estimateurs de shrinkage sont présentés dans la Fig. 11, la FA est fixée à une valeur de  $PFA = 0.03$ . Le paramètre de régularisation  $\beta$  a été fixé à  $\beta = 0.8$  pour les estimateurs de shrinkage. Nous avons optimisé empiriquement les résultats sur  $\beta$  menant à la valeur choisie. Dans ce cas, les FPE et notamment les shrinkage FPE sont capables de localiser les cibles artificielles et présentent une diminution du nombre de fausses alarmes. Cette amélioration est due au fait que les FPE traitent les valeurs aberrantes et les échantillons impulsifs afin pour eux d'avoir une moindre contribution au processus de caractérisation du fond, tandis que les estimateurs SMV-SCM (et sa version diagonal loaded) souffrent de la présence de pixels fortement réfléchants dans les données secondaires. Remarquez que les shrinkage FPE permettent une meilleure détection par rapport aux FPE.

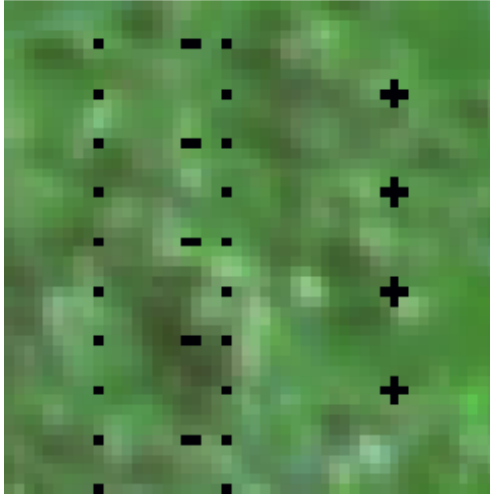
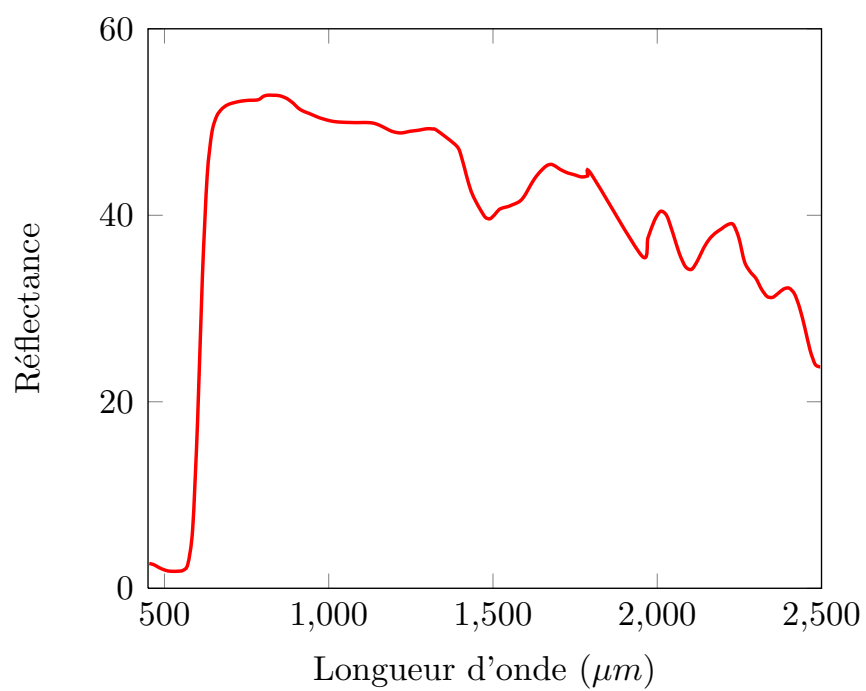


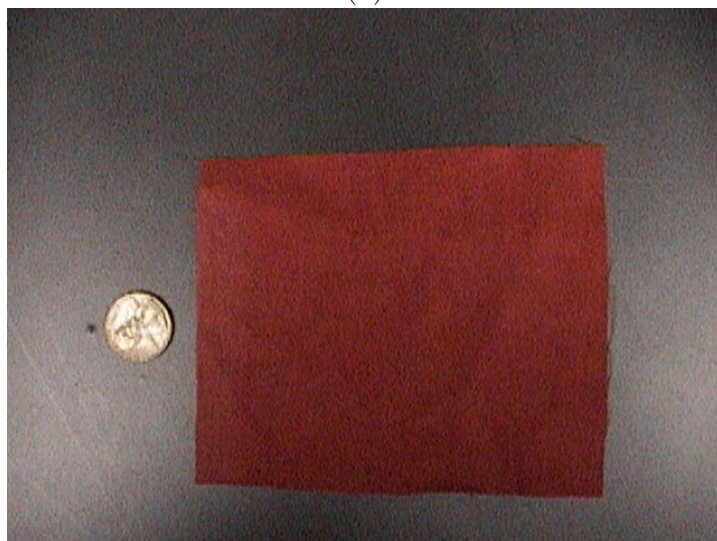
Figure 9: Jeu de données d'origine et la position des anomalies artificielles.

## 0.6 Conclusion

Des détecteurs de cibles classiques sont habituellement obtenus dérivant le rapport de vraisemblance sous hypothèse gaussienne et remplaçant des paramètres inconnus du fond par leurs estimations. Dans la plupart des applications, des signaux d'interférence sont supposés être gaussiens de moyenne nulle ou avec un vecteur moyenne connu qui peut être enlevé et de matrice de covariance inconnue. Lorsque le vecteur moyenne est inconnu, il doit être estimé conjointement avec la matrice de covariance, comme c'est le cas par exemple dans l'imagerie hyperspectrale. Dans ce travail, les versions adaptatives du filtre adapté classique et du filtre adapté normalisé, ainsi que deux versions du détecteur Kelly sont calculées puis sont analysées dans le



(a)



(b)

Figure 10: (a) Endmember utilisé dans l'expérience qui correspond à (b) un panneau de tissu.

cas où le vecteur moyenne du fond n'est pas connu. Plus précisément, des expressions théoriques pour la régulation de la fausses alarme sont dérivées et la propriété TFAC est vérifiée pour permettre au détecteur d'être indépendant des paramètres de nuisance. Il est intéressant de souligner que ces expressions sont essentielles pour régler automatiquement le seuil du détecteur. Dans le cas contraire, des méthodes numériques ou des simulations de Monte-Carlo doivent être utilisées conduisant à des résultats moins précis.

Cette thèse propose également un nouveau type de détecteurs robustes permettant de surmonter la non-gaussianité et l'hétérogénéité des données hyperspectrales réels. Lors de la prise en compte de l'hétérogénéité et de la non-gaussianité, les distributions elliptiques offrent des modèles fiables pour la caractérisation du fond. Grâce à cette hypothèse non-gaussienne, ce travail met en évidence le fait que les procédures d'estimation robustes sont une alternative intéressante aux estimateurs gaussiens classiques. Les  $M$ -estimateurs pour le vecteur moyenne et la matrice de dispersion sont décrits. Les performances des estimateurs robustes ont été étudiées. Cette analyse révèle la supériorité de  $M$ -estimateurs dans un environnement non-gaussien. L'objectif de cette thèse est alors non seulement de rappeler les méthodes bien connues pour la détection de cibles, mais aussi de proposer des moyens de les étendre au cadre non-gaussien. En outre, les expressions théoriques pour la régulation de la fausse alarme sont prévus et la propriété TFAC est atteinte dans la classe des distributions elliptiques.



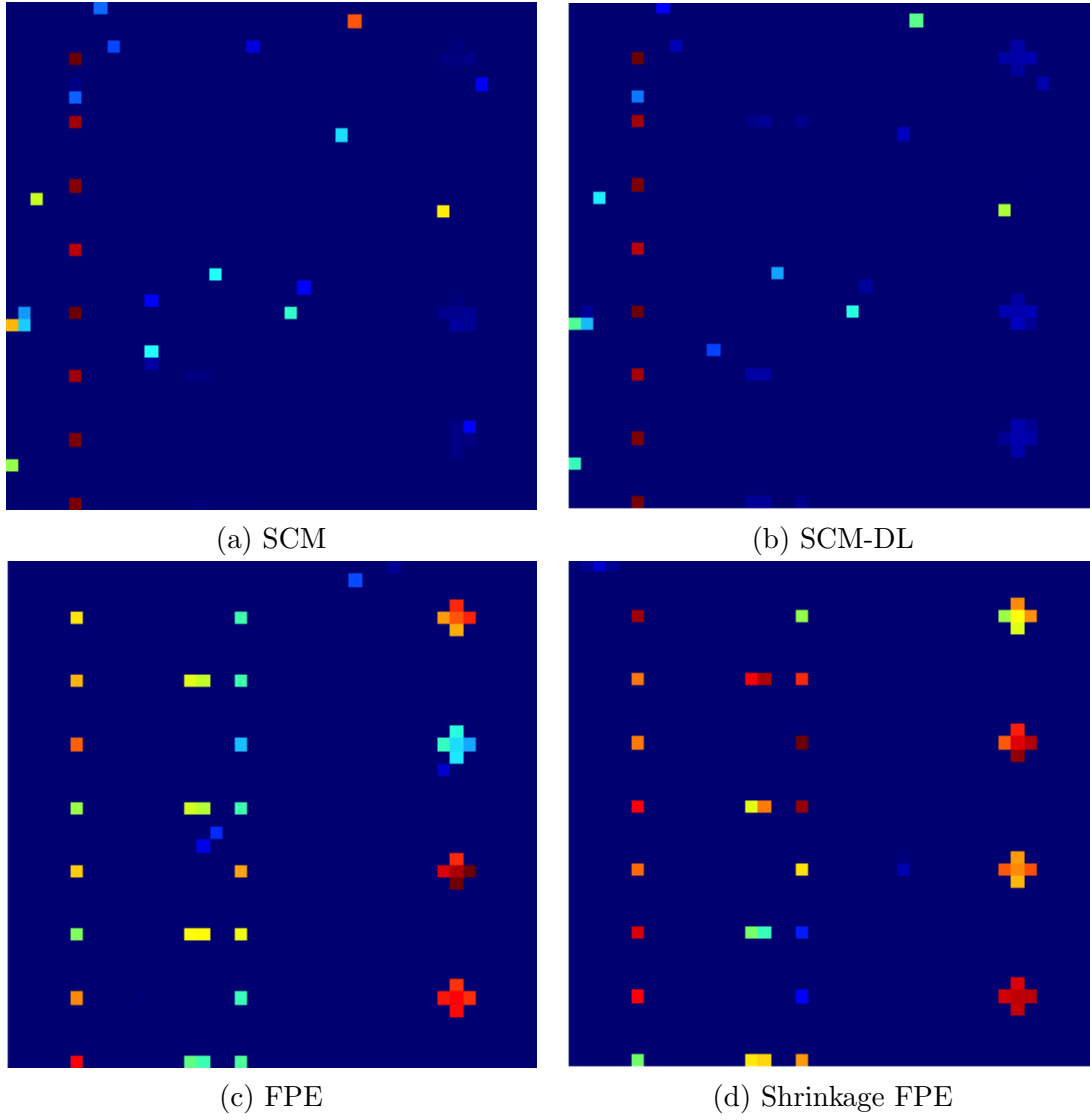


Figure 11: Kelly AD construit avec des estimateurs robustes pour cibles artificielles dans une HSI réelle.





# Contents

<b>Résumé</b>	<b>vii</b>
0.1 Introduction . . . . .	vii
0.2 État de l’art . . . . .	vii
0.2.1 Notions préliminaires . . . . .	viii
0.2.2 Techniques classiques de détection de cibles . . . . .	ix
0.3 Détection de cibles gaussienne . . . . .	xii
0.3.1 Détection Gaussienne de moyenne différent de zéro . . . . .	xii
0.3.2 Détection des anomalies . . . . .	xv
0.4 Détection de cibles non-gaussienne . . . . .	xviii
0.4.1 Distributions elliptiques . . . . .	xix
0.4.2 Estimation robuste . . . . .	xix
0.4.3 Détection non-gaussienne adaptative . . . . .	xxii
0.4.4 Détection des anomalies . . . . .	xxiv
0.5 Application à l’imagerie hyperspectrale . . . . .	xxvi
0.5.1 Image Blind Test HYMAP . . . . .	xxvi
0.6 Conclusion . . . . .	xxix
<b>Contents</b>	<b>xxxv</b>
<b>List of Figures</b>	<b>xxxix</b>
<b>Introduction</b>	<b>1</b>
<b>1 State of the art</b>	<b>5</b>
1.1 Hyperspectral Imaging . . . . .	5
1.2 Target Detection and Anomaly Detection . . . . .	8
1.3 Dimensionality Reduction . . . . .	9
1.4 Preliminary Notions . . . . .	11
1.5 Classical Target Detection Techniques . . . . .	13
1.5.1 Adaptive Matched Filter . . . . .	15
1.5.2 Adaptive Kelly detector . . . . .	17
1.5.3 Adaptive Normalized Matched Filter . . . . .	18
1.6 Summary . . . . .	19

<b>2</b>	<b>Target Detection in Gaussian Background</b>	<b>21</b>
2.1	Introduction . . . . .	21
2.2	Adaptive non-zero mean Gaussian Detection . . . . .	22
2.2.1	Adaptive Matched Filter Detector . . . . .	22
2.2.2	Kelly Detector . . . . .	24
2.2.3	Adaptive Normalized Matched Filter . . . . .	29
2.2.4	Simulations . . . . .	30
2.3	Anomaly Detection . . . . .	34
2.3.1	Reed-Xiaoli Detector . . . . .	34
2.3.2	Kelly Anomaly Detector . . . . .	37
2.3.3	Normalized-RXD and Uniform Target Detector . . . . .	41
2.3.4	Generalized Kelly Anomaly Detector . . . . .	42
2.3.5	Simulations . . . . .	42
2.4	Summary . . . . .	44
<b>3</b>	<b>Target Detection in non-Gaussian Background</b>	<b>47</b>
3.1	Introduction . . . . .	47
3.2	Elliptically Contoured Distributions . . . . .	49
3.2.1	Definition . . . . .	49
3.2.2	The Class of Spherically Invariant Random Vectors . . . . .	51
3.2.3	Examples of Complex Elliptical Distributions . . . . .	52
3.3	Robust Estimation Procedures . . . . .	54
3.3.1	Maximum Likelihood Estimates . . . . .	55
3.3.2	$M$ -Estimators . . . . .	56
3.3.3	Robust shrinkage estimators . . . . .	63
3.3.4	Simulations . . . . .	65
3.4	Adaptive non-Gaussian Detection . . . . .	69
3.4.1	Adaptive Normalized Matched Filter . . . . .	70
3.4.2	Simulations . . . . .	72
3.5	Anomaly Detection . . . . .	82
3.5.1	Robust Kelly Anomaly Detector . . . . .	82
3.5.2	Binary Partition Trees . . . . .	83
3.6	Summary . . . . .	85
<b>4</b>	<b>Application to Hyperspectral Imaging</b>	<b>87</b>
4.1	Introduction . . . . .	87
4.2	NASA Hyperion Image . . . . .	88
4.2.1	False Alarm Regulation . . . . .	88
4.2.2	Performance Evaluation . . . . .	89
4.2.3	Anomaly Detection . . . . .	93
4.3	Blind Test HYMAP Image . . . . .	98
4.3.1	False Alarm Regulation . . . . .	101
4.3.2	Anomaly Detection . . . . .	103
4.4	Rural scene . . . . .	109
4.4.1	False Alarm Regulation . . . . .	109

4.4.2	Anomaly Detection . . . . .	109
4.5	MUSE project . . . . .	113
4.6	Summary . . . . .	115
<b>Conclusion</b>		<b>117</b>
<b>A A class of robust estimates for detection in HSI</b>		<b>121</b>
<b>B Adaptive non-Zero Mean Gaussian Detection</b>		<b>127</b>
<b>C Hyperspectral Anomaly Detectors</b>		<b>139</b>
<b>Bibliography</b>		<b>153</b>



# List of Figures

1	PFA par rapport au seuil de l'AMF lorsque (1) $\mu$ et $\Sigma$ sont connus (MF) (courbes rouges et noires) (2) seulement $\mu$ est connue (courbes grises et bleues) (3) Proposition 0.3.1: les deux $\mu$ et $\Sigma$ sont inconnus (courbes jaunes et vertes). . . . .	xv
2	PFA par rapport au seuil de le détecteur "plug-in" de Kelly lorsque (1) seulement $\mu$ est connue (courbes grises et bleues) (2) Proposition 0.3.2: les deux $\mu$ et $\Sigma$ sont inconnus (courbes jaunes et vertes). . . .	xvi
3	PFA par au rapport seuil de l'ANMF lorsque (1) $\mu$ et $\Sigma$ sont connus (NMF) (courbes rouges et noires) (2) seulement $\mu$ est connue (courbes grises et bleues) (3) Proposition 0.3.3: les deux $\mu$ et $\Sigma$ sont inconnus (courbes jaunes et vertes). . . . .	xvi
4	PFA par rapport au seuil pour l'ANMF sous une distribution gaussienne pour $m = 10$ et $N = 50$ lorsque (1) les SCM-SMV sont utilisés (courbes rouges et noires) (2) Proposition 0.4.1: les FPE sont utilisés (courbes jaunes et vertes). . . . .	xxiv
5	PFA par rapport au seuil pour l'ANMF sous une $K$ -distribution avec le paramètre de forme $\nu = 0.3$ pour $m = 10$ and $N = 50$ lorsque (1) les SCM-SMV sont utilisés (courbes rouges et noires) (2) Proposition 0.4.1: les FPE sont utilisés (courbes jaunes et vertes). . . . .	xxv
6	Composition de la vraie couleur de la scène HYMAP. . . . .	xxvi
7	Fonction de répartition complémentaire de la distance de Mahalanobis $\mathbb{P}(T^2 > \lambda)$ pour l'image HYMAP (bleu) et la relation théorique pour la distribution normale multivariée (jaune et rouge). . . . .	xxvii
8	PFA par rapport au seuil pour l'ANMF pour une scène HSI relle et $m = 8$ et $N = 98$ (1) les SCM-SMV sont utilisés (courbes rouges et noires) et (2) les FPE sont utilisés (courbes jaunes et vertes). . . . .	xxviii
9	Jeu de données d'origine et la position des anomalies artificiels. . . . .	xxix
10	(a) Endmember utilisé dans l'expérience qui correspond à (b) un panneau de tissu. . . . .	xxx
11	Kelly AD construit avec des estimateurs robustes pour cibles artificielles dans une HSI réelle. . . . .	xxxii
1.1	Datacube for the Pavia scene. . . . .	6
1.2	Image at a single wavelength (45 <sup>th</sup> band). . . . .	7
1.3	Spectra of one of the pixels in the image (pixel (1,1)). . . . .	8



1.4	Sliding window mask over the observation vector (green dot) with a guard window (red) and useful secondary data (blue). . . . .	15
2.1	PFA versus threshold for the AMF when (1) $\mu$ and $\Sigma$ are known (MF) (red and black curves) (2) only $\mu$ is known (gray and blue curves) (3) Proposition 2.2.1: both $\mu$ and $\Sigma$ are unknown (yellow and green curves). . . . .	31
2.2	PFA versus threshold for the "plug-in" Kelly detector when (1) only $\mu$ is known (gray and blue curves) (2) Proposition 2.2.2: both $\mu$ and $\Sigma$ are unknown (yellow and green curves) . . . . .	31
2.3	PFA versus threshold for the ANMF when (1) $\mu$ and $\Sigma$ are known (NMF) (red and black curves) (2) only $\mu$ is known (gray and blue curves) (3) Proposition 2.2.3: both $\mu$ and $\Sigma$ are unknown (yellow and green curves) . . . . .	32
2.4	Probability of detection for different SNR values and $PFA = 10^{-3}$ in Gaussian case. . . . .	33
2.5	Graphical interpretation in two-dimensional space for the $\Lambda_{KellyAD}$ detector. . . . .	39
2.6	PFA versus threshold for the $\Lambda_{KellyAD}$ when (1) $\mu$ and $\Sigma$ are known (Mahalanobis) (red and black curves) (2) only $\mu$ is known (gray and blue curves) (3) Proposition 2.3.1: both $\mu$ and $\Sigma$ are unknown (yellow and green curves). . . . .	43
2.7	Probability of detection for different SNR values and $PFA = 10^{-3}$ in Gaussian case. . . . .	44
3.1	Exceedance plots are essentially a cumulative histogram of Mahalanobis distance values, and they provide a useful way to visualize the fat tails of multivariate distributions. This figure shows the theoretical results when Gaussian distribution is assumed (yellow and red curves) and experimental hyperspectral data distribution which diverts from the expected behavior for Hyperion NASA image. . . . .	48
3.2	Comparison between probability density functions of Gaussian distribution and $t$ -distribution with different values of $\nu$ . . . . .	54
3.3	Empirical consistency of the FPE for the joint estimation problem. . . . .	60
3.4	Comparison of weighting functions for the SMV-SCM, Huber, $t$ -distribution and FPE with $m = 3$ . . . . .	61
3.5	Performance comparison for different estimators for samples drawn from Gaussian distribution for $m = 10$ and $\rho = 0.4$ . . . . .	66
3.6	Performance comparison for different estimators for samples drawn from Gaussian distribution for $N = 200$ and $\rho = 0.4$ . . . . .	67
3.7	Performance comparison for different estimators for samples drawn from Gaussian distribution for $m = 10$ and $N = 200$ . . . . .	67
3.8	Performance comparison for different estimators for samples drawn from $K$ -distribution with shape parameter $\nu = 0.5$ for $m = 10$ and $\rho = 0.4$ . . . . .	68

3.9	Performance comparison for different estimators for samples drawn from $K$ -distribution with shape parameter $\nu = 0.5$ for $N = 200$ and $\rho = 0.4$ . . . . .	69
3.10	Performance comparison for different estimators for samples drawn from $K$ -distribution with shape parameter $\nu = 0.5$ for $m = 10$ and $N = 200$ . . . . .	70
3.11	PFA-threshold for the ANMF under Gaussian distribution for $m = 10$ and $N = 50$ when (1) the SCM-SMV are used (red and black curves) (2) Proposition 3.4.1: the Huber estimates with $q = 0.75$ are used (purple and orange curves) . . . . .	73
3.12	PFA versus threshold for the ANMF under a Gaussian distribution for $m = 10$ and $N = 50$ when (1) the SCM-SMV are used (red and black curves) (2) Proposition 3.4.1: the Student $M$ -esimtators are used (blue and gray curves) . . . . .	74
3.13	PFA versus threshold for the ANMF under a Gaussian distribution for $m = 10$ and $N = 50$ when (1) the SCM-SMV are used (red and black curves) (2) Proposition 3.4.1: the FPE are used (yellow and green curves) . . . . .	75
3.14	PFA versus threshold for the ANMF under a Gaussian distribution for $m = 3$ and $N = 21$ for (1) the SCM-SMV estimates (red) (2) the FPE (yellow) and (3) the FPE built $m + 1/mN$ (dark green). . . . .	76
3.15	PFA-threshold under $K$ -distribution with shape parameter $\nu = 0.3$ for $m = 10$ and $N = 50$ when (1) the SCM-SMV are used (red and black) (2) Proposition 3.4.1: the Huber estimators with $q = 0.75$ are used (orange and purple) . . . . .	77
3.16	PFA-threshold under $K$ -distribution with shape parameter $\nu = 0.3$ for $m = 10$ and $N = 50$ when (1) the SCM-SMV are used (red and black) (2) Proposition 3.4.1: the Student $M$ -estimators are used (gray and blue curves) . . . . .	77
3.17	PFA versus threshold for the ANMF under a $K$ -distribution with shape parameter $\nu = 0.3$ for $m = 10$ and $N = 50$ when (1) the SCM-SMV are used (red and black curves) (2) Proposition 3.4.1: the FPE are used (yellow and green curves) . . . . .	78
3.18	PFA versus threshold for the ANMF under a Gaussian distribution for the FPE obtained with $N = 50$ secondary data and for different values of $m$ . . . . .	78
3.19	PFA versus threshold for the ANMF under a Gaussian distribution with $m = 5$ for the FPE and for different values of $N$ secondary data. . . . .	79
3.20	PFA versus threshold for the ANMF under a Gaussian distribution with $m = 3, N = 21$ for the FPE and for different values of $\rho$ secondary data CFAR MATRIX. . . . .	80
3.21	PFA versus threshold for the ANMF under a Gaussian distribution with $m = 3, N = 21$ for the FPE and for different values of $\mu$ secondary data. . . . .	80

3.22	PFA versus threshold for the ANMF under a Gaussian distribution with $m = 3$ and $N=21$ for the FPE for different underlying distributions Gaussian and $K$ -distribution with shape parameter $\nu$ . . . . .	81
3.23	Probability of detection for different SNR values and $PFA = 10^{-3}$ in non-Gaussian environment. . . . .	82
3.24	Construction of the Binary Partition Tree (BPT). . . . .	83
4.1	True color composition of the Hyperion scene. . . . .	89
4.2	Q-Q Plot of the data sample versus the Normal theoretical distribution. . . . .	90
4.3	Histogram of the data sample and comparison with the Gaussian probability density function. . . . .	90
4.4	AMF false-alarm regulation for a real HSI image . . . . .	91
4.5	Kelly false-alarm regulation for a real HSI image . . . . .	91
4.6	ANMF false-alarm regulation for a real HSI image . . . . .	92
4.7	Q-Q Plot of data sample versus the Normal theoretical distribution. . . . .	92
4.8	Probability of detection for different SNR values and $PFA = 10^{-2}$ for real HSI data. . . . .	93
4.9	Kelly AD complementary CDF of the Mahalanobis distance for a real HSI image . . . . .	94
4.10	Endmember used in the experiment which corresponds to land materials. . . . .	94
4.11	Anomaly detection for artificial targets in real HSI. . . . .	96
4.12	Kelly AD built with robust estimates for artificial targets in real HSI. . . . .	97
4.13	True color composition of the HyMap scene. . . . .	98
4.14	Q-Q Plot of data sample versus the Normal theoretical distribution. . . . .	99
4.15	Histogram of the data sample and comparison with the Gaussian probability density function. . . . .	100
4.16	Complementary CDF of the Mahalanobis distance $\mathbb{P}(T^2 > \lambda)$ for the Hymap image (blue) and the theoretical relationship for multivariate Normal distribution (yellow and red). . . . .	100
4.17	PFA versus threshold for the ANMF for real HSI scene and $m = 8$ and $N = 98$ when (1) the SCM-SMV are used (red and black curves) (2) the FP estimates are used (yellow and green curves). . . . .	102
4.18	Original data set and position of the artificial anomalies. . . . .	103
4.19	(a) Endmember used in the experiment which corresponds to (b) a fabric panel. . . . .	104
4.20	Kelly AD built with robust estimates for artificial targets in real HSI. . . . .	105
4.21	Location of the targets in the image. . . . .	106
4.22	PD-PFA plot comparing classical sliding window against BPT strategy . . . . .	107
4.23	Detection maps for sliding window using (a) Gaussian estimates, (b) FPE and (c) BPT-approach . . . . .	108
4.24	True color composition of the Rural scene. . . . .	110
4.25	PFA versus threshold for the ANMF for real HSI scene and $m = 8$ and $N = 98$ when (1) the SCM-SMV are used (red and black curves) (2) the FP estimates are used (yellow and green curves). . . . .	111
4.26	Kelly AD built with (a) Gaussian SMV-SCM and (b) robust FPE. . . . .	112

4.27 Classical and Fixed-point anomaly detection in a hyperspectral image of $300 \times 300$ in 3578 channels. . . . .	114
--	-----



# Acronyms

<b>AD</b>	Anomaly Detection
<b>AMF</b>	Adaptive Matched Filter
<b>ANMF</b>	Adaptive Normalized Matched Filter
<b>BPT</b>	Binary Partition Trees
<b>CDF</b>	Cumulative Density Function
<b>CE</b>	Complex Elliptical distributions
<b>CFAR</b>	Constant False Alarm Rate
<b>CM</b>	Covariance Matrix
<b>CSCM</b>	Centered Sample Covariance Matrix
<b>FPE</b>	Fixed Point Estimates
<b>GLRT</b>	Generalized Likelihood Ratio
<b>HSI</b>	Hyperspectral Imaging
<b>IID</b>	Independent Identically Distributed
<b>LR</b>	Likelihood Ratio
<b>LRT</b>	Likelihood Ratio Test
<b>MF</b>	Matched Filter
<b>ML</b>	Maximum Likelihood
<b>MLE</b>	Maximum Likelihood Estimates
<b>MV</b>	Mean Vector
<b>NMSE</b>	Normalized Mean Square Error
<b>N-RXD</b>	Normalized Reed-Xiaoli Detector
<b>PCA</b>	Principal Component Analysis

<b>RXD</b>	Reed-Xiaoli Detector
<b>SCM</b>	Sample Covariance Matrix
<b>SMV</b>	Sample Mean Vector
<b>SNR</b>	Signal-to-Noise Ratio
<b>UTD</b>	Uniform Target Detector

# Notation

$\mathbb{R}$	Set of real numbers.
$\mathbb{C}$	Set of complex numbers.
$\mathbb{E}$	denotes the expectation,
$T$	denotes the transpose operator,
$H$	denotes the hermitian operator,
$*$	denotes the conjugate operator.
$\sim$	means “distributed as”,
$\stackrel{d}{=}$	means “has the same distribution as” .
$\xrightarrow{d}$	denotes the convergence in distribution,
$\otimes$	denotes the Kronecker product,
$\text{vec}$	denotes the operator that transforms a $m \times n$ matrix into a vector of dimension $mn$ by concatenating the columns,
$\mathcal{H}_0$	denotes the signal absent (or null) hypothesis,
$\mathcal{H}_1$	denotes the signal present (or alternative) hypothesis,
$\mathcal{H}_1$	decides $\mathcal{H}_1$ if the result is greater and $\mathcal{H}_0$ if the result is lower than a certain value,
$\mathcal{H}_0$	
$\mathbf{I}_m$	denotes the identity matrix of dimensions $m \times m$ ,
$\mathbf{0}_{m,n}$	denotes the zero matrix of dimensions $m \times n$ .
$\mathbf{K}$	is the commutation matrix such that $\mathbf{K}\text{vec}(\mathbf{A}) = \text{vec}(\mathbf{A}^T)$ .
$\Im(\mathbf{x})$	denotes the imaginary part of $\mathbf{x}$ ,
$\Re(\mathbf{x})$	denotes the real part of $\mathbf{x}$ ,
$\text{cov}(\mathbf{x})$	denotes the covariance of $\mathbf{x}$ ,
$\ \cdot\ _F$	denotes the Frobenius norm,
$\text{Tr}$	denotes the trace operator,
${}_2F_1(a, b; c; z)$	denotes the hypergeometric function.





# Introduction

Hyperspectral imaging (HSI) extends from the fact that for any given material, the amount of emitted radiation varies with wavelength. HSI sensors measure the radiance of the materials within each pixel area at a very large number of contiguous spectral bands and provide image data containing both spatial and spectral information (see [Manolakis et al. \(2003\)](#) for more details and references therein). A large number of hyperspectral detection algorithms have been derived and explored over the past years (see for e.g. [Chang \(2003\)](#), [Manolakis et al. \(2014\)](#)). Most of these methods rely on classical Gaussian distribution assumption and need for the statistical characterization of the background usually through first and second order parameters (i.e. the mean vector and the covariance matrix). Classical detection methods assume that background is zero mean or with known mean vector that can be exploited. However, when the mean vector is unknown, as it is the case for hyperspectral imaging, it has to be included in the detection process. We propose in this work an extension of classical detection methods for both covariance matrix and mean vector unknown.

However, the actual distribution of the background pixels may differ from the theoretically predicted under Gaussian hypothesis. In fact, as stated in [Manolakis and Marden \(2002\)](#), the empirical distribution usually has heavier tails compared to the Gaussian distribution, and these tails strongly influence the observed false-alarm rate of the detector. The class of elliptically contoured distributions has already been popularized for background characterization in HSI [Manolakis and Marden \(2002\)](#). Although non-Gaussian models are assumed for background modeling and detectors design, the parameters estimation is still performed using classical Gaussian-based estimators; for example the covariance matrix generally determined by the Sample Covariance Matrix and the mean vector with the Sample Mean Vector. We analyze here some robust estimation procedures ( $M$ -estimators introduced in [Huber \(1964\)](#), [Maronna \(1976\)](#), [Tyler \(1987\)](#)) more suitable when non-Gaussian distributions are assumed. Moreover, target detection methods are extended to the non-Gaussian distributed background providing a unified framework for target detection in HSI.

Many of these problems have been widely analyzed in radar detection domain. In this context, several authors have studied conventional detection schemes in Gaussian environment [Scharf and Friedlander \(1994\)](#), [Kraut et al. \(2001\)](#), [Kelly \(1986\)](#). These techniques constitute the basis for adaptive target detection for non-zero mean Gaussian distribution. In addition, the occurrence of impulsive noise has led to the development of techniques for treatment in case of non-Gaussian clutter (see for e.g.

Gini and Greco (2002), Pascal et al. (2008b)). These methods found in radar signal processing serve as a reference and starting point for this work. This document is organized as follows.

The first chapter introduces the State of the art. The main characteristics of HSI acquisition and hyperspectral data processing are described. In particular, target detection and anomaly detection techniques and their applications are detailed. Next, the basic preliminary notions on signal processing are recalled and classical target detection methods, widely investigated in detection theory and used all over this document, are described in the last section.

In Chapter 2, we derive the expression of each adaptive detector, the Adaptive Matched Filter, the Kelly's detection test and the Adaptive Normalized Matched Filter (ANMF), under the Gaussian assumption where both the mean vector and the covariance matrix are assumed to be unknown. The main contribution of this chapter is the exact derivation of the distribution of each proposed detection scheme under null hypothesis, i.e. when no target is supposed to be present. Thus, through Gaussian assumption, closed-form expressions for the false-alarm regulation are obtained, which allow to theoretically set the detection threshold for a given Probability of False Alarm (PFA). Furthermore, this chapter reviews classical anomaly detection schemes such as the widely spread Reed-Xiaoli Detector and some of its variations. Moreover, the Mahalanobis distance based detector is rigorously derived from a Kelly's test-based approach. Finally, this detector is analyzed and its exact distribution is derived when both mean vector and covariance matrix are unknown. The Kelly's detection test applied to anomaly detection problems constitute an original result of this work.

Chapter 3 provides a summary on elliptical distributions, their characterization and generation. We detail some examples of elliptical distributions and emphasize the fact that real hyperspectral data cannot be described only with the Gaussian model. In addition, this chapter introduces robust parameter estimation tools, the  $M$ -estimators, for the scatter matrix and location vector. The main contribution of this chapter and of this thesis is the analysis of target detection techniques in non-Gaussian distributed background. For this purpose, we describe the different target detection methods studied, the ANMF and the Kelly's anomaly detector. Further, we obtain theoretical expressions for false alarm regulation which highlight the Constant False Alarm Rate (CFAR) properties of the detection schemes. More precisely, the detectors' performance are analyzed over simulations and compared to those of the conventional Gaussian detectors. The key result of this work can be summarized as follows: The proposed detection methods jointly used with robust estimates allow not only to overcome the heterogeneity and non-Gaussianity of the data but also to reach the same performance than the conventional detector on homogeneous and Gaussian background. At last, we study the use of robust estimation methods in anomaly detectors leading to some great improvement in the detection process.

Chapter 4 represents the more applied contribution of this work. The different detection methods studied in Gaussian and non-Gaussian case are investigated for four hyperspectral images. False-alarm regulation expressions are validated for the different detection schemes, the AMF, the ANMF, classical Kelly's test and Kelly's anomaly detector on real data. Finally, the improvement brought by robust estimation procedures is studied for detection tasks in different scenarios.

We conclude with a summary of the results presented and some perspectives, new issues, opportunities and paths, this work suggests.



# 1

## State of the art

### 1.1 Hyperspectral Imaging

Hyperspectral imaging (HSI), also known as hyperspectral spectroscopy [Goetz et al. \(1985\)](#) provides rich information both spatially and spectrally. Hyperspectral spectrometers are a class of optical sensors that measure spectral radiance from every pixel in a scene. The result is the so-called hyperspectral image. The resulting spectra on each point of the image can contain more than 200 contiguous narrow channels covering visible and near-infrared wavelengths at a high spectral resolution [Green et al. \(1998\)](#). This fact enhances the capability of HSI to discriminate different materials through their unique spectral signature. With the increasing development of these sensors delivering high spatial and spectral resolution, HSI has gained interest in the last decades for target detection and recognition purposes [Bioucas-Dias et al. \(2013\)](#), [Plaza et al. \(2009\)](#). These sensors are usually deployed on satellites or aircrafts and are used for a large number of remote sensing applications: border surveillance, mine detection, agricultural monitoring, terrain exploration for minerals or soil types characterization, etc.

Hyperspectral images contain a wealth of data, but interpreting them needs for a depth comprehension of exactly what properties of ground materials we are trying to measure, and how they are related to the sensor measurements. The analysis of light that is emitted or reflected from different materials and the fluctuation of the amount of radiation with wavelength is known as spectroscopy. In optical remote sensing applications, one studies the spectrum of the sunlight that is propagated through the atmosphere and diffusely reflected, transmitted and/or absorbed by the ground materials on the Earth surface. Scene information is contained in the measured radiance as a function of continuous space, wavelength and time variables. In practice,

sensors have limitations due to finite-resolution recordings. In terms of geometrical properties of the imaging system, the spatial resolution of a sensor is given by the instantaneous field of view of the spectrometer. The field of view is a function of the optics of the sensor, the size of the detector components and the altitude. The spatial resolution of the sensor determines the size of smallest object that can be distinguished on the Earth surface and relates how well the spatial details can be described in the image. The spectral resolution is defined by the width of spectral bands used to measure the radiance at difference wavelengths. The table in 1.1 mentions some of the hyperspectral sensors currently being operated for investigation or commercial purposes. Most of this spectrometers measure radiation in the solar illumination portion of the electromagnetic spectrum ( $0.4\mu\text{m}$  to  $2.5\mu\text{m}$ ). This limits their use to daylight hours. When the range of the hyperspectral sensor is extended into thermal infrared region ( $8\mu\text{m}$  to  $14\mu\text{m}$ ), materials emit more radiation than they reflect from the sun allowing spectrometers to operate all daylong, as it is the case for the DAIS 2115.

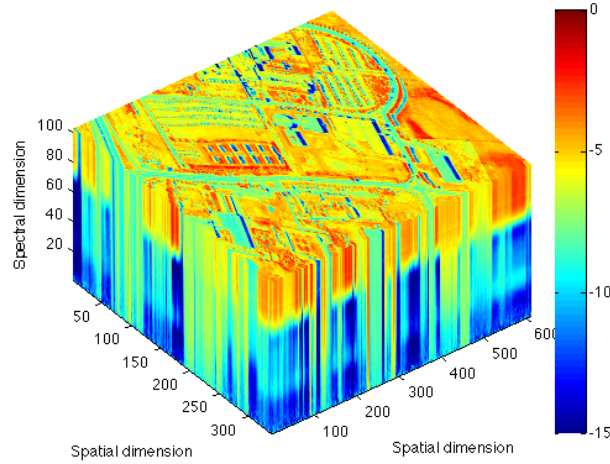


Figure 1.1: Datacube for the Pavia scene.

An equivalent interpretation of HSI is given by the obtention of a stack of images representing the radiance in the corresponding wavelength. Thereupon, the HSI sensors provide three-dimensional data cube with spatial-spatial-spectral components. Fig. 1.1 shows an example of a data cube. Thus, if the values of all pixels in the same spectral band are depicted in spatial coordinates, a grayscale image is obtained which shows the spatial distribution of the scene in the corresponding spectral wavelength (see Fig. 1.2). If all the bands are extracted in the same spatial location and plotted as a function of wavelength, the average of all the materials present in the corresponding ground resolution cell forms the spectrum (see Fig. 1.3). The scene in Fig. 1.1 was acquired by the ROSIS sensor with spectral coverage ranging from  $0.43\mu\text{m}$  to  $0.86\mu\text{m}$  during a flight campaign over Pavia, northern Italy, in 2003. The number of spectral bands is 103 for  $610 \times 340$  pixels. Moreover, the spatial resolution is 1.3 meters.

Sensor	Organization	Country	Number of Bands	Wavelength Range ( $\mu m$ )
AVIRIS	NASA	United States	224	0.4-2.5
AISA	Spectral Imaging Ltd.	Finland	286	0.45-0.9
CASI	Itres Research	Canada	288	0.43-0.87
DAIS 2115	GER Corp.	United States	211	0.4-12.0
HYMAP	Integrated Spectronics Pty Ltd.	Australia	128	0.4-2.45
PROBE-1	Earth Search Sciences Inc.	United States	128	0.4-2.45

Table 1.1: Research and Commercial Imaging Spectrometers

Figure 1.2: Image at a single wavelength (45<sup>th</sup> band).

In the spectral representation, the pixel  $\mathbf{x}$  can be seen as a  $m$ -dimensional vector  $\mathbf{x} = [x_1, \dots, x_m]^T$  in  $\mathbb{R}^m$  [Manolakis et al. \(2003\)](#). Each spectral band is assigned to one dimension and the vector lies inside the positive cone of  $\mathbb{R}^m$ . Note that a change in the level of illumination can modify the norm of the vector but not its orientation which is related to the shape of the spectra. If every material can be characterized by a unique spectrum, this could be used as a spectral mark for discrimination purposes. In the spatial representation, each band corresponds to a matrix  $\mathbf{X}$  in  $\mathbb{R}^{N_1 \times N_2}$ , where  $N_1$  and  $N_2$  are the two spatial dimensions of the image.



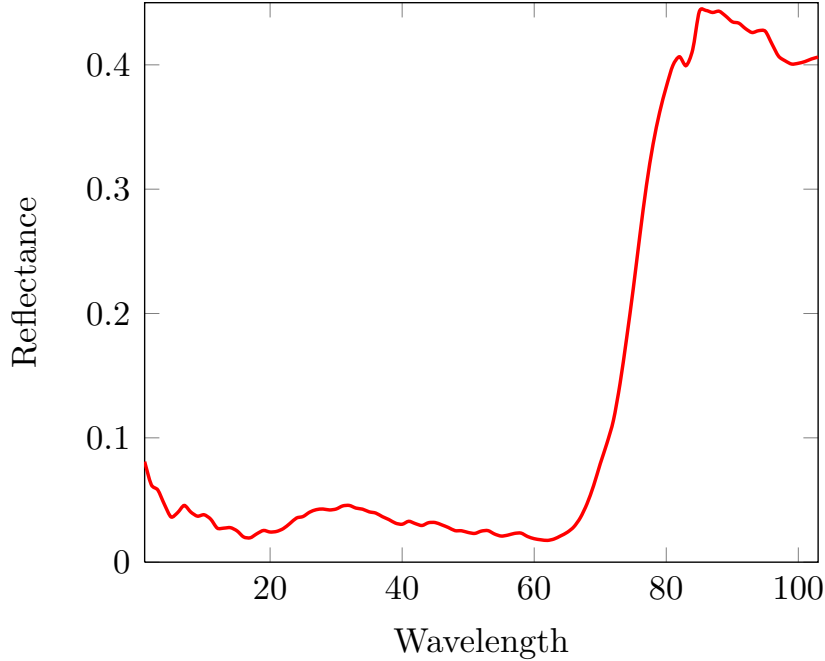


Figure 1.3: Spectra of one of the pixels in the image (pixel (1,1)).

These spectral and spatial representations have been widely turn to good account in several HSI applications: feature extraction, unmixing [Keshava and Mustard \(2002\)](#), [Chen et al. \(2013\)](#), [Bioucas-Dias et al. \(2012\)](#), classification [Harsanyi and Chang \(1994\)](#), segmentation [Valero et al. \(2011\)](#) detection [Nasrabadi \(2014\)](#), and closely related to the high-dimensionality of HSI data, sparse and redundant modeling [Theiler et al. \(2011\)](#). This work is focused on target detection tasks and a review on target detection applications for HSI is provided in next section.

## 1.2 Target Detection and Anomaly Detection

As detailed before, hyperspectral processing involves various applications such as unmixing, classification, detection, dimensionality reduction, ... Among them, hyperspectral detection is an active research topic that has led to many publications e.g. [Manolakis and Shaw \(2002\)](#), [Stein et al. \(2002\)](#), [Chang and Chiang \(2002\)](#), [Kwon and Nasrabadi \(2006\)](#). More precisely, hyperspectral target detection methods are commonly used to detect targets embedded in background and that generally cannot be solved by spatial resolution [Matteoli et al. \(2010a\)](#). There are two different methodologies for target detection purposes in the HSI literature [Manolakis et al. \(2003\)](#): Anomaly Detection [Stein et al. \(2002\)](#), [Chang and Chiang \(2002\)](#) and Target Detection [Manolakis and Shaw \(2002\)](#).

When the spectral signature of the desired target is known, it can be used as steering vector in *Target Detection* techniques [Manolakis et al. \(2003\)](#). However, despite of the application, the aim of target detection techniques is to locate small rare objects

that constitute a very small region compared to the background. The problem of target detection can be stated as a hypothesis test which decides between the presence or the absence of a target. Furthermore, Detection Theory [Kay \(1998\)](#) arises in many different military and civilian applications and has been widely investigated in several signal processing domains such as radar, sonar, communications, see [Gini et al. \(2001\)](#) for the different references. The typical framework proposed in detection theory is based on the Neyman-Pearson approach [Van Trees \(2004\)](#). The design criterion consists in maximizing the probability of detection for a given probability of false alarm. Moreover, the decision strategy is given by a Likelihood Ratio Test (LRT) which is subject to the probability density functions (PDFs) conditioned on the two hypotheses. Nevertheless, the LRT supposes that all the parameters of the distributions are perfectly known and in real applications such information is usually unavailable. Therefore, the suboptimal decision approach generally used is the Generalized LRT (GLRT) [Kay \(1993\)](#) which employs the available independent and identically distributed (IID) signal-free secondary data to estimate the unknown parameters according to the Maximum Likelihood (ML) strategy. Unfortunately, one cannot characterize one material with an unique deterministic spectrum [Manolakis et al. \(2007\)](#). The observed spectra samples from the same material are never identical, because of variations in atmospheric conditions, in the surface material, sensor noise, adjacent materials, etc. Thus, the inherent spectral variability prevents to characterize homogenous regions with only one spectral signature. The resulting detection schemes have to be robust against this spectral variability.

In many practical situations, there is not enough information about the target to detect, thus *Anomaly Detection* methods are widely used [Eismann et al. \(2009\)](#). Anomaly detectors search for pixels in the image with spectral characteristics that differ from the background. It can be interpreted as a particular case of target detection in which no a priori information about the spectra of the targets of interest is available. The type of interesting targets can differ significantly from one application to another, e.g. in forestry applications infected trees are the anomalies of interest, whereas in defense and intelligence applications the anomalies to be detected are usually vehicles. Note that, since anomaly detectors do not use any a priori knowledge, they cannot distinguish between true targets and detections of bright pixels of the background or targets that are not of interest. Anomaly detection arises in many other domains other than hyperspectral imaging such as biomedical imaging [Chen et al. \(1999\)](#), [Tarassenko et al. \(1995\)](#), network and data security [Lazarevic et al. \(2003\)](#) and data mining [Song et al. \(2007\)](#).

## 1.3 Dimensionality Reduction

One of the major challenges in hyperspectral data exploitation is the high-dimensionality of the data cubes provided by the sensors. Because hyperspectral pixels usually consist of hundreds of bands and the data cubes are typically hundreds of megabytes, the amount of data to analyze can be overwhelming. On the other hand, as spectral

bands are narrow and continuous, there is generally a high correlation between contiguous bands. Consequently, the number of bands acquired by the sensor is much higher than the dimensionality of the subspace where the hyperspectral information lies. Many methods have been proposed to reduce the dimensionality. A popular technique consists in choosing the best bands according to a certain criteria, such as the Signal-to-Noise Ratio (SNR), separation between classes, or reconstruction accuracy (see e.g. [Chang et al. \(1999\)](#), [Mausel et al. \(1990\)](#)). In this work, we have used linear projection methods to reduce the dimension [Chen et al. \(2011a\)](#). It involves the projection of the data into a lower dimensional subspace where most of the information is contained. Although other sophisticated methods for dimensionality reduction could be studied, they are beyond the scope of this work and constitute part of perspectives to be further investigated.

Let us consider a  $m$ -dimensional pixel  $\mathbf{x}$ . The vector can be reduced to a  $k$ -dimensional vector (and  $k < m$ ) by projecting it through a  $m \times k$  matrix  $\mathbf{P}$ ,  $\mathbf{y} = \mathbf{P}\mathbf{x}$ . Thus, the choice of  $\mathbf{P}$  leads to different dimensionality reduction schemes. Those explored in this work, due to their simplicity, are:

- **Random Subset of Bands.** Let  $\mathbf{Q}$  be a random permutation matrix of size  $m \times m$  with only one entry of 1 in each row and each column and 0's elsewhere. The projection matrix is obtained as a random subset of  $k$  rows of  $\mathbf{Q}$ ,
- **Random Projection.** The entries of the matrix  $P_{i,j}$  are independent identically distributed samples from a Gaussian distribution  $\mathcal{N}(0, 1/k)$ ,
- **Uniform Downsampling.** Let  $\mathbf{p} = [1\ 0 \dots 0]^T$  be a  $d$ -dimensional row vector where  $d$  is the downsampling rate ( $d \approx m/k$ ). Thus, the projection matrix can be written as:

$$\mathbf{P} = \begin{bmatrix} \mathbf{p} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{p} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{p} \end{bmatrix}, \quad (1.1)$$

where  $\mathbf{0}$  is a  $d$ -dimensional zero vector. If  $m$  is not an integer multiple of  $k$ , the last column will consist of  $m - (m/k)d < d$  entries.

- **Averaging.** Let  $\mathbf{p} = [d\ d \dots d]^{-T}$  a  $d$ -dimensional row vector where  $d$  is the downsampling rate. The projection matrix  $\mathbf{P}$  takes the same form as in Eq. (1.1).

The resulting data set in the lower dimensional space is highly correlated in the spectral dimension for Random Projection and Averaging methods. Thus, Uniform Downsampling and Random Subset of Bands techniques are preferred. However, the choice of the most discriminating bands utterly depends on the application.

Before detailing the classical detection methods, let us set the framework and describe the classical hypotheses in signal processing.

## 1.4 Preliminary Notions

In most signal processing applications, one considers that interference signals are characterized as a random process and has to assume a distribution for the received data. For instance in radar domain, it is common to represent the clutter with a certain distribution and statistically differentiate the target from this distribution. Similarly, in HSI, the reflectance for a given material suffers from an inherent variability. This variability is usually described by using probabilistic models and different materials will lead to different distributions. This principle is used for target detection purposes, classification, unmixing, etc.

In general, the statistical parameters of these distributions are hardly available. Thus, they have to be estimated from the data itself. The mean vector (MV, first order moment) is often supposed to be zero or perfectly known, and in that situation it can be exploited. However, it is not the case for HSI. Indeed, hyperspectral data represent radiance or reflectance and thus, they are positive and the mean vector has to be included for characterization. The covariance matrix (CM, second order moment) and its estimation are an important research topic in multivariate statistics and in signal processing. The aim of this work is to jointly study these two parameters.

In most signal processing literature, signals are usually characterized through complex distributions. Hence, target detection methods are derived considering complex distributed data and classical results on the detector's distribution are found in the complex case. As hyperspectral data are real, the analysis could have been done in the real case. However, the detectors distribution are unknown for zero-mean Gaussian distribution in the real case. For this purpose, Hilbert transform is used to render hyperspectral data complex. The analytic representation of the signal is obtained as:  $\mathbf{x}_a = \mathbf{x} + j\mathcal{H}(\mathbf{x})$ , where  $\mathbf{x}$  is the original real-valued vector and  $\mathcal{H}(\cdot)$  is the Hilbert transform. The main idea is that the negative components of the Fourier transform of a real-valued function are superfluous, due to the Hermitian symmetry of the Fourier transformation. This method discards these negative components without loss of information, obtaining a complex-valued function instead. The conversion from complex back to real is just a matter of discarding the imaginary part and the real component of the analytic representation is still the original data. Therefore, this technique does not change the nature of the original hyperspectral data. The analytical representation will be considered for classical target detection purposes, while anomaly detection methods are studied in the real case as they were originally derived for HSI.

Before going further, let us define the terms of which knowledge are essential to read this thesis and will be used throughout this document.

### Definition 1.4.1. Mean vector

*The mean vector (or expectation) of a complex random vector  $\mathbf{x} = \mathbf{u} + j\mathbf{v}$  is defined as:*

$$\mathbb{E}[\mathbf{x}] = \mathbb{E}[\mathbf{u}] + j\mathbb{E}[\mathbf{v}]. \quad (1.2)$$

**Definition 1.4.2.** *Covariance Matrix*

The covariance matrix of a complex random vector  $\mathbf{x} = \mathbf{u} + j\mathbf{v}$  of dimension  $m$  is defined as:

$$\begin{aligned}\mathbf{M} = \text{cov}(\mathbf{x}) &= \mathbb{E}[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])^H] \\ &= \text{cov}(\mathbf{u}) + \text{cov}(\mathbf{v}) + j(\text{cov}(\mathbf{v}, \mathbf{u}) - \text{cov}(\mathbf{u}, \mathbf{v})),\end{aligned}\quad (1.3)$$

and  $\mathbf{M}$  belongs to the class of positive semidefinite Hermitian  $m \times m$  matrices.

It has been often assumed that signals, interferences, noise and background are modeled as Gaussian stochastic processes. Indeed, this assumption makes sense in many applications [Kay \(1998\)](#). Let us now recall the definition of a Gaussian vector.

**Definition 1.4.3.** *Complex Gaussian Distribution*

A  $m$ -dimensional vector  $\mathbf{x} = \mathbf{u} + j\mathbf{v}$  has a complex Gaussian distribution with mean  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$  if  $\mathbf{x} = (\mathbf{u}^T, \mathbf{v}^T)^T \in \mathbb{R}^{2m}$  has a normal distribution [van den Bos \(1995\)](#). If  $\text{rank}(\boldsymbol{\Sigma}) = m$ , the probability density function exists and is of the form

$$f_{\mathbf{x}}(\mathbf{x}) = \pi^{-m} |\boldsymbol{\Sigma}|^{-1} \exp\{-(\mathbf{x} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\}. \quad (1.4)$$

In the following, it will be referred to as Gaussian distribution or normal distribution indistinctively and denoted  $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . We assume that vectors are circular of second order and that the pseudo-covariance matrix is zero ( $\mathbb{E}[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])^T] = \mathbf{0}$ ), thus it will be omitted. The complex Gaussian distribution is called circular, without going into further details. However, this assumption is not always valid and we refer to [Schreier and Scharf \(2010\)](#) for non-circular signal processing.

The Gaussian distribution benefits from the fact that the optimal estimators in the sense of Maximum Likelihood (ML) for the statistical parameters are easily obtained. Along with their well-known properties and their simplicity of analysis, the Sample Mean Vector (SMV) and the Sample Covariance Matrix (SCM) are the most extended estimators and they are easily implemented from a given set of samples. Let  $\mathbf{x}_1, \dots, \mathbf{x}_N$  be an IID  $N$ -sample, where  $\mathbf{x}_i \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . Thus, the SMV and the SCM can be written as:

$$\hat{\boldsymbol{\mu}}_{SMV} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i, \quad (1.5)$$

$$\hat{\boldsymbol{\Sigma}}_{SCM} = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^H. \quad (1.6)$$

Further, we shall denote the Centered SCM (CSCM) as:

$$\hat{\boldsymbol{\Sigma}}_{CSCM} = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^H. \quad (1.7)$$

Some key properties about the distribution of the estimators are recapped in the following.

**Definition 1.4.4.** *Complex Wishart distribution*

Let  $\mathbf{x}_1, \dots, \mathbf{x}_N$  be an IID  $N$ -sample, where  $\mathbf{x}_i \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . Let us define  $\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}}_{SMV}$  and  $\hat{\mathbf{W}} = N \hat{\boldsymbol{\Sigma}}_{SCM}$  referred to as a Wishart matrix. Thus one has (see [Gupta and Nagar \(2000\)](#) for the real case):

- $\hat{\boldsymbol{\mu}}$  and  $\hat{\mathbf{W}}$  are independently distributed;
- $\hat{\boldsymbol{\mu}} \sim \mathcal{CN}(\boldsymbol{\mu}, \frac{1}{N} \boldsymbol{\Sigma})$ ;
- $\hat{\mathbf{W}} \sim \mathcal{CW}(N - 1, \boldsymbol{\Sigma})$  is Wishart distributed with  $N - 1$  degrees of freedom.

The asymptotic distribution of the Wishart matrix is (e.g. [Bilodeau and Brenner \(1999\)](#)):

$$\sqrt{N} \text{vec}(\hat{\mathbf{W}} - \boldsymbol{\Sigma}) \xrightarrow{d} \mathcal{CN}(\mathbf{0}_{m^2, 1}, (\boldsymbol{\Sigma}^T \otimes \boldsymbol{\Sigma}), (\boldsymbol{\Sigma}^T \otimes \boldsymbol{\Sigma}) \mathbf{K}). \quad (1.8)$$

where  $\mathbf{K}$  is the commutation matrix. Therefore, the performances obtained with the SMV-SCM estimators are perfectly known. These estimators are unbiased, consistent and asymptotically Gaussian.

After providing the framework on Gaussian distributions, we discuss now the general background for adaptive target detection. First, we recall classical zero mean Gaussian detection schemes. Then, we will extend in Chapter 2 these tests for the non-zero mean case studied in this work.

## 1.5 Classical Target Detection Techniques

In signal processing, the problem of target detection is generally stated as a binary hypothesis test which decides the presence or absence of a target in the cell under test. Therefore, for a fixed cell of study, the received signal is a vector  $\mathbf{x}$  of dimension  $m$  referred to as the observation vector. The known complex signal characterizing the intended target is denoted by  $\mathbf{s}$ , and it is corrupted by an additive noise  $\mathbf{b}$  referred as clutter or background. The detection problem can be written as:

$$\begin{cases} \mathcal{H}_0 : \mathbf{x} = \mathbf{b}, & \mathbf{x}_i = \mathbf{b}_i, i = 1, \dots, N \\ \mathcal{H}_1 : \mathbf{x} = \mathbf{s} + \mathbf{b}, & \mathbf{x}_i = \mathbf{b}_i, i = 1, \dots, N, \end{cases} \quad (1.9)$$

where the  $\mathbf{x}_i$  are the so-called secondary data (signal-free) used to estimate the noise parameters, and the signal  $\mathbf{s}$  can be written in the form  $\alpha \mathbf{p}$ , where  $\alpha$  is an unknown complex scalar amplitude, and  $\mathbf{p}$  is the steering vector describing the signal which is sought.

Under the signal absent (or null) hypothesis  $\mathcal{H}_0$ , the signal  $\mathbf{x}$  only contains the background. This is constituted, among others, by various reflections from the ground and environmental components, as well as thermal noise. Under the signal present (or alternative) hypothesis  $\mathcal{H}_1$ , in addition to this background, the signal  $\mathbf{x}$  contains the target  $\mathbf{s}$  to be detected. The goal of detection is to determine which of these two hypotheses is more likely, while minimizing the following probabilities:

- the non-detection probability PND, which is the probability of not detecting the target (hypothesis  $\mathcal{H}_0$ ) when in fact the signal is present,
- the false-alarm probability PFA, which is the probability of deciding the target is present (hypothesis  $\mathcal{H}_1$ ) when in fact it is not.

One can also define the probability of detection,  $PD = 1 - PND$ .

In practice, it is impossible to simultaneously minimize these two errors. A good compromise is given by the Neyman-Pearson criterion which aims to maximize the probability of detection PD (or equivalently minimize PND) for a fixed PFA. The optimal test according to this criterion (see e.g. [Kay \(1998\)](#)) is the Likelihood Ratio (LR) given by:

$$\Lambda(\mathbf{x}) = \frac{p(\mathbf{x}|\mathcal{H}_1)}{p(\mathbf{x}|\mathcal{H}_0)} \underset{\mathcal{H}_0}{\gtrless} \eta. \quad (1.10)$$

The detection threshold  $\eta$  is determined according to a fixed value for the false-alarm probability  $PFA = k$  and computed solving one of the followings equations:

$$PFA = \mathbb{P}(\Lambda(\mathbf{x}; \mathcal{H}_0) > \eta) = k, \quad (1.11)$$

$$PFA = \int_{\mathcal{D}_0} p(\mathbf{x}|\mathcal{H}_0) d\mathbf{x}, \quad (1.12)$$

where  $\mathcal{D}_0$  is the set of  $\mathbf{x}$  contained in the decision domain of the hypothesis  $\mathcal{H}_0$ . The aim of this chapter is to provide analytic expressions for these equations. Otherwise, one has to numerically solve Eq. (1.11) and Eq. (1.12) or obtain the threshold through Monte-Carlo simulations which is a less accurate method and/or more time consuming.

**Remark 1.5.1.** *The relationship between the PFA and the detection threshold  $\eta$  is nothing else than the complementary cumulative density function of the detector  $\Lambda(\mathbf{x}; \mathcal{H}_0)$ . Thus, it completely characterizes the statistical distribution of this detector under hypothesis  $\mathcal{H}_0$ .*

Once the threshold is settled, the PD can be calculated according to:

$$PD = \mathbb{P}(\Lambda(\mathbf{x}; \mathcal{H}_1) > \eta), \quad (1.13)$$

$$PD = \int_{\mathcal{D}_1} p(\mathbf{x}|\mathcal{H}_1) d\mathbf{x}, \quad (1.14)$$

where  $\mathcal{D}_1$  is the set of  $\mathbf{x}$  contained in the decision domain of the hypothesis  $\mathcal{H}_1$ . In this case, it is more unusual to obtain a closed-form expression for the PD. This problem is beyond the scope of this work. Since the background statistics, i.e. the MV and the CM, are assumed to be unknown, they have to be estimated from  $\mathbf{x}_1, \dots, \mathbf{x}_N \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  a sequence of  $N$  IID signal-free secondary data. Then, the adaptive detector is obtained by replacing in Eq. (1.10) the unknown parameters by their estimates leading to the Generalized Likelihood Ratio Test (GLRT). In practice,

an estimate may be obtained from the pixels surrounding pixel cell under test, which play the role of the  $N$  IID signal-free secondary data and replacing the parameters by their estimates in the GLRT detector yields the two-step GLRT approach. The sample size  $N$  has to be chosen large enough to ensure the invertibility of the covariance matrix and small enough to justify both spectral homogeneity (stationarity) and spatial homogeneity. The use of a sliding mask provides a more realistic scenario than when estimating the parameters using all the pixels in the image. For a given pixel, a window centered on this pixel compose the secondary data used to estimate the statistical parameters. In addition, a small window with the size of the expected targets can be used as guard window. An example of such a window is displayed in Fig. 1.4. However, it requires the information on the size and shape of the targets and this information is hardly available. Let us now recall the detectors under interest in this work.

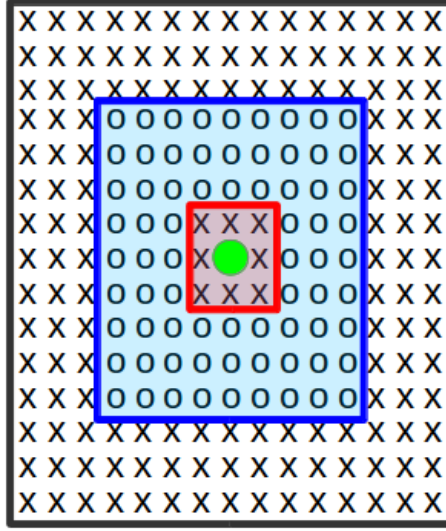


Figure 1.4: Sliding window mask over the observation vector (green dot) with a guard window (red) and useful secondary data (blue).

### 1.5.1 Adaptive Matched Filter

The Matched Filter (MF) detector is the optimal linear filter for maximizing the SNR in the presence of additive Gaussian noise with known parameters [Kay \(1998\)](#). Hence, the signal model can be written as:

$$\begin{cases} \mathcal{H}_0 : & \mathbf{x} = \mathbf{b} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ \mathcal{H}_1 : & \mathbf{x} = \alpha \mathbf{p} + \mathbf{b} \sim \mathcal{CN}(\alpha \mathbf{p} + \boldsymbol{\mu}, \boldsymbol{\Sigma}). \end{cases} \quad (1.15)$$

The LR according to the signal model is given by:

$$\Lambda(\mathbf{x}, \alpha) = \frac{\exp \left[ -(\mathbf{x} - (\alpha \mathbf{p} + \boldsymbol{\mu}))^H \boldsymbol{\Sigma}^{-1} (\mathbf{x} - (\alpha \mathbf{p} + \boldsymbol{\mu})) \right]}{\exp \left[ -(\mathbf{x} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right]} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta. \quad (1.16)$$



Taking logarithms and ignoring the constant terms, one obtains:

$$(\mathbf{x} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) - (\mathbf{x} - (\alpha \mathbf{p} + \boldsymbol{\mu}))^H \boldsymbol{\Sigma}^{-1} (\mathbf{x} - (\alpha \mathbf{p} + \boldsymbol{\mu})) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda. \quad (1.17)$$

and  $\lambda = \log(\eta)$ . Since the complex amplitude is unknown, it has to be estimated from the observation vector  $\mathbf{x}$  and the background parameters according to:

$$\hat{\alpha} = \frac{\Re\{\mathbf{p}^H \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\}}{\mathbf{p}^H \boldsymbol{\Sigma}^{-1} \mathbf{p}}. \quad (1.18)$$

Replacing this value in Eq. (1.16) and after some manipulations, the resulting MF detection scheme is:

$$\Lambda_{MF} = \frac{|\mathbf{p}^H \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})|^2}{(\mathbf{p}^H \boldsymbol{\Sigma}^{-1} \mathbf{p})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda. \quad (1.19)$$

**Remark 1.5.2.** Note that it differs from the classical MF by the term  $\boldsymbol{\mu}$ , the background mean, but without any consequence since  $\mathbf{x} - \boldsymbol{\mu} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$ .

The “PFA-threshold” curve corresponds to the statistical behavior in general of a detector under null hypothesis. Moreover, the “PFA-threshold” relationship is given by [Kay \(1998\)](#):

$$PFA_{MF} = \exp(-\lambda). \quad (1.20)$$

The widely used Adaptive Matched Filter (AMF), denoted  $\Lambda_{AMF\hat{\boldsymbol{\Sigma}}}^{(N)}$  to underline the dependency with  $N$ , is usually built replacing the covariance matrix  $\boldsymbol{\Sigma}$  by its estimate  $\hat{\boldsymbol{\Sigma}}$  obtained from the  $N$  secondary data (two-step GLRT approach). The mean vector is generally supposed to be known. Thus, the adaptive version becomes:

$$\Lambda_{AMF\hat{\boldsymbol{\Sigma}}}^{(N)} = \frac{|\mathbf{p}^H \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{x} - \boldsymbol{\mu})|^2}{(\mathbf{p}^H \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{p})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda. \quad (1.21)$$

Then, the theoretical “PFA-threshold” relationship is given by [Robey et al. \(1992\)](#) for an estimate following a Wishart distribution obtained with  $\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Sigma}}_{SCM}$ :

$$PFA_{AMF\hat{\boldsymbol{\Sigma}}} = {}_2F_1\left(N - m + 1, N - m + 2; N + 1; -\frac{\lambda}{N}\right), \quad (1.22)$$

where  ${}_2F_1(\cdot)$  is the hypergeometric function [Abramowitz et al. \(1964\)](#) defined as,

$${}_2F_1(a, b; c; z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 \frac{t^{b-1}(1-t)^{c-b-1}}{(1-tz)^a} dt. \quad (1.23)$$

This detector holds the CFAR properties in the sense that its false alarm expression only depends on the dimension of the vector  $m$  and the number of secondary data used for the estimation  $N$ . Note that it is also independent of the noise covariance matrix  $\boldsymbol{\Sigma}$ , therefore the detector is said to be CFAR-matrix. However, its performance strongly relies on the good fit of the Gaussian model and the false alarm rate is highly increased when normal assumption is not verified.

### 1.5.2 Adaptive Kelly detector

The Kelly detector was derived in Kelly (1986). It is based on the Generalized Likelihood Ratio Test (GLRT) assuming Gaussian distribution and the same signal model than the AMF in Eq. (1.15). In this case, only the covariance matrix  $\mathbf{\Sigma}$  is unknown, the mean vector is assumed to be known. Therefore,  $N + 1$   $m$ -dimensional vectors are observed under each hypothesis. Under the hypothesis  $\mathcal{H}_0$ , the  $N + 1$  vectors are all Gaussian noise, while under  $\mathcal{H}_1$ , just  $N$  vectors contain only the background. Thus, the joint probability density function (PDF) of the  $N$  secondary data and the observation vector  $\mathbf{x}$  under the two hypotheses  $\mathcal{H}_i$  can be written as the product of  $N + 1$  multivariate Gaussian densities from each vector, separating the observation vector from the  $N$  secondary data:

$$f_i(\mathbf{x}) = \left( \frac{1}{\pi^m |\mathbf{\Sigma}|} \exp[-\text{Tr}(\mathbf{\Sigma}^{-1} \mathbf{T}_i)] \right)^{N+1}, \quad (1.24)$$

where  $\text{Tr}$  is the trace operator and  $\mathbf{T}_i$  is the composite sample covariance matrix constructed from both the secondary data and observation vector:

$$\mathbf{T}_0 = \frac{1}{N+1} \left( (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^H + \hat{\mathbf{W}} \right), \quad (1.25)$$

$$\mathbf{T}_1 = \frac{1}{N+1} \left( (\mathbf{x} - (\alpha \mathbf{p} + \boldsymbol{\mu}))(\mathbf{x} - (\alpha \mathbf{p} + \boldsymbol{\mu}))^H + \hat{\mathbf{W}} \right), \quad (1.26)$$

and  $\hat{\mathbf{W}} = N \hat{\mathbf{\Sigma}}_{CSCM}$ . The first step is to maximize with respect to (w.r.t) the unknown covariance matrix  $\mathbf{\Sigma}$ . Thus, the matrix maximizing the PDF  $f_i$  is simply  $\mathbf{T}_i$ . When this estimator is replaced in the PDF, one obtains:

$$\max_{\mathbf{\Sigma}} f_i = \left( \frac{1}{(\pi e)^m |\mathbf{T}_i|} \right)^{N+1}. \quad (1.27)$$

It remains to maximize this expression over the complex unknown signal amplitude  $\alpha$ . And the GLRT neglecting the exponent  $N + 1$  is given by:

$$\Lambda(\mathbf{x}, \alpha) = \frac{|\mathbf{T}_0|}{|\mathbf{T}_1|} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \eta. \quad (1.28)$$

Then, by maximizing the PDF under both hypotheses and by maximizing the LR w.r.t the complex signal amplitude, and after some manipulations, the resulting adaptive Kelly detector scheme takes the following form:

$$\Lambda_{Kelly}^{(N)} = \frac{|\mathbf{p}^H \hat{\mathbf{\Sigma}}_{CSCM}^{-1} (\mathbf{x} - \boldsymbol{\mu})|^2}{\left( \mathbf{p}^H \hat{\mathbf{\Sigma}}_{CSCM}^{-1} \mathbf{p} \right) \left( N + (\mathbf{x} - \boldsymbol{\mu})^H \hat{\mathbf{\Sigma}}_{CSCM}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda, \quad (1.29)$$

where  $\lambda = 1 - \eta^{-\frac{1}{N+1}}$ . As shown in Kelly (1986), the PFA for the Kelly test is given by:

$$PFA_{Kelly} = (1 - \lambda)^{N-m+1}. \quad (1.30)$$

The Kelly detector is a CFAR test, in which the PFA is independent of the true covariance matrix. However, it has no known optimality property in the sense of maximizing the probability of detection for a given probability of false alarm as explained in [Kelly \(1986\)](#). The AMF and the Kelly detector are based on the same assumptions about the nature of the observations. It is therefore interesting to compare their detection performance for a given PFA. The main difference between Eq. (1.21) and Eq. (1.29) is the presence of the additional term  $\frac{1}{N}(\mathbf{x} - \boldsymbol{\mu})^H \hat{\boldsymbol{\Sigma}}_{CSCM}^{-1}(\mathbf{x} - \boldsymbol{\mu})$ . This difference can be seen as a measure of the error in the estimate of the covariance matrix from the  $N$  reference data. In particular, when  $N$  is large ( $N \rightarrow \infty$ ), this term tends to 0,  $\hat{\boldsymbol{\Sigma}}$  tends to its true value,  $\boldsymbol{\Sigma}$  and the GLRT tends to MF. Thus, for large values of  $N$ , the performances are substantially the same.

### 1.5.3 Adaptive Normalized Matched Filter

The Normalized Matched Filter (NMF) is obtained when considering that the covariance matrix is different under the two hypotheses. That is to say that the background has the same covariance structure but different variance.

$$\begin{cases} \mathcal{H}_0 : & \mathbf{x} = \mathbf{b} \sim \mathcal{CN}(\boldsymbol{\mu}, \sigma_0^2 \boldsymbol{\Sigma}) \\ \mathcal{H}_1 : & \mathbf{x} = \alpha \mathbf{p} + \mathbf{b} \sim \mathcal{CN}(\alpha \mathbf{p} + \boldsymbol{\mu}, \sigma_1^2 \boldsymbol{\Sigma}). \end{cases} \quad (1.31)$$

Thus, the ML estimates of  $\sigma_j^2$  are easily derived from  $\hat{\sigma}_j^2 = \arg \max_{\sigma} \{f(\mathbf{x}|\sigma, \mathcal{H}_j)\}$ , ( $j = 0, 1$ ) and assuming normal distribution, one has:

$$\hat{\sigma}_0^2 = \frac{1}{2m}(\mathbf{x} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}), \quad (1.32)$$

$$\hat{\sigma}_1^2 = \frac{1}{2m}(\mathbf{x} - (\alpha \mathbf{p} + \boldsymbol{\mu}))^H \boldsymbol{\Sigma}^{-1}(\mathbf{x} - (\alpha \mathbf{p} + \boldsymbol{\mu})), \quad (1.33)$$

After replacing the complex amplitude  $\alpha$  by its estimate Eq. (1.18) when building the LR and after some manipulations, one obtains [Scharf and Friedlander \(1994\)](#):

$$\Lambda_{NMF} = \frac{|\mathbf{p}^H \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})|^2}{(\mathbf{p}^H \boldsymbol{\Sigma}^{-1} \mathbf{p}) ((\mathbf{x} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}))} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda, \quad (1.34)$$

where  $\lambda = 1 - \eta^{-\frac{1}{m}}$  and for which one has [Scharf and Friedlander \(1994\)](#):

$$PFA_{NMF} = (1 - \lambda)^{m-1}. \quad (1.35)$$

The ANMF is generally obtained when the unknown noise covariance matrix is replaced by an estimate according to the two-step GLRT strategy [Kraut et al. \(2001\)](#):

$$\Lambda_{ANMF\hat{\boldsymbol{\Sigma}}}^{(N)} = \frac{|\mathbf{p}^H \hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{x} - \boldsymbol{\mu})|^2}{(\mathbf{p}^H \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{p}) ((\mathbf{x} - \boldsymbol{\mu})^H \hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{x} - \boldsymbol{\mu}))} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda. \quad (1.36)$$

And the PFA for an estimate following a Wishart distribution obtained with  $\hat{\Sigma} = \hat{\Sigma}_{SCM}$  can be written according to [Kraut et al. \(2001\)](#):

$$PFA_{ANMF\hat{\Sigma}} = (1 - \lambda)^{a-1} {}_2F_1(a, a - 1; b - 1; \lambda), \quad (1.37)$$

where  $a = N - m + 2$  and  $b = N + 2$ .

This detection test has been widely studied in signal processing and it has been derived independently by different authors following different approaches [Korado \(1968\)](#), [Gini \(1997\)](#), [Kraut and Scharf \(1999\)](#). It also benefits from great popularity in HSI target detection [Manolakis et al. \(2009, 2013\)](#). This detection test holds important CFAR properties as its distribution is independent of the noise covariance matrix.

## 1.6 Summary

This chapter sets out the main characteristics of the hyperspectral imaging. The acquisition process and the different representations for hyperspectral data are detailed. Hyperspectral sensors are a new class of optical sensors that measures the spectrum from each point in an image. There has been an increasing interest in the last decades in the use of HSI for a wide range of applications, leading to the proliferation of new signal processing methods for hyperspectral data exploitation such as unmixing, classification, feature extraction, etc. Notably, this thesis is focused on the detection of small objects or targets in hyperspectral images. The two principal target detection techniques in HSI, target detection and anomaly detection, are investigated in the following chapters.

We have also introduced the basic blocks for classical target detection in Gaussian environment. These techniques are found in many signal processing applications, particularly in radar domain. We propose in [Chapter 2](#) to extend these classical methods to the case where the background mean vector is unknown and has to be included in the detection process. Moreover, when Gaussian assumption is not fulfilled, [Chapter 3](#) describes detection schemes in non-Gaussian framework. Finally, this work intends to provide a unified framework for target detection in HSI .



# 2

## Adaptive Target Detection in Gaussian Background

### 2.1 Introduction

Interestingly, target detection methods have been extensively developed and analyzed in the signal processing and radar processing [Gini et al. \(2001\)](#), [Kelly \(1986\)](#), [Kraut et al. \(2001\)](#), [Robey et al. \(1992\)](#). In all these works as well as in several signal processing applications, signals are assumed to be Gaussian with zero mean or with a known MV that can be removed. In such a context, previous chapter has recapped several well-known algorithms in Statistical Detection Theory [Kay \(1998\)](#), for instance the MF and its adaptive versions, the Kelly detector [Kelly \(1986\)](#) or the ANMF [Kraut and Scharf \(1999\)](#). However, when the MV of the noise background is unknown, these techniques are no longer adapted and improved methods have to be derived by taking into account the mean vector estimation. For this purpose, the aim of this chapter is to generalize them to the case when the mean vector is unknown and has to be estimated.

More precisely, this chapter deals with the classical AMF, the Kelly detection test and the ANMF. Further, last section is devoted to anomaly detection methods, it reviews classical anomaly detection schemes such as the widely spread Reed-Xiaoli Detector and some of its variations. All these detectors have been derived under Gaussian assumptions and benefit from great popularity in HSI target detection literature, see e.g. [Manolakis et al. \(2009, 2013\)](#). To evaluate the detector performance, the classical process, according to the Neyman-Pearson criterion is first to regulate

the false-alarm, by setting a detection threshold for a given PFA. Since the PFA is the complementary CDF of the detection test, this process is equivalent to the derivation of the detection test distribution. Then, the probability of detection is evaluated for different SNRs. Therefore, keeping the false-alarm rate constant (CFAR) is essential to set a proper detection threshold [Gini and Greco \(2002\)](#), [Conte et al. \(2002\)](#). The aim is to build a CFAR detector which provides detection thresholds that are relatively immune to noise and background variation, and allow target detection with a constant false-alarm rate. The theoretical analysis of CFAR methods for adaptive detectors is a challenging problem since in adaptive schemes, the statistical distribution of the detectors is not always available in a closed-form expression.

## 2.2 Adaptive non-zero mean Gaussian Detection

In this section, let us now assume that the mean vector  $\mu$  is unknown as it is the case for instance in HSI and let us derive the new corresponding detection schemes. Then, using standard calculus on Wishart distributions, recapped in [Section 1.4](#), the distributions of each detection test are provided.

### 2.2.1 Adaptive Matched Filter Detector

When both the covariance matrix and the mean vector are unknown, they are replaced by their estimates from the secondary data (two-step GLRT) in [Eq. \(1.19\)](#) leading to the AMF detector of the following form:

$$\Lambda_{AMF\hat{\Sigma},\hat{\mu}}^{(N)} = \frac{|\mathbf{p}^H \hat{\Sigma}^{-1} (\mathbf{x} - \hat{\mu})|^2}{(\mathbf{p}^H \hat{\Sigma}^{-1} \mathbf{p})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda, \quad (2.1)$$

where the notation  $\Lambda_{AMF\hat{\Sigma},\hat{\mu}}^{(N)}$  is used to stress now the dependency on the estimated mean vector  $\hat{\mu}$ . The distribution of this detection test is given in the next Proposition, through its PFA.

**Proposition 2.2.1.** *Under Gaussian assumptions, the theoretical relationship between the PFA and the threshold is given by*

$$PFA_{AMF\hat{\Sigma},\hat{\mu}} = {}_2F_1 \left( N - m, N - m + 1; N; -\frac{\lambda'}{N - 1} \right), \quad (2.2)$$

where  $\lambda' = \frac{(N-1)}{(N+1)}\lambda$ ,  $\hat{\Sigma} = \hat{\Sigma}_{SCM}$  and  $\hat{\mu} = \hat{\mu}_{SMV}$ , recapped in [Section 1.4](#).

Before turning into the proof, let us comment on this result.

- Interestingly, this detector also holds the CFAR property in the sense that its false-alarm expression depends only on the dimension  $m$  and on the number of secondary data  $N$ , but not on the noise parameters  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$ . Note that the only effect of estimating the mean is the loss of one degree of freedom and the modification of the threshold compared to Eq. (1.22). Obviously, the impact of these modifications decrease as the number of secondary data  $N$  used to estimate the unknown parameters increases.
- Moreover, the result has been obtained when using the MLEs of the unknown parameters but the proof can be easily extended to other covariance matrix estimators such as  $\hat{\boldsymbol{\Sigma}} = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^H$  which is the unbiased covariance matrix estimate or  $\hat{\boldsymbol{\Sigma}} = \frac{1}{N+1} \sum_{i=1}^N (\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^H$ .

*Proof.* For simplicity matters, the following notations are used:  $\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Sigma}}_{SCM}$  and  $\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}}_{SMV}$ .

Since the derivation of the PFA is done under hypothesis  $\mathcal{H}_0$ , let us set  $\forall i = 1, \dots, N, \mathbf{x}_i \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  and  $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , where all these vectors are independent. Now, let us denote

$$\hat{\mathbf{W}}_{N-1} = \sum_{i=1}^N (\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^H \sim \mathcal{CW}(N-1, \boldsymbol{\Sigma}), \quad (2.3)$$

Since  $\hat{\boldsymbol{\mu}} \sim \mathcal{CN}(\boldsymbol{\mu}, \frac{1}{N}\boldsymbol{\Sigma})$  and  $\hat{\boldsymbol{\mu}}$  is independent of  $\mathbf{x}$  by construction, one has  $\mathbf{x} - \hat{\boldsymbol{\mu}} \sim \mathcal{CN}(\mathbf{0}, \frac{N+1}{N}\boldsymbol{\Sigma})$ . This can be equivalently rewritten as

$$\sqrt{N/(N+1)}(\mathbf{x} - \hat{\boldsymbol{\mu}}) \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}). \quad (2.4)$$

Now, let us set  $\mathbf{y} = \sqrt{\frac{N}{N+1}}(\mathbf{x} - \hat{\boldsymbol{\mu}})$  with  $\mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$ .

When computing the SCM, one has

$$\hat{\boldsymbol{\Sigma}}_{SCM} = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^H = \frac{1}{N} \hat{\mathbf{W}}_{N-1}. \quad (2.5)$$

As we jointly estimate the mean and the covariance matrix, a degree of freedom is lost, compared with the only covariance matrix estimation problem.

Let us now consider the classical AMF test (i.e.  $\boldsymbol{\mu}$  known) built from  $N-1$  secondary data, rewritten in terms of  $\hat{\mathbf{W}}_{N-1}$ :

$$\Lambda_{AMF}^{(N-1)} = (N-1) \frac{|\mathbf{p}^H \hat{\mathbf{W}}_{N-1}^{-1} \mathbf{y}|^2}{(\mathbf{p}^H \hat{\mathbf{W}}_{N-1}^{-1} \mathbf{p})}, \quad (2.6)$$



where  $\mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \Sigma)$  and whose “PFA-threshold” relationship is given by Eq. (1.22) where  $N$  is replaced by  $N - 1$ .

Now, for the joint estimation problem, the AMF can be rewritten as:

$$\begin{aligned}\Lambda_{AMF\hat{\Sigma},\hat{\mu}}^{(N)} &= N \frac{|\mathbf{p}^H \hat{\mathbf{W}}_{N-1}^{-1} (\mathbf{x} - \hat{\mu})|^2}{(\mathbf{p}^H \hat{\mathbf{W}}_{N-1}^{-1} \mathbf{p})} \\ &= N \frac{N+1}{N} \frac{|\mathbf{p}^H \hat{\mathbf{W}}_{N-1}^{-1} \mathbf{y}|^2}{(\mathbf{p}^H \hat{\mathbf{W}}_{N-1}^{-1} \mathbf{p})} \\ &= \frac{(N+1)}{(N-1)} \Lambda_{AMF\hat{\Sigma}}^{(N-1)},\end{aligned}\tag{2.7}$$

where  $(\mathbf{x} - \hat{\mu})$  has been replaced by  $\sqrt{N+1/N} \mathbf{y}$  with  $\mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \Sigma)$ , as previously.

Hence, one can determine the false-alarm relationship:

$$\begin{aligned}PFA_{AMF\hat{\Sigma},\hat{\mu}} &= \mathbb{P} \left( \Lambda_{AMF\hat{\Sigma},\hat{\mu}}^{(N)} > \lambda; \mathcal{H}_0 \right) \\ &= \mathbb{P} \left( \frac{(N+1)}{(N-1)} \Lambda_{AMF\hat{\Sigma}}^{(N-1)} > \lambda; \mathcal{H}_0 \right) \\ &= \mathbb{P}(\Lambda_{AMF\hat{\Sigma}}^{(N-1)} > \lambda'; \mathcal{H}_0),\end{aligned}\tag{2.8}$$

where  $\lambda' = \frac{(N-1)}{(N+1)}\lambda$ , which leads to the conclusion.  $\square$

## 2.2.2 Kelly Detector

The Kelly detector for both the unknown mean vector and the unknown covariance matrix has now to be derived since it is not the previous Kelly in which an estimate of the mean vector is plugged. Following the same lines as in Kelly (1986), we now assume that both the mean vector and the covariance matrix are unknown. The likelihood functions under  $\mathcal{H}_0$  and  $\mathcal{H}_1$  are given in Eq. (1.24). Under  $\mathcal{H}_0$  and  $\mathcal{H}_1$ , the maxima are achieved at:

$$\max_{\Sigma, \mu} f_i = \left( \frac{1}{(\pi e)^m |\mathbf{T}_i|} \right)^{N+1}, \text{ for } i = 0, 1,\tag{2.9}$$

where now

$$(N+1)\mathbf{T}_0 = (\mathbf{x} - \hat{\mu}_0)(\mathbf{x} - \hat{\mu}_0)^H + \sum_{i=1}^N (\mathbf{x}_i - \hat{\mu}_0)(\mathbf{x}_i - \hat{\mu}_0)^H,\tag{2.10}$$

$$(N+1)\mathbf{T}_1 = (\mathbf{x} - \alpha\mathbf{p} - \hat{\mu}_1)(\mathbf{x} - \alpha\mathbf{p} - \hat{\mu}_1)^H + \sum_{i=1}^N (\mathbf{x}_i - \hat{\mu}_1)(\mathbf{x}_i - \hat{\mu}_1)^H,\tag{2.11}$$

and

$$\hat{\boldsymbol{\mu}}_0 = \frac{1}{N+1} \left( \mathbf{x} + \sum_{i=1}^N \mathbf{x}_i \right), \quad (2.12)$$

$$\hat{\boldsymbol{\mu}}_1 = \frac{1}{N+1} \left( \mathbf{x} - \alpha \mathbf{p} + \sum_{i=1}^N \mathbf{x}_i \right). \quad (2.13)$$

And neglecting the exponent  $N+1$ , one obtains the following LR:

$$\Lambda(\mathbf{x}, \alpha) = \frac{|\mathbf{T}_0|}{|\mathbf{T}_1|} \stackrel{\mathcal{H}_1}{\geq} \eta. \quad (2.14)$$

Then, as this LR still depends on the unknown amplitude  $\alpha$  of the signal, it has to be maximized w.r.t  $\alpha$ , which is equivalent to minimize  $\mathbf{T}_1$  w.r.t  $\alpha$ . A way to do this is to introduce the following sample covariance matrix:

$$\hat{\mathbf{S}}_0 = \sum_{i=1}^N (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_0)(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_0)^H. \quad (2.15)$$

Then,  $(N+1)|\mathbf{T}_0|$  can be written as:

$$(N+1)|\mathbf{T}_0| = |\hat{\mathbf{S}}_0| \left( 1 + (\mathbf{x} - \hat{\boldsymbol{\mu}}_0)^H \hat{\mathbf{S}}_0^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_0) \right). \quad (2.16)$$

In the same way, and after some manipulations,  $(N+1)|\mathbf{T}_1|$  becomes:

$$\begin{aligned} (N+1)|\mathbf{T}_1| &= |\hat{\mathbf{S}}_0| \left( \sum_{i=1}^N (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_1)^H \hat{\mathbf{S}}_0^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_1) + (\mathbf{x} - \alpha \mathbf{p} - \hat{\boldsymbol{\mu}}_1)^H \hat{\mathbf{S}}_0^{-1} (\mathbf{x} - \alpha \mathbf{p} - \hat{\boldsymbol{\mu}}_1) \right) \\ &= |\hat{\mathbf{S}}_0| (A + B). \end{aligned} \quad (2.17)$$

Now, let us rewrite the two terms  $A$  and  $B$  to separate the terms involving  $\alpha$ . By recalling that  $\hat{\boldsymbol{\mu}}_1 = \hat{\boldsymbol{\mu}}_0 - \frac{1}{N+1} \alpha \mathbf{p}$ , one obtains:

$$A = 1 + \frac{N|\alpha|^2}{(N+1)^2} \mathbf{p}^H \hat{\mathbf{S}}_0^{-1} \mathbf{p} + \frac{2}{N+1} \Re \left\{ \bar{\alpha} \mathbf{p}^H \hat{\mathbf{S}}_0^{-1} \sum_{i=1}^N (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_0) \right\}, \quad (2.18)$$

$$B = (\mathbf{x} - \hat{\boldsymbol{\mu}}_0)^H \hat{\mathbf{S}}_0^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_0) + \frac{N^2|\alpha|^2}{(N+1)^2} \mathbf{p}^H \hat{\mathbf{S}}_0^{-1} \mathbf{p} - \frac{2N}{N+1} \Re \left\{ \bar{\alpha} \mathbf{p}^H \hat{\mathbf{S}}_0^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_0) \right\}. \quad (2.19)$$

Notice that  $\mathbf{x} - \hat{\boldsymbol{\mu}}_0 = -\sum_{i=1}^N (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_0)$ , then rearranging the expression of  $(N+1)|\mathbf{T}_1|$  leads to:

$$\frac{(N+1)|\mathbf{T}_1|}{|\hat{\mathbf{S}}_0|} = \frac{(N+1)|\mathbf{T}_0|}{|\hat{\mathbf{S}}_0|} + \frac{N|\alpha|^2}{(N+1)} \mathbf{p}^H \hat{\mathbf{S}}_0^{-1} \mathbf{p} - 2 \Re \left\{ \bar{\alpha} \mathbf{p}^H \hat{\mathbf{S}}_0^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_0) \right\}. \quad (2.20)$$

Now, the term depending on  $\alpha$  can be rewritten as follows:

$$\frac{N}{(N+1)} \mathbf{p}^H \hat{\mathbf{S}}_0^{-1} \mathbf{p} \left| \alpha - \frac{N+1}{N} \frac{\mathbf{p}^H \hat{\mathbf{S}}_0^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_0)}{\mathbf{p}^H \hat{\mathbf{S}}_0^{-1} \mathbf{p}} \right|^2 - \frac{N+1}{N} \frac{\left| \mathbf{p}^H \hat{\mathbf{S}}_0^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_0) \right|^2}{\mathbf{p}^H \hat{\mathbf{S}}_0^{-1} \mathbf{p}}. \quad (2.21)$$

Minimizing  $|\mathbf{T}_1|$  w.r.t  $\alpha$  is equivalent to cancel the square term in the previous equation. Thus, the GLRT can now be written according to the following definition.

**Definition 2.2.1** (The generalized Kelly detector). *Under Gaussian assumptions, the extension of the Kelly's test when both the mean vector and the covariance matrix of the background are unknown takes the following form:*

$$\Lambda = \frac{\beta(N) \left| \mathbf{p}^H \hat{\mathbf{S}}_0^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_0) \right|^2}{(\mathbf{p}^H \hat{\mathbf{S}}_0^{-1} \mathbf{p}) \left( 1 + (\mathbf{x} - \hat{\boldsymbol{\mu}}_0)^H \hat{\mathbf{S}}_0^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_0) \right)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda, \quad (2.22)$$

where  $\beta(N) = \frac{N+1}{N}$ ,  $\lambda = \frac{\eta-1}{\eta}$  and

- $\hat{\mathbf{S}}_0 = \sum_{i=1}^N (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_0)(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_0)^H,$
- $\hat{\boldsymbol{\mu}}_0 = \frac{1}{N+1} \left( \mathbf{x} + \sum_{i=1}^N \mathbf{x}_i \right).$

Let us now comment on this new detector. One can notice that both the covariance matrix  $\hat{\mathbf{S}}_0$  as well as the mean  $\hat{\boldsymbol{\mu}}_0$  estimates depend on the data  $\mathbf{x}$  under test, which is not the case in other classical detectors where the unknown parameters are estimated from signal-free secondary data. Indeed,  $\hat{\mathbf{S}}_0$  depends on  $\hat{\boldsymbol{\mu}}_0$  which depends on  $\mathbf{x}$ . Consequently,  $\hat{\mathbf{S}}_0$  and  $\mathbf{x} - \hat{\boldsymbol{\mu}}_0$  are not independent. Moreover, the covariance matrix estimate  $\hat{\mathbf{S}}_0$  is not Wishart-distributed due to the non-standard mean estimate  $\hat{\boldsymbol{\mu}}_0$ . Thus, the derivation of this ratio distribution is very difficult.

As for previous detector, it would be intuitive to conjecture that the proposed test behaves as the classical Kelly's test but for  $N-1$  degrees of freedom. To prove that let us first rewrite Eq. (2.22) as follows:

$$\Lambda = \frac{\left| \mathbf{p}^H \hat{\mathbf{S}}_0^{-1} \mathbf{y} \right|^2}{(\mathbf{p}^H \hat{\mathbf{S}}_0^{-1} \mathbf{p}) \left( 1 + \frac{N}{N+1} \mathbf{y}^H \hat{\mathbf{S}}_0^{-1} \mathbf{y} \right)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda, \quad (2.23)$$

where we use:

- $(\mathbf{x} - \hat{\boldsymbol{\mu}}_0) = \frac{N}{N+1}(\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}),$
- $\hat{\boldsymbol{\mu}}_{SMV} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i,$
- $\mathbf{y} = \sqrt{\frac{N}{N+1}}(\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}) \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}).$

Now, let us set  $\hat{\mathbf{S}}_0^{(i)} = \sum_{i=1}^N (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_0^{(i)})(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_0^{(i)})^H$  where  $\hat{\boldsymbol{\mu}}_0^{(i)} = 1/N(\sum_{j \neq i}^N \mathbf{x}_j + \mathbf{x})$ . Then, the test becomes:

$$\frac{\frac{N+1}{N} \left| \mathbf{p}^H (\hat{\mathbf{S}}_0^{(i)})^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}) \right|^2}{\left( \mathbf{p}^H (\hat{\mathbf{S}}_0^{(i)})^{-1} \mathbf{p} \right) \left( 1 + (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})^H (\hat{\mathbf{S}}_0^{(i)})^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}) \right)}. \quad (2.24)$$

One can notice that each  $\mathbf{x}_i$  (including  $\mathbf{x}$ ) plays the same role, thus the distribution of this test is the same for every permutation of the  $(N+1)$ -sample  $(\mathbf{x}, \mathbf{x}_1, \dots, \mathbf{x}_N)$ . However, the dependency between the covariance matrix estimate and the data under test  $\mathbf{x}$  still remains.

To fill this gap, another way of taking advantage of the Kelly's detector when the mean vector is unknown can be to use the classical scheme recalled in Eq. (1.29) and to plug the classical estimator of the mean (two-step GLRT), based only on the secondary data, i.e.  $\hat{\boldsymbol{\mu}}_{SMV} = 1/N \sum_{i=1}^N \mathbf{x}_i$ . This leads to the plug-in Kelly's detector:

$$\Lambda_{Kelly, \hat{\boldsymbol{\Sigma}}, \hat{\boldsymbol{\mu}}}^{(N)} = \frac{|\mathbf{p}^H \hat{\boldsymbol{\Sigma}}_{SCM}^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})|^2}{\left( \mathbf{p}^H \hat{\boldsymbol{\Sigma}}_{SCM}^{-1} \mathbf{p} \right) \left( N + (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})^H \hat{\boldsymbol{\Sigma}}_{SCM}^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}) \right)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda. \quad (2.25)$$

In this case, the distribution can be derived. This is the purpose of the following proposition.

**Proposition 2.2.2.** *The theoretical relationship between the PFA and the threshold is given by*

$$PFA_{Kelly, \hat{\boldsymbol{\Sigma}}, \hat{\boldsymbol{\mu}}} = \frac{\Gamma(N)}{\Gamma(N-m+1) \Gamma(m-1)} \int_0^1 \left[ 1 + \frac{\lambda}{1-\lambda} \left( 1 - \frac{u}{N+1} \right) \right]^{m-N} u^{N-m} (1-u)^{m-2} du, \quad (2.26)$$

where  $\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Sigma}}_{SCM}$  and  $\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}}_{SMV}$ .

*Proof.* The detection test rewritten with  $\hat{\Sigma}_{SCM}^{-1} = N \mathbf{W}_{N-1}^{\hat{1}}$  becomes:

$$\Lambda_{Kelly \hat{\Sigma}, \hat{\mu}}^{(N)} = \frac{N^2 \left| \mathbf{p}^H \hat{\mathbf{W}}_{N-1}^{-1} (\mathbf{x} - \hat{\mu}) \right|^2}{N \left( \mathbf{p}^H \mathbf{W}_{N-1}^{\hat{1}} \mathbf{p} \right) \left( N + N \mathbf{y}^H \mathbf{W}_{N-1}^{\hat{1}} (\mathbf{x} - \hat{\mu}) \right)}, \quad (2.27)$$

and replacing  $(\mathbf{x} - \hat{\mu})$  by  $\sqrt{\frac{N+1}{N}} \mathbf{y}$ , one obtains:

$$\begin{aligned} \Lambda_{Kelly \hat{\Sigma}, \hat{\mu}}^{(N)} &= \frac{\frac{N+1}{N} N^2 \left| \mathbf{p}^H \hat{\mathbf{W}}_{N-1}^{-1} \mathbf{y} \right|^2}{N \left( \mathbf{p}^H \mathbf{W}_{N-1}^{\hat{1}} \mathbf{p} \right) \left( N + \frac{N+1}{N} N \mathbf{y}^H \mathbf{W}_{N-1}^{\hat{1}} \mathbf{y} \right)} \\ &= \frac{\left| \mathbf{p}^H \hat{\mathbf{W}}_{N-1}^{-1} \mathbf{y} \right|^2}{\left( \mathbf{p}^H \mathbf{W}_{N-1}^{\hat{1}} \mathbf{p} \right) \left( \frac{N}{N+1} + \mathbf{y}^H \mathbf{W}_{N-1}^{\hat{1}} \mathbf{y} \right)}, \end{aligned} \quad (2.28)$$

with  $\mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \Sigma)$ .

The classical Kelly detector obtained when the mean vector is known is recalled here, built with  $N - 1$  zero-mean Gaussian data, and written with  $\hat{\mathbf{W}}_{N-1}$ :

$$\Lambda_{Kelly \hat{\Sigma}}^{(N-1)} = \frac{\left| \mathbf{p}^H \hat{\mathbf{W}}_{N-1}^{-1} \mathbf{y} \right|^2}{\left( \mathbf{p}^H \mathbf{W}_{N-1}^{\hat{1}} \mathbf{p} \right) \left( 1 + \mathbf{y}^H \mathbf{W}_{N-1}^{\hat{1}} \mathbf{y} \right)}. \quad (2.29)$$

It is worth pointing out that the term  $N/(N+1)$  resulting from the mean estimation in  $\Lambda_{Kelly \hat{\Sigma}, \hat{\mu}}^{(N)}$  does not appear in the classical Kelly detector Eq. (2.29). This fact prevents from relating the two expressions. Thus, a proof similar to the Proposition 2.2.1 is not feasible.

According to Kraut et al. (2001), Richmond (2000), an equivalent LR can be expressed as:

$$\hat{\kappa}^2 = \frac{\Lambda_{Kelly \hat{\Sigma}, \hat{\mu}}^{(N)}}{1 - \Lambda_{Kelly \hat{\Sigma}, \hat{\mu}}^{(N)}} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \frac{\lambda}{1 - \lambda}. \quad (2.30)$$

Following the same development proposed in Kraut et al. (2001), the statistic  $\hat{\kappa}^2$  can be identified as the ratio  $\theta/\beta$  between two independent scalar random variables  $\theta$  and  $\beta$ . For this particular development of Kelly distribution with non-centered data, the scalar random variable  $\beta$  is found to have the same distribution as the function  $1 - u/(N+1)$  where  $u$  is a random variable following a complex central beta distribution with  $N - m + 1, m - 1$  degrees of freedom:

$$u \sim f_u(u) = \frac{\Gamma(N)}{\Gamma(N - m + 1) \Gamma(m - 1)} u^{N-m} (1 - u)^{m-2},$$

whereas the PDF of the variable  $\theta$  is distributed according to the complex  $F$ -distribution with  $1, N - m$  degrees of freedom scaled by  $1/(N - m)$ :

$$\theta \sim f_\theta(\theta) = (N - m) (1 + \theta)^{m-N-1}. \quad (2.31)$$

One can now derive the cumulative density function of the Kelly test as:

$$\begin{aligned} \mathbb{P} \left( \Lambda_{\text{Kelly}}^{(N)} \hat{\Sigma}, \hat{\mu} \leq \lambda \right) &= \mathbb{P} \left( \hat{\kappa}^2 \leq \frac{\lambda}{1 - \lambda} \right) = \mathbb{P} \left( \theta \leq \beta \frac{\lambda}{1 - \lambda} \right) \\ &= \int_0^1 \left[ \int_0^{\frac{\lambda}{1-\lambda} (1-u/(N+1))} f_\theta(v) dv \right] f_u(u) du. \end{aligned} \quad (2.32)$$

Solving the integral one obtains the ‘‘PFA-threshold’’ relationship:

$$\begin{aligned} PFA_{\text{Kelly}} \hat{\Sigma}, \hat{\mu} &= \\ \frac{\Gamma(N)}{\Gamma(N - m + 1) \Gamma(m - 1)} \int_0^1 \left[ 1 + \frac{\lambda}{1 - \lambda} \left( 1 - \frac{u}{N + 1} \right) \right]^{m-N} u^{N-m} (1 - u)^{m-2} du. \end{aligned} \quad (2.33)$$

$$(2.34)$$

However, the final expression cannot be further simplified and a closed-form expression as those obtained for the other detectors cannot be determined.  $\square$

### 2.2.3 Adaptive Normalized Matched Filter

Similarly, the ANMF for both mean vector and covariance matrix estimation becomes:

$$\Lambda_{\text{ANMF}} \hat{\Sigma}, \hat{\mu} = \frac{|\mathbf{p}^H \hat{\Sigma}^{-1} (\mathbf{x} - \hat{\mu})|^2}{(\mathbf{p}^H \hat{\Sigma}^{-1} \mathbf{p}) \left( (\mathbf{x} - \hat{\mu})^H \hat{\Sigma}^{-1} (\mathbf{x} - \hat{\mu}) \right)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda. \quad (2.35)$$

**Proposition 2.2.3.** *The theoretical relationship between the PFA and the threshold is given by*

$$PFA_{\text{ANMF}} \hat{\Sigma}, \hat{\mu} = (1 - \lambda)^{a-1} {}_2F_1(a, a - 1; b - 1; \lambda), \quad (2.36)$$

where  $a = (N - 1) - m + 2$  and  $b = (N - 1) + 2$ ,  $\hat{\Sigma} = \hat{\Sigma}_{SCM}$  and  $\hat{\mu} = \hat{\mu}_{SMV}$ .

*Proof.* The proof is similar to the proof of Proposition 2.2.1. The main difference is due to the normalization term  $(\mathbf{x} - \hat{\mu})^H \hat{\Sigma}^{-1} (\mathbf{x} - \hat{\mu})$ . Indeed, the correction factor  $N/(N - 1)$  appears both at the numerator and at the denominator, and consequently,

it can be simplified. The same argument is also true for the factor  $N$  that arises from the covariance matrix estimates, i.e. since the detector is homogeneous in terms of covariance matrix estimates, this scalar also disappears. Thus, the distribution of the ANMF with an estimate of the mean is exactly the same as in Eq. (1.37) where  $N$  is replaced by  $N - 1$ .  $\square$

## 2.2.4 Simulations

In this section, we validate the theoretical analysis on simulated data. The experiments were conducted on  $m = 5$  dimensional Gaussian vectors, for different values of  $N$ , the number of secondary data and the computations have been made through  $10^6$  Monte-Carlo trials. The true covariance is chosen as a Toeplitz matrix whose entries are  $\Sigma_{i,j} = \rho^{|i-j|}$  and where  $\rho = 0.4$ . The mean vector is arbitrarily set to have all entries equal to  $(3 + 4j)$ .

### 2.2.4.1 False Alarm Regulation

The FA regulation is presented for previous detection schemes having a closed-form expression, i.e. for all except the generalized Kelly detector.

Fig. 2.1 shows the false-alarm regulation for the MF, the AMF when only covariance matrix is unknown and the AMF where both the covariance matrix and the mean vector are unknown for different values of  $N$ . The perfect agreement between the green and yellow curves illustrates the results of Proposition 2.2.1. Moreover, remark that when  $N$  increases both AMFs get closer to each other, and they approach the known parameters case characterized by the MF. Since the mean vector (Eq. (1.5)) and the covariance matrix (Eq. (1.6)) estimates tend to the true value of the mean  $\boldsymbol{\mu}$  and the true value of the covariance matrix  $\boldsymbol{\Sigma}$ , when  $N$  tends to infinity we also have the convergence in the probability of  $\Lambda_{AMF\hat{\boldsymbol{\Sigma}},\hat{\boldsymbol{\mu}}}$  to the MF Eq. (1.19).

$$\mathbb{P}(\Lambda_{AMF\hat{\boldsymbol{\Sigma}},\hat{\boldsymbol{\mu}}} > \lambda) \underset{N \rightarrow \infty}{\sim} \mathbb{P}(\Lambda_{AMF\hat{\boldsymbol{\Sigma}}} > \lambda) \xrightarrow[N \rightarrow \infty]{} \mathbb{P}(\Lambda_{MF} > \lambda) \quad (2.37)$$

Consequently, when  $N$  tends to infinity, one can observe a convergence of Eq. (2.2) to Eq. (1.22), and both of them to Eq. (1.20), as shown on Fig. 2.1. This fact also occurs for the ANMF and for the Kelly test in Eq. (2.26) which converges to the classical Kelly in Eq. (1.30).

Fig. 2.2 and Fig. 2.3 present the FA regulation for the Kelly detector and the ANMF respectively, under Gaussian assumption. For clarity purposes, the results are displayed in terms of the threshold  $\eta$  from Eq. (1.29),  $\eta = (1 - \lambda)^{-(N+1)}$ , and Eq. (1.34),  $\eta = (1 - \lambda)^{-m}$ , respectively and a logarithmic scale is used. This validates results of Proposition 2.2.2 and 2.2.3 for the SCM-SMV. The same behavior when  $N \rightarrow \infty$  that appears in Fig. 2.1 is still valid for the Kelly's detection test and for the ANMF.

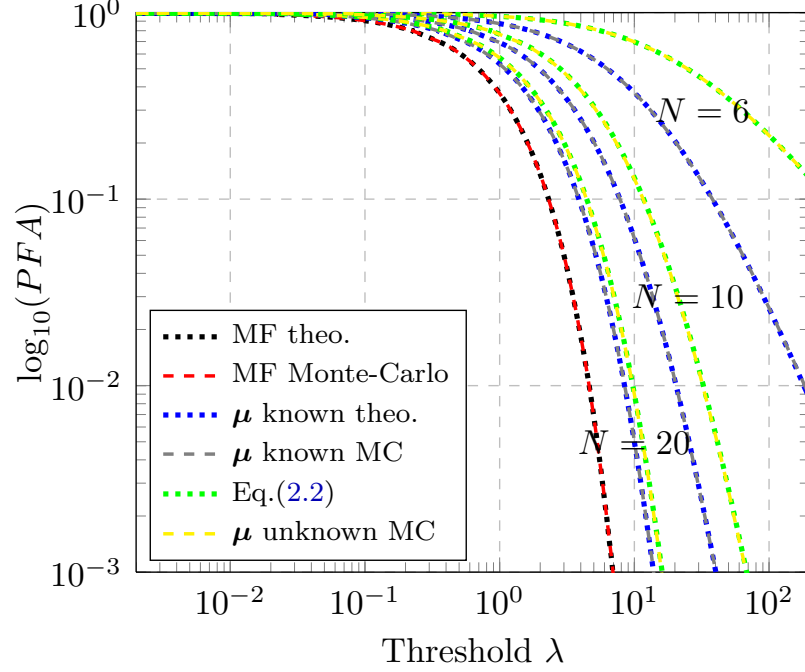


Figure 2.1: PFA versus threshold for the AMF when (1)  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  are known (MF) (red and black curves) (2) only  $\boldsymbol{\mu}$  is known (gray and blue curves) (3) Proposition 2.2.1: both  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  are unknown (yellow and green curves).

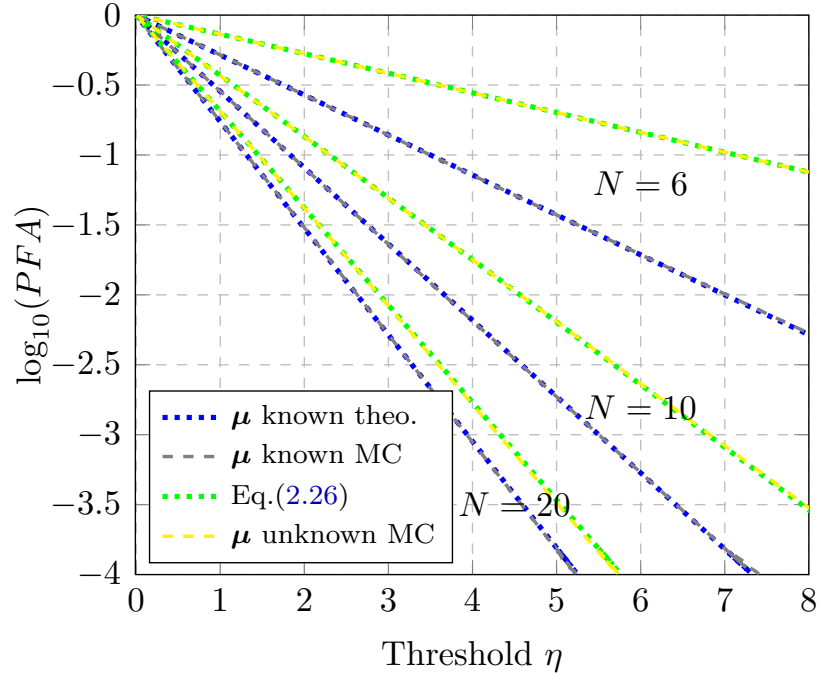


Figure 2.2: PFA versus threshold for the "plug-in" Kelly detector when (1) only  $\boldsymbol{\mu}$  is known (gray and blue curves) (2) Proposition 2.2.2: both  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  are unknown (yellow and green curves)



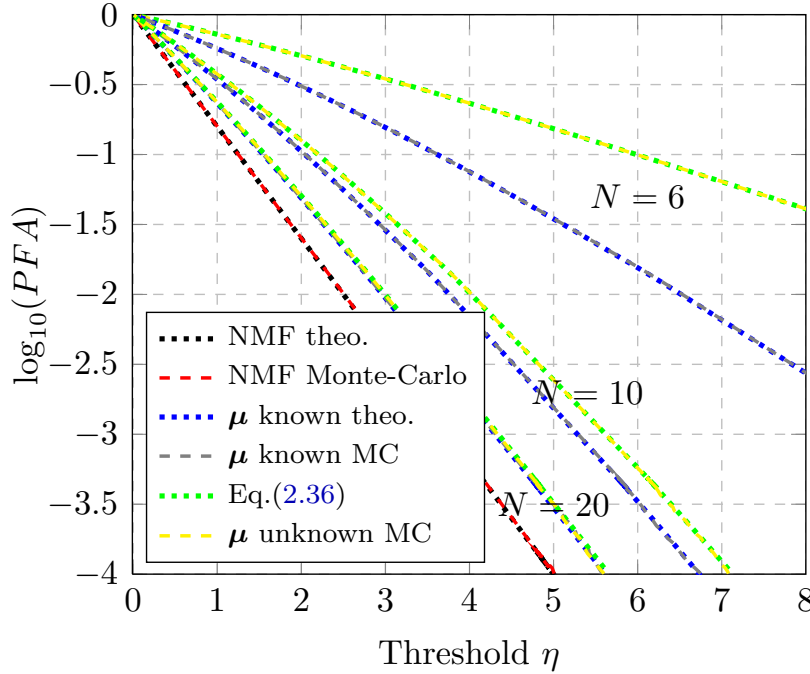


Figure 2.3: PFA versus threshold for the ANMF when (1)  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  are known (NMF) (red and black curves) (2) only  $\boldsymbol{\mu}$  is known (gray and blue curves) (3) Proposition 2.2.3: both  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  are unknown (yellow and green curves)

Remark that the derived relationships given by Eq. (2.2) and Eq. (2.36) are quite similar to those for which the mean is known. However, as illustrated in Fig. 2.1 and Fig. 2.3, there is an important difference for small values of  $N$ .

It is worth pointing out that the theoretical “PFA-threshold” relationships presented above depend only on the size of the vectors  $m$  and the number of secondary data used to estimate the parameters  $N$ . Thus, the detector outcome will not depend on the true value of the covariance matrix or the mean vector. These three detectors hold the CFAR property with respect to the background parameters. However, their distribution strongly relies on the underlying distribution of the background, i.e. if Gaussian assumption is not fulfilled the “PFA-threshold” relationship will divert from the theoretical results derived in this paper.

#### 2.2.4.2 Performance Evaluation

The four detection schemes are compared in terms of probability of detection. Firstly, one sets the probability of false alarm to a specific value. Here we set  $PFA = 10^{-3}$  with  $m = 5$  and  $N = 10$ . Then, the threshold is adjusted to reach the desired PFA, according to the false alarm regulation curves described above. For the generalized Kelly detector, the threshold is empirically computed to ensure the same  $PFA = 10^{-3}$ . The SNR is obtained according to:  $SNR = |\alpha^2| \mathbf{p}^H \boldsymbol{\Sigma}^{-1} \mathbf{p}$ . Hence, one can add artificial targets with steering vector  $\mathbf{p}$  and the variation on the amplitude

$\alpha$  will lead to the different SNR values. The artificial targets signature used for the simulations is the unity vector  $\mathbf{p} = [1, \dots, 1]^T$  which corresponds to the non-prior approach that does not introduce any information into the detector.

Fig. 2.4 presents the detection probability versus the SNR. When data follow a multivariate normal distribution, the detectors delivering the best performance results are the Kelly detectors (“Plug-in” and generalized). Actually, these detectors lead to very similar performance with a small improvement of the generalized (resp. “plug in”) one at low (resp. high) SNR. As expected, the AMF and the ANMF require a higher SNR to achieve same probability of detection. For instance, for a given  $PD = 0.8$ , the gain of Kelly’s test is of 3 dBs compared to the ANMF.

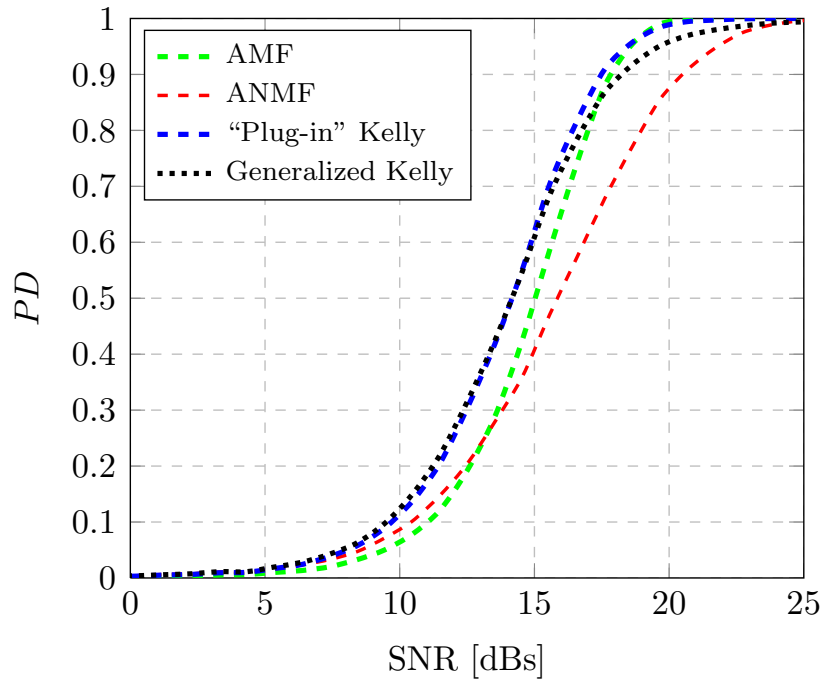


Figure 2.4: Probability of detection for different SNR values and  $PFA = 10^{-3}$  in Gaussian case.

## 2.3 Anomaly Detection

Target detection methods presented above require the knowledge of the spectra steering vector  $\mathbf{p}$  of the desired targets. One could be interested in a large number of possible targets each with different signatures. Thus, the variety of sought spectra corresponding to the different kind of targets and the difficulties due to the atmospheric compensation for the measured spectral signatures (used as steering vectors) have led to the derivation of new algorithms that intend to distinguish unusual materials in a scene without reference to target signatures. In this section, we are focused on anomaly detection (see e.g. [Chang and Chiang \(2002\)](#) and references therein). It can be interpreted as a particular case of target detection in which no a priori information about the spectra of the targets of interest is available. Hence, the aim of anomaly detection is to locate objects in the image that are anomalous with respect to the background. The type of interesting targets can differ significantly from one application to another, e.g. in forestry applications infected trees are the anomalies of interest, whereas in defense and intelligence applications the anomalies to be detected are usually vehicles. Note that, since anomaly detectors do not use any a priori knowledge, they cannot distinguish between true targets and detections of bright pixels of the background or targets that are not of interest. This fact makes extremely difficult to define a false alarm rate for the detectors as highlighted in [Matteoli et al. \(2010b\)](#).

Anomalies are defined with reference to a model of the background. As for the previous target detection methods, the background model is developed adaptively using reference data. Most of these methods rely on classical Gaussian distribution assumption and need for the statistical characterization of the background usually through first and second order parameters (i.e. the mean vector and the covariance matrix). In this case, the reference data are taken either from a local neighborhood around the observation vector either using all the pixels in the image. Both approaches have their benefits (see e.g. [Stein et al. \(2002\)](#)). Local strategy provides more realistic scenario for the background characterization. However, it can be susceptible to the presence of false alarms due to isolated anomalies. While the global approach is not likely to generate this kind of false alarms, it will decrease the detection capability for isolated targets. Local procedures will be considered in the following for the different detection schemes.

Anomaly detection methods studied in this section are analyzed in the real case as they were originally derived for HSI. Let us now recap here the most popular Gaussian based anomaly detectors (see e.g. [Nasrabadi \(2014\)](#) for a complete survey in anomaly detection methods).

### 2.3.1 Reed-Xiaoli Detector

The Reed-Xiaoli Detector (RXD) proposed in [Reed and Yu \(1990\)](#) is commonly considered as the benchmark anomaly detector for hyperspectral data. The considered

signal model can be written as:

$$\begin{cases} \mathcal{H}_0 : \mathbf{x}_i = \mathbf{b}_i, & , i = 1, \dots, N \\ \mathcal{H}_1 : \mathbf{x}_i = \alpha_i \mathbf{p} + \mathbf{b}_i, & , i = 1, \dots, N, \end{cases} \quad (2.38)$$

where  $\mathbf{x}_i$  are the  $N$  available data vectors on the image of dimension  $m$ .  $\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$  represents the residual background,  $\mathbf{p}$  is the spectral signature of the possible anomalous material assumed to be unknown; and  $\alpha_i$  stands for the amplitude of the intended targets through the  $N$  available data, i.e. it is a known vector  $\boldsymbol{\alpha} = [\alpha_1 \dots \alpha_N]^T$  of dimension  $N$  that indicates the strength and position of the sought targets over the image. Remark that each vector from the available data can potentially contain an anomaly while in classical detection problem secondary data are assumed to be signal-free as stated in Eq. (1.9).

Thus, one can arrange the vector data on a matrix as:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1(1) & \dots & \mathbf{x}_N(1) \\ \vdots & \ddots & \vdots \\ \mathbf{x}_1(m) & \dots & \mathbf{x}_N(m) \end{bmatrix}, \quad (2.39)$$

and the joint PDF under real Gaussian distribution takes the form:

$$f_i(\mathbf{X}) = \left( \frac{1}{(2\pi)^{\frac{m}{2}} |\mathbf{\Sigma}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2} \text{Tr}(\mathbf{\Sigma}^{-1} \mathbf{T}_i)\right] \right)^N \quad \text{for } i = 0, 1. \quad (2.40)$$

Similarly to Kelly (1986), the covariance matrix is supposed to be unknown and the LR has to be maximized on a first step w.r.t. the matrix  $\mathbf{\Sigma}$ . Thus, the matrices maximizing the PDF  $f_i$  are simply  $\mathbf{T}_i$  and can be written according to:

$$\mathbf{T}_0 = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i^T = \frac{1}{N} \mathbf{X} \mathbf{X}^T, \quad (2.41)$$

$$\mathbf{T}_1 = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \alpha_i \mathbf{p})(\mathbf{x}_i - \alpha_i \mathbf{p})^T = \frac{1}{N} (\mathbf{X} - \mathbf{p} \boldsymbol{\alpha})(\mathbf{X} - \mathbf{p} \boldsymbol{\alpha})^T, \quad (2.42)$$

and the GLRT, neglecting the exponent  $N/2$  is given by:

$$\Lambda(\mathbf{X}, \mathbf{p}) = \frac{|\mathbf{T}_0|}{|\mathbf{T}_1|} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta. \quad (2.43)$$

It remains to maximize this expression over the unknown spectral signature  $\mathbf{p}$  and the resulting MLE takes the form:

$$\hat{\mathbf{p}} = \frac{\mathbf{X} \boldsymbol{\alpha}^T}{\boldsymbol{\alpha} \boldsymbol{\alpha}^T}. \quad (2.44)$$

Thus, replacing  $\mathbf{p}$  by Eq. (2.44) into Eq. (2.43) yields the test:

$$\Lambda(\mathbf{X}) = \frac{|\mathbf{X}\mathbf{X}^T|}{\left| \mathbf{X}\mathbf{X}^T - \frac{(\mathbf{X}\boldsymbol{\alpha}^T)(\mathbf{X}\boldsymbol{\alpha}^T)^T}{\boldsymbol{\alpha}\boldsymbol{\alpha}^T} \right|}, \quad (2.45)$$

which is equivalent to the benchmark detector:

$$\Lambda(\mathbf{X}) = \frac{(\mathbf{X}\boldsymbol{\alpha}^T)^T(\mathbf{X}\mathbf{X}^T)^{-1}(\mathbf{X}\boldsymbol{\alpha}^T)}{\boldsymbol{\alpha}\boldsymbol{\alpha}^T}. \quad (2.46)$$

Since hyperspectral data are not zero mean, let us now consider that the background  $\mathbf{b}_i$  is distributed according to  $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  and the mean vector  $\boldsymbol{\mu}$  is supposed to be known. In the case just one anomaly in the data under test is intended to be detected, the corresponding amplitude vector can be written as  $\boldsymbol{\alpha}_i = [0 \dots 0 \ 1 \ 0 \dots 0]^T$  where 1 is at the  $i^{th}$  position and the previous detector takes the form:

$$\Lambda_{RXD} = (\mathbf{x}_i - \boldsymbol{\mu})^T \hat{\boldsymbol{\Sigma}}_{SCM}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda. \quad (2.47)$$

The distribution of the previous test is detailed in [Reed and Yu \(1990\)](#) and under  $\mathcal{H}_0$  hypothesis takes the following form:

$$\Lambda_{RXD} \sim f_{RXD|\mathcal{H}_0}(\lambda) = \frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-m}{2}\right) \Gamma\left(\frac{m}{2}\right)} (1-\lambda)^{\frac{N-m-2}{2}} \lambda^{\frac{m-2}{2}} \quad (2.48)$$

which corresponds to the  $\beta$  distribution with  $\frac{m}{2}$  and  $\frac{N-m}{2}$  degrees of freedom. The PFA is obtained after integrating:

$$PFA = \int_{\lambda_0}^1 f_{RXD|\mathcal{H}_0} d\lambda \quad (2.49)$$

for a chosen threshold  $\lambda_0$ .

Finally, since the mean vector is usually unknown, it can be replaced on the detector in by its estimate  $\hat{\boldsymbol{\mu}}_{SMV}$ . The resulting detector, commonly known as two-step GLRT, yields:

$$\Lambda_{ARXD} = (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{SMV})^T \hat{\boldsymbol{\Sigma}}_{SCM}^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{SMV}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda. \quad (2.50)$$

The covariance matrix estimation  $\hat{\boldsymbol{\Sigma}}_{SCM}$  in Eq. (2.50), is performed over all the data set, i.e. including the vector  $\mathbf{x}_i$  under test. In the following, the test in Eq. (2.50) will be referred as the Adaptive RXD (ARXD), to underline the fact that the unknown

mean vector is replaced by its estimate.

Let us now discuss the presence of one anomaly  $\mathbf{s}_0$  in the secondary data and its impact on the covariance matrix estimate for the RXD scheme. According to the model in Eq. (2.38), the anomalous signal can be written as  $\mathbf{s}_0 = \alpha_0 \mathbf{p}_0$  where  $\alpha_0$  is the amplitude of the signal; and the received signal  $\mathbf{x} = \mathbf{b} + \mathbf{s}_0$ . Let us assume without loss of generality that the mean of the noise  $\mathbf{b}$  is  $\boldsymbol{\mu} = 0$ . In presence of an anomaly, the covariance matrix estimate becomes:

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N} \left( \sum_{i=1}^{N-1} \mathbf{b}_i \mathbf{b}_i^T + \mathbf{s}_0 \mathbf{s}_0^T \right) = \hat{\boldsymbol{\Sigma}}_{\mathbf{bb}} + \frac{\alpha_0^2}{N} \mathbf{p}_0 \mathbf{p}_0^T. \quad (2.51)$$

By the matrix inversion lemma, the inverse of the covariance estimate yields:

$$\hat{\boldsymbol{\Sigma}}^{-1} = \hat{\boldsymbol{\Sigma}}_{\mathbf{bb}}^{-1} - \frac{\hat{\boldsymbol{\Sigma}}_{\mathbf{bb}}^{-1} \mathbf{p}_0 \mathbf{p}_0^T \hat{\boldsymbol{\Sigma}}_{\mathbf{bb}}^{-1}}{N \alpha_0^{-2} + \mathbf{p}_0^T \hat{\boldsymbol{\Sigma}}_{\mathbf{bb}}^{-1} \mathbf{p}_0}. \quad (2.52)$$

In case the test pixel contains an anomaly, we consider that the background signal is negligible compared to the strong target. Thus, the quadratic form can be written as:

$$\mathbf{s}_0^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{s}_0 = \left( \frac{\alpha^2}{N + \alpha^2 \mathbf{p}_0^T \hat{\boldsymbol{\Sigma}}_{\mathbf{bb}}^{-1} \mathbf{p}_0} \right) \mathbf{p}_0^T \hat{\boldsymbol{\Sigma}}_{\mathbf{bb}}^{-1} \mathbf{p}_0. \quad (2.53)$$

On the other hand, if the test pixel is only constituted of background, one obtains:

$$\mathbf{b}_i^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{b}_i = \left( \frac{N}{N + \alpha^2 \mathbf{p}_0^T \hat{\boldsymbol{\Sigma}}_{\mathbf{bb}}^{-1} \mathbf{p}_0} \right) \mathbf{b}_i^T \hat{\boldsymbol{\Sigma}}_{\mathbf{bb}}^{-1} \mathbf{b}_i. \quad (2.54)$$

Let us now analyze the effect of a great amplitude of the anomaly present in the secondary data:

$$\lim_{\alpha \rightarrow \infty} \left( \frac{\alpha^2}{N + \alpha^2 \mathbf{p}_0^T \hat{\boldsymbol{\Sigma}}_{\mathbf{bb}}^{-1} \mathbf{p}_0} \right) \mathbf{p}_0^T \hat{\boldsymbol{\Sigma}}_{\mathbf{bb}}^{-1} \mathbf{p}_0 = 1 \quad (2.55)$$

$$\lim_{\alpha \rightarrow \infty} \left( \frac{N}{N + \alpha^2 \mathbf{p}_0^T \hat{\boldsymbol{\Sigma}}_{\mathbf{bb}}^{-1} \mathbf{p}_0} \right) \mathbf{b}_i^T \hat{\boldsymbol{\Sigma}}_{\mathbf{bb}}^{-1} \mathbf{b}_i = 0 \quad (2.56)$$

Therefore, if the target is very strong, it will be still detected according to Eq. (2.55). However, the dynamic range of the image is reduced to the values between 0 and 1 due to the presence of an outlier in the secondary data. Finally, the outlier will be correctly suppressed if the signal only contains the background as in Eq. (2.56).

### 2.3.2 Kelly Anomaly Detector

Let us now detail a classical anomaly detector often mistakenly referred to as the RXD.

Following the development proposed in Kelly (1986) and recalled in Section 1.5.2, we

assume the same signal model as in Eq. (1.15), i.e. an observation vector  $\mathbf{x}$  that may contain the target and  $N$  signal free secondary data  $\mathbf{x}_i$ . However, the amplitude of the signal  $\alpha$  is supposed to be known and the unknown parameter is now the steering vector  $\mathbf{p}$ . As in classical Kelly detector, the covariance matrix is unknown. In order to be coherent with Section 2.3.1, let us consider without loss of generality real Gaussian distribution (instead of complex distribution). The likelihood functions under  $\mathcal{H}_0$  and  $\mathcal{H}_1$  are analogous to those given in Eq. (1.24) and their maxima w.r.t to the covariance matrix are achieved taking Eq. (1.25) and Eq. (1.26), in the real case. Hence, the LR takes the same form than in Eq. (2.43) and one has to maximize it now w.r.t. the spectral signature  $\mathbf{p}$ . Then, one can write from Eq. (1.25) and Eq. (1.26),  $(N+1)|\mathbf{T}_i|$  according to:

$$(N+1)|\mathbf{T}_0| = |\hat{\mathbf{W}}| \left( 1 + (\mathbf{x} - \boldsymbol{\mu})^T \hat{\mathbf{W}}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right), \quad (2.57)$$

$$(N+1)|\mathbf{T}_1| = |\hat{\mathbf{W}}| \left( 1 + (\mathbf{x} - (\alpha\mathbf{p} + \boldsymbol{\mu}))^T \hat{\mathbf{W}}^{-1} (\mathbf{x} - (\alpha\mathbf{p} + \boldsymbol{\mu})) \right). \quad (2.58)$$

And the minimum of Eq. (2.58) is obtained at:

$$\hat{\mathbf{p}} = \frac{\mathbf{x} - \boldsymbol{\mu}}{\alpha}. \quad (2.59)$$

After replacing  $\mathbf{p}$  by Eq. (2.59) into Eq. (2.43), it is easy to show that the resulting GLRT is equivalent to:

$$\Lambda_{KellyAD \hat{\Sigma}} = (\mathbf{x} - \boldsymbol{\mu})^T \hat{\Sigma}_{CSCM}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda. \quad (2.60)$$

The quadratic form in Eq. (2.60) corresponds to the Mahalanobis distance detailed in [Mahalanobis \(1936\)](#). It performs statistically as an outlier detector. When Gaussian assumption is valid, the quadratic form  $(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})$  follows a  $\chi^2$ -distribution with  $m$  degrees of freedom for  $\Sigma$  and  $\boldsymbol{\mu}$  perfectly known. In case the parameter  $\Sigma$  is replaced by its MLE, the CSCM, the distribution of the quadratic form can be written according to (see [Bilodeau and Brenner \(1999\)](#)):

$$\Lambda_{KellyAD \hat{\Sigma}}^{(N)} = (\mathbf{x} - \boldsymbol{\mu})^T \hat{\Sigma}_{CSCM}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \sim T^2, \quad (2.61)$$

becomes a Hotelling  $T^2$  distribution and thus,

$$\frac{N-m+1}{mN} \Lambda_{KellyAD \hat{\Sigma}}^{(N)} \sim F_{m, N-m+1} \quad (2.62)$$

where  $F_{m, N-m+1}$  is the non-central  $F$ -distribution with  $m$  and  $N-m+1$  degrees of freedom [Weinstein \(2010\)](#) and the superscript  $^{(N)}$  is used to stress the dependence on the number of secondary data  $N$ . For high values of  $N$ , ( $N > 10m$ ), the distribution can be approximated by the  $\chi^2$ -distribution.

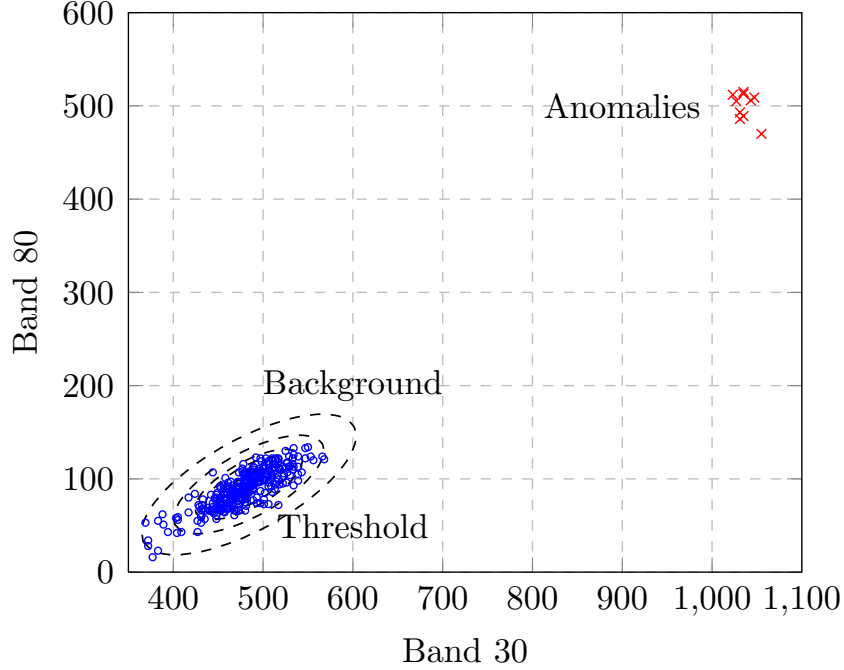


Figure 2.5: Graphical interpretation in two-dimensional space for the  $\Lambda_{KellyAD}$  detector.

As discussed above, when the mean vector is unknown, it can be replaced on the detector (two-step GLRT) by its MLE leading to:

$$\Lambda_{KellyAD \hat{\Sigma}, \hat{\mu}}^{(N)} = (\mathbf{x} - \hat{\mu}_{SMV})^T \hat{\Sigma}_{SCM}^{-1} (\mathbf{x} - \hat{\mu}_{SMV}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda. \quad (2.63)$$

Fig. 2.5 provides a graphical interpretation of the detector in a two dimensional space.  $\Lambda_{KellyAD \hat{\Sigma}}$  decision surfaces correspond to ellipsoids in multidimensional spaces. Dots in blue correspond to water pixels and isolated red crosses represent the anomalous pixels for ground materials.

**Remark 2.3.1.** Interestingly, note that  $\Lambda_{RXD}$  (resp.  $\Lambda_{ARXD}$ ) and the  $\Lambda_{KellyAD \hat{\Sigma}}$  (reps.  $\Lambda_{KellyAD \hat{\Sigma}, \hat{\mu}}^{(N)}$ ) differ only from the fact that the vector  $\mathbf{x}$  under test is also present in the covariance matrix estimation in Eq. (2.50). Therefore, in  $\Lambda_{RXD}$ , the  $N$  secondary data are not assumed to be signal free and the proposed detector aims to compare every sample to the covariance matrix over all the samples. While in the second approach,  $\Lambda_{KellyAD \hat{\Sigma}}$ , one intends to differentiate the observation vector from the background statistically characterized using  $N$  samples. Hence,  $N + 1$  vectors are available in the latter and  $\Lambda_{KellyAD \hat{\Sigma}}$  does not represent anymore a benchmark structure. Often, the local Kelly detector is mistakenly referred as the local RXD when the users, either remove the vector  $\mathbf{x}_i$  from the secondary data or they prevent it to be part of this set by using a guard window.



The distribution of this detection test is given in the next Proposition.

**Proposition 2.3.1.** *The distribution of the detector under Gaussian assumption is given by*

$$\frac{N-m}{m(N+1)} \Lambda_{\text{KellyAD } \hat{\Sigma}, \hat{\mu}}^{(N)} \sim F_{m, N-m}, \quad (2.64)$$

with  $F_{m, N-m}$  is the non-central  $F$ -distribution with  $m$  and  $N-m$  degrees of freedom and  $\hat{\Sigma} = \hat{\Sigma}_{SCM}$  and  $\hat{\mu} = \hat{\mu}_{SMV}$ .

*Proof.* The proof is similar to the proof of Proposition 2.2.1. For simplicity matters, the following notations are used:  $\hat{\Sigma} = \hat{\Sigma}_{SCM}$  and  $\hat{\mu} = \hat{\mu}_{SMV}$ .

Let us set  $\forall i = 1, \dots, N$ ,  $\mathbf{x}_i \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$  and  $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ , where all these vectors are independent. Now, let us denote

$$\hat{\mathbf{W}}_{N-1} = \sum_{i=1}^N (\mathbf{x}_i - \hat{\mu})(\mathbf{x}_i - \hat{\mu})^T = N \hat{\Sigma}_{SCM}.$$

Note that as an application of the Cochran theorem (see e.g. [Anderson \(1984\)](#)), one has

$$\hat{\mathbf{W}}_{N-1} \stackrel{\text{dist.}}{=} \sum_{i=1}^{N-1} (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T = (N-1) \hat{\Sigma}_{SCM}.$$

Since  $\hat{\mu} \sim \mathcal{N}(\boldsymbol{\mu}, \frac{1}{N}\Sigma)$ , and  $\hat{\mu}$  is independent of  $\mathbf{x}$  by construction, one has  $\mathbf{x} - \hat{\mu} \sim \mathcal{N}(\mathbf{0}, \frac{N+1}{N}\Sigma)$ . This can be equivalently rewritten as

$$\mathbf{y} = \sqrt{N/(N+1)}(\mathbf{x} - \hat{\mu}) \sim \mathcal{N}(\mathbf{0}, \Sigma).$$

As we jointly estimate the mean and the covariance matrix, a degree of freedom is lost, compared with the only covariance matrix estimation problem.

Let us now consider  $\Lambda_{\text{KellyAD } \hat{\Sigma}}^{(N-1)}$  (i.e.  $\boldsymbol{\mu}$  known) built from  $N-1$  secondary data, rewritten in terms of  $\hat{\mathbf{W}}_{N-1}$ :

$$\Lambda_{\text{KellyAD } \hat{\Sigma}}^{(N-1)} = (N-1) \left( (\mathbf{x} - \boldsymbol{\mu})^T \hat{\mathbf{W}}_{N-1}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right)$$

where  $(\mathbf{x} - \boldsymbol{\mu}) \sim \mathcal{N}(\mathbf{0}, \Sigma)$  and whose distribution is given by Eq. (2.62) where  $N$  is replaced by  $N-1$ .

Now, for the joint estimation problem, the  $\Lambda_{\text{KellyAD } \hat{\Sigma}, \hat{\mu}}$  can be rewritten as:

$$\begin{aligned} \Lambda_{\text{KellyAD } \hat{\Sigma}, \hat{\mu}}^{(N)} &= N \left( (\mathbf{x} - \hat{\mu})^T \hat{\mathbf{W}}_{N-1}^{-1} (\mathbf{x} - \hat{\mu}) \right) \\ &= N \frac{N+1}{N} \left( \mathbf{y}^T \hat{\mathbf{W}}_{N-1}^{-1} \mathbf{y} \right) \\ &\stackrel{\text{dist.}}{=} \frac{N+1}{N-1} \Lambda_{\text{KellyAD } \hat{\Sigma}}^{(N-1)} \end{aligned}$$

This concludes the proof.  $\square$

The “PFA-threshold” relationship is easily obtained as the complementary CDF of the detector distribution, as pointed in Remark 1.5.1.

**Remark 2.3.2.** *It is worth pointing out from Eq. (2.63) that  $\Lambda_{\text{KellyAD}}^{\hat{\Sigma}, \hat{\mu}}$  performs similarly to a matched filter structure applied to  $\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}$ :*

$$\Lambda(\mathbf{x}) = c \mathbf{H}^T (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}), \quad (2.65)$$

where  $\mathbf{H}^T$  is the matched signal and  $c$  a constant that can be also a function on  $\mathbf{x}$ . The expression in Eq. (2.65) is completely characterized by the matched signal  $\mathbf{H}^T$  and the scale constant  $c$ . Hence, one can identify from Eq. (2.63) the matched signal  $\mathbf{H}^T = (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})^T \hat{\Sigma}_{SCM}^{-1}$  and  $c = 1$ .

### 2.3.3 Normalized-RXD and Uniform Target Detector

Following the same approach than in Eq. (2.65), one can derive many different anomaly detection schemes. We recall here two popular variants of the Mahalanobis distance described in Chang and Chiang (2002): the Normalized-RXD (N-RXD) and the Uniform Target Detector (UTD).

The N-RXD takes the form:

$$\Lambda_{N-RXD} = \frac{(\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})^T}{\|\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}\|} \hat{\Sigma}_{SCM}^{-1} \frac{(\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})}{\|\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}\|} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda, \quad (2.66)$$

where  $\|\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}\|^2 = (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})^T (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})$  stands for the Euclidean norm of the vector. The detection test in Eq. (2.66) can be immediately identified as the normalized version of  $\Lambda_{\text{KellyAD}}$ . In addition,  $\Lambda_{N-RXD}$  takes also the form of a matched filter specified in Eq. (2.65) with matched signal  $\mathbf{H}^T = (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})^T \hat{\Sigma}_{SCM}^{-1}$  the same as in Eq. (2.60) and a different scale constant  $c = \|\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}\|^{-2}$ .

The UTD is another widespread anomaly detection test. It was firstly introduced in Harsanyi (1993) and can be defined as:

$$\Lambda_{UTD} = (\mathbf{1} - \hat{\boldsymbol{\mu}}_{SMV})^T \hat{\Sigma}_{SCM}^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda. \quad (2.67)$$

with  $\mathbf{1} = [1, \dots, 1]^T$  is the  $m$ -dimensional unity vector. Once again the detector in Eq. (2.67) can be interpreted as a matched filter where  $\mathbf{H}^T = (\mathbf{1} - \hat{\boldsymbol{\mu}}_{SMV})^T \hat{\Sigma}_{SCM}^{-1}$  is the matched signal. If there is no a priori information about the target spectra, the non-prior approach is the one that does not introduce any information into the detector and consists on assuming uniform distribution for the spectra over all the bands.

### 2.3.4 Generalized Kelly Anomaly Detector

In the previous detection schemes, it has not been taken into account in the derivation of the test that both the mean vector  $\boldsymbol{\mu}$  and the covariance matrix  $\boldsymbol{\Sigma}$  were unknown. One simply replaced the mean vector by a plug-in estimate in the detector (two-step GLRT). In case both covariance matrix and mean vector are unknown, we need to derive a new detector. This strategy is similar to the one proposed in Section 2.2.2 for the generalized Kelly detection test. The likelihood functions for the two competing hypothesis are described by Eq. (1.24) and their maxima are achieved when taking the same results detailed in Section 2.2.2 for both the covariance matrix (Eq. (2.10)) and the mean vector (Eq. (2.12)). Then, the LR can be written according to Eq. (2.43) and it remains to maximize it w.r.t.  $\mathbf{p}$  which is equivalent to minimize Eq. (2.20). The minima is obtained by taking:

$$\hat{\mathbf{p}} = \frac{N+1}{N} \frac{(\mathbf{x} - \boldsymbol{\mu}_0)}{\alpha}. \quad (2.68)$$

Hence, the resulting detector takes the form:

$$\Lambda_{G-KellyAD} = (\mathbf{x} - \hat{\boldsymbol{\mu}}_0)^H \hat{\mathbf{S}}_0^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_0) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda, \quad (2.69)$$

where  $\hat{\mathbf{S}}_0 = \sum_{i=1}^N (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_0)(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_0)^H$ , and  $\hat{\boldsymbol{\mu}}_0 = \frac{1}{N+1} \left( \mathbf{x} + \sum_{i=1}^N \mathbf{x}_i \right)$ . Once again the mean vector estimate  $\boldsymbol{\mu}_0$  and the covariance matrix  $\mathbf{S}_0$  depend on the data under test  $\mathbf{x}$ . Hence,  $\mathbf{x} - \hat{\boldsymbol{\mu}}_0$  and  $\hat{\mathbf{S}}_0$  are not independent. Remark that one can write  $(\mathbf{x} - \hat{\boldsymbol{\mu}}_0) = \frac{N}{N+1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})$ . Neglecting the multiplicative constants, the test in Eq. (2.69) appears to be equivalent to the classical  $\Lambda_{RXD}$  obtained throughout a different approach but built with  $N+1$  available data.

### 2.3.5 Simulations

In this section, we validate the theoretical analysis on simulated data. Firstly, we validate through Monte-Carlo simulations the distribution of  $\Lambda_{KellyAD}$  detailed above. The experiments have been conducted on Gaussian vectors of dimension  $m = 5$  and for different values of  $N$ . The computations have been made through  $10^6$  Monte-Carlo trials. The true covariance is chosen as a Toeplitz matrix whose entries are  $\Sigma_{i,j} = \rho^{|i-j|}$  and where  $\rho = 0.4$ . The mean vector is arbitrarily set to have all entries equal to 3. Exceedance plot shows the fraction of points in the data set whose Mahalanobis distance is larger than the indicated value. This is essentially a cumulative histogram of Mahalanobis distance values which correspond to the ‘‘PFA-threshold’’ relationship. Remark that the definition of false alarms is not unique and it depends on the application. Thus, we will rather refer to the distribution of the detector in target absent hypothesis.

Fig. 2.6 illustrates the distribution of the detector under null hypothesis. The case

where both the covariance matrix and the mean vector are perfectly known corresponds to the  $\chi^2$ -distribution and the adaptive versions of the quadratic form become a Hotelling  $T^2$ . The perfect agreement between the the green and yellow curves bears out the results of Proposition 2.3.1.

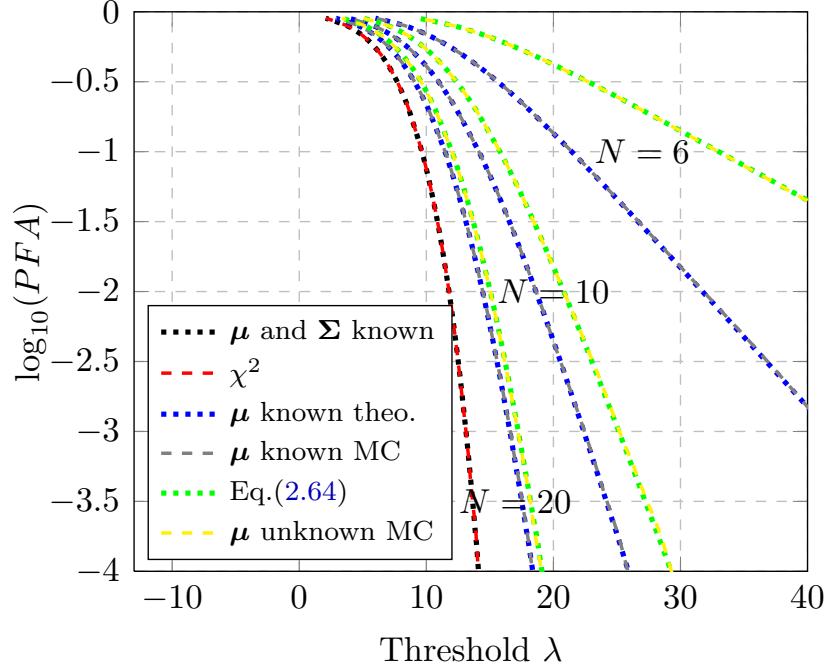


Figure 2.6: PFA versus threshold for the  $\Lambda_{KellyAD}$  when (1)  $\mu$  and  $\Sigma$  are known (Mahalanobis) (red and black curves) (2) only  $\mu$  is known (gray and blue curves) (3) Proposition 2.3.1: both  $\mu$  and  $\Sigma$  are unknown (yellow and green curves).

Furthermore, we compare in Fig. 2.7 the five proposed anomaly detectors in terms of probability of detection. The experiments were realized on Gaussian vectors of dimension  $m = 5$ , for  $N = 10$  and the artificial targets signature used for the simulations is the unity vector  $\mathbf{p} = [1, \dots, 1]^T$ . The SNR is obtained according to:  $SNR = |\alpha^2| \mathbf{p}^H \Sigma^{-1} \mathbf{p}$ . On a first step, the threshold is determined to ensure a  $PFA = 10^{-3}$  obtained empirically from the data. The best results are obtained for the Mahalanobis-based detectors, i.e. the classical RXD, Kelly AD and generalized Kelly AD. The two detectors derived according to Kelly's approach perform fundamentally the same and slightly better than the RXD. This improvement may be due to the fact that  $N + 1$  data are available for the Kelly's strategies, while only  $N$  samples are used in the classical RXD. The matched filter based detectors deliver poor performance. In the case of the UTD, as the matching signal is the unity vector, the detector is shown to be not optimal even in the case the artificial targets signature used for the simulations is the unity vector. The N-RXD presents an almost flat curve. The normalization factor grows as the SNR gets higher. The outcome of this detector can be assimilated to the residual background level and its use in Gaussian environment

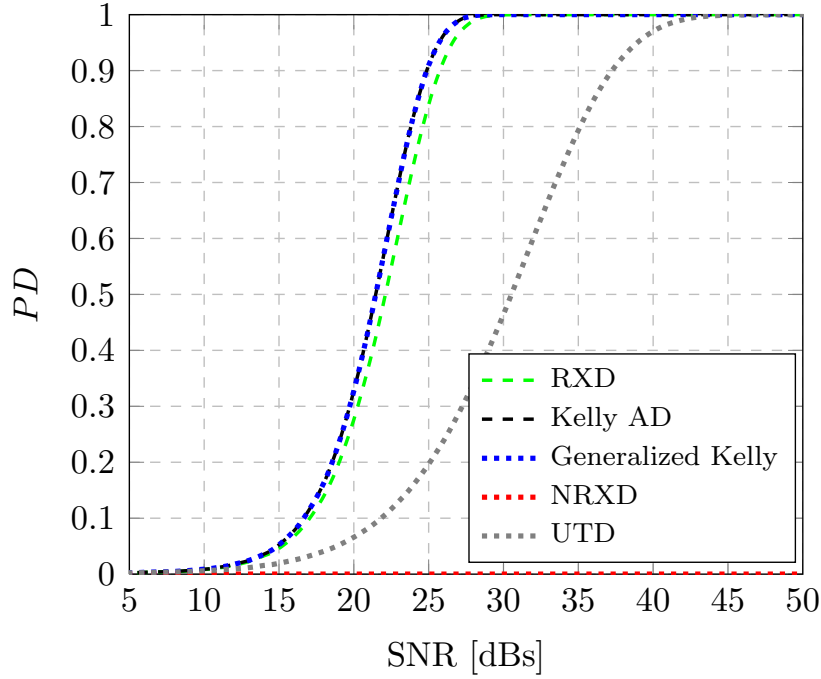


Figure 2.7: Probability of detection for different SNR values and  $PFA = 10^{-3}$  in Gaussian case.

should be avoided.

## 2.4 Summary

Four adaptive detection schemes, the AMF, Kelly detectors with a “plug-in” and a generalized versions as well as the ANMF, have been analyzed in the case where both the covariance matrix and the mean vector are unknown and need to be estimated. In this context, theoretical closed-form expressions for false-alarm regulation have been derived under Gaussian assumptions for the SCM-SMV estimates for three detection schemes. The resulting “PFA-threshold” expressions highlight the CFARness of these detectors since they only depend on the size and the number of data, but not on the unknown parameters. The theoretical analysis has been validated through Monte Carlo simulations and the performances of the detectors have been compared in terms of probability of detection.

Finally, the last section of this chapter is dedicated to anomaly detection. The classical RXD test is explored and compared to other four detectors. The different advantages and drawbacks for the different detection schemes are commented. Furthermore, the comparison is performed through Monte Carlo simulations in Gaussian context.

One difficulty for the background detection statistic is to assume a tractable model or at least to account for robustness to deviation from the assumed theoretical model in the detection scheme. Since Gaussian assumption is not always fulfilled for real hyperspectral data, alternative robust estimation techniques are proposed in next chapter. However, it is essential to notice that the derivations for many results in robust detection contexts strongly rely on the results obtained in the Gaussian context. For instance, this is the case of [Pascal et al. \(2006\)](#) in which the derivation of a robust detector distribution is based on its Gaussian counterpart.



# 3

## Adaptive Target Detection in non-Gaussian Background

### 3.1 Introduction

In the previous chapter, a complete analysis for detection problems has been derived under Gaussian assumption. However, it has been shown in [Manolakis and Marden \(2002\)](#), [Frontera-Pons et al. \(2012a\)](#) that the Gaussian distribution is not always a good model for background characterization in HSI (see Fig. 3.1). To fill this gap, one of the most general and acknowledged models for background statistics characterization is the family of Elliptically Contoured Distributions, originally introduced by Kelker in [Kelker \(1970\)](#) and extended to the complex case in [Krishnaiah and Lin \(1986\)](#), [Ollila et al. \(2012\)](#). They account for non-Gaussianity by providing a long tailed alternative to the multivariate Normal model. They are proven to represent a more accurate characterization of HSI data than models based on the Gaussian assumption [Theiler et al. \(2010\)](#), [Niu et al. \(2010\)](#).

Although elliptical distributions have already been introduced for background modeling in HSI [Manolakis and Marden \(2002\)](#), the parameters estimation is often performed using classical Gaussian based estimators; for example the covariance matrix, generally determined by the SCM and the mean vector with the SMV. As mentioned previously in Section 1.4, these classical estimators correspond to the Maximum Likelihood Estimators (MLEs) for Gaussian assumption. However, they lead to sub-optimal detection schemes when the noise is a non-Gaussian process. When working on the elliptical distributions framework, the model can be used to assess the robustness of statistical procedures and to derive alternative robust estimators of the parameters: the mean vector and the covariance matrix [Gini and Greco \(2002\)](#).



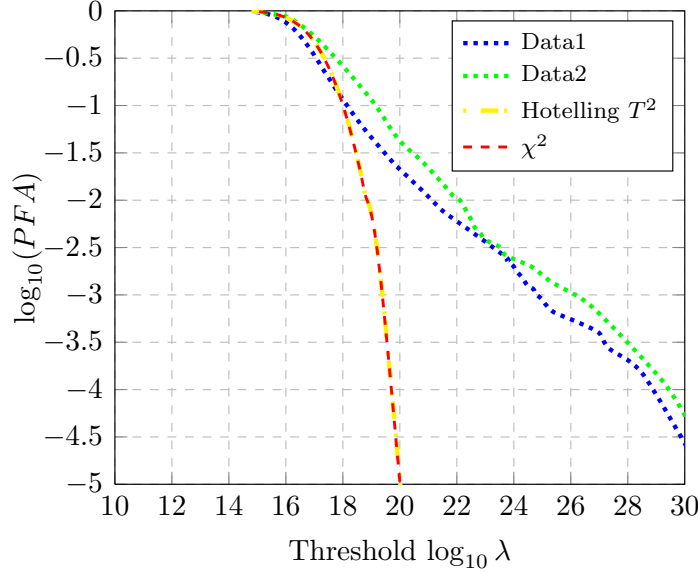


Figure 3.1: Exceedance plots are essentially a cumulative histogram of Mahalanobis distance values, and they provide a useful way to visualize the fat tails of multivariate distributions. This figure shows the theoretical results when Gaussian distribution is assumed (yellow and red curves) and experimental hyperspectral data distribution which diverts from the expected behavior for Hyperion NASA image.

In statistics, one is generally interested on the scatter matrix which corresponds to a covariance structure estimate that is proportional to the covariance matrix and on the location estimate related to the mean vector. Robust location and scatter  $M$ -estimates were firstly introduced as a generalization of the MLEs. Up until now, they have been extensively studied in statistics literature [Huber \(1964\)](#), [Maronna \(1976\)](#), [Bilodeau and Brenner \(1999\)](#) and have been used in several signal processing applications, such as radar detection [Ollila et al. \(2012\)](#), [Mahot et al. \(2012\)](#), [Couillet et al. \(2012\)](#). When the underlying distribution is unknown,  $M$ -estimators provide an alternative approach for robust parameter estimation of elliptical populations. These can then be used as plug-in estimators (two-step GLRT) in place of the unknown mean vector and of the covariance matrix. This is a simple but often efficient method to obtain robust properties for signal processors derived under the Gaussian assumption. We note that the use of robust signal processors is of fundamental interest due to the occurrence of impulsive environments and outliers in real hyperspectral images.

The first part of this chapter reviews the main characteristics of the family of Elliptically Contoured Distributions and provides some examples of distributions. In Section 3.3 the corresponding robust  $M$ -estimators and their statistical properties are detailed. On Section 3.4, we describe the different target detection techniques analyzed. Section 3.5 presents the anomaly detectors in non-Gaussian environment. Finally, Section 3.6 summarizes this work.

## 3.2 Elliptically Contoured Distributions

The family of elliptical distributions is highly appreciated in signal processing, for its high flexibility and the fact that it can represent a multitude of particular distributions. In many areas, when it turned out that the Gaussian model did not perfectly fit the data, it is quite natural that it has turned to this family which included also the Gaussian case. Thus, fairly extensive studies have been conducted on them in many areas dealing with random data. These include the results obtained in finance-related applications (e.g. [Eberlein and Keller \(1995\)](#), [Frahm \(2004\)](#), [Frahm et al. \(2005\)](#)). In hyperspectral applications, elliptical distributions have been proposed for data modeling in [Manolakis and Marden \(2002\)](#). Many other authors such [Theiler et al. \(2010\)](#), [Frontera-Pons et al. \(2012b\)](#) have acknowledged this class of distributions trying to promote their use. Thus, this large family deserves some attention since it allows to represent all kinds of distributions slight or heavy tail, circular or not as well as the compound Gaussian distributions. It provides a multivariate location-scale class of distributions that primarily serve as long tailed alternatives to the multivariate Normal model. Let us now present the class of complex elliptically contoured distributions [Krishnaiah and Lin \(1986\)](#), [Ollila et al. \(2012\)](#), [Frahm \(2004\)](#).

### 3.2.1 Definition

**Definition 3.2.1.** *A  $m$ -dimensional random complex vector  $\mathbf{z}$  has a complex elliptical (CE) distribution if its characteristic function is of the form:*

$$\Phi_{\mathbf{z}}(\mathbf{c}) = \exp(j \Re(\mathbf{c}^H \boldsymbol{\mu})) \phi(\mathbf{c}^H \boldsymbol{\Sigma} \mathbf{c}), \quad (3.1)$$

for some function  $\phi : \mathbb{R}^+ \rightarrow \mathbb{R}$ , called characteristic generator, a positive semidefinite matrix  $\boldsymbol{\Sigma}$ , called scatter matrix and  $\boldsymbol{\mu} \in \mathbb{C}^m$  the location vector. We shall write  $\mathbf{z} \sim \mathcal{CE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ .

Remark that the previous definition in [3.2.1](#) permits singular CE distributions as the scatter matrix  $\boldsymbol{\Sigma}$  could be singular ( $\text{rank}(\boldsymbol{\Sigma}) \leq m$ ). In addition, the scatter matrix and the characteristic generator do not uniquely define a particular CE distribution and a scale constraint, either on  $\boldsymbol{\Sigma}$  either on  $\phi$ , needs to be imposed for identifiability purposes.

The family of  $\mathcal{CE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$  is closed under affine transformations for any given  $\phi$ . Let  $\mathbf{z} \sim \mathcal{CE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ , then:

$$\mathbf{Bz} + \mathbf{b} \sim \mathcal{CE}(\mathbf{B}\boldsymbol{\mu} + \mathbf{b}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^H, \phi), \quad (3.2)$$

for  $\mathbf{B} \in \mathbb{C}^{k \times m}$  and  $\mathbf{b} \in \mathbb{C}^k$ . Hence, every affinely transformed and particularly every linearly combined elliptical random vector is elliptical, too. An interesting fact is that the generating variates of affinely transformed random vectors always remain invariant. Thus, affinely transformed random vectors not only are elliptical but even closed under the corresponding location-scale family, i.e. the corresponding transformed random vector is of the same type that  $\mathbf{z}$ .

Therefore, a random vector  $\mathbf{x} \sim \mathcal{CE}(\mathbf{0}, \mathbf{I}_m, \phi)$  is spherically distributed since  $\Phi_X(\mathbf{c}) = \phi(\mathbf{c}^H \mathbf{c})$ ; and every affine transformation of a spherical random vector has an elliptical distribution. The converse is true, according to the following theorem, when the transformation matrix has full rank.

**Theorem 3.2.1.** (Stochastic Representation Theorem)

A  $m$ -dimensional random vector  $\mathbf{z} \sim \mathcal{CE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$  with  $\text{rank}(\boldsymbol{\Sigma}) = k \leq m$  if and only if it admits the following representation:

$$\mathbf{z} \stackrel{d}{=} \boldsymbol{\mu} + \mathcal{R} \mathbf{A} \mathcal{U}^{(k)}, \quad (3.3)$$

where  $\mathcal{U}^{(k)}$  is a  $k$ -dimensional random vector uniformly distributed on the unit complex  $k$ -sphere  $\mathbb{CS}^k$ ;  $\mathcal{R}$  is a non-negative random variable called generating variate, being stochastically independent of  $\mathcal{U}^{(k)}$ ;  $\boldsymbol{\mu} \in \mathbb{C}^m$  and  $\boldsymbol{\Sigma} = \mathbf{A} \mathbf{A}^H$  is a factorization of  $\boldsymbol{\Sigma}$  where  $\mathbf{A} \in \mathbb{C}^{m \times k}$  with  $\text{rank}(\mathbf{A}) = k$ .

The generating variate  $\mathcal{R}$  determines the distribution's shape, in particular the tail of the distribution. Indeed, the generating variate can be connected to the characteristic generator via its CDF as detailed in [Frahm \(2004\)](#). Hence, an  $m$ -dimensional random vector  $\mathbf{x}$  is spherically distributed if and only if  $\mathbf{x} \stackrel{d}{=} \mathbf{O} \mathbf{x}$  for all unitary matrix  $\mathbf{O}$ .

The stochastic representation provides a simple manner to simulate elliptically distributed random vectors. The uniform spherical distribution can be easily obtained from a complex Normal distributed random vector,  $\mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_m)$ , when dividing it by its length,  $\mathcal{U}^{(k)} \stackrel{d}{=} \mathbf{y} / \|\mathbf{y}\|_2$ . Then, the transformation matrix  $\mathbf{A}$  on the vector  $\mathcal{U}^{(k)}$  produces elliptically contoured density surfaces. Thereupon, some knowledge on the CDF of  $\mathcal{R}$  is required to completely determine the shape of the distribution and include the location vector  $\boldsymbol{\mu}$ . Remark that the dispersion on the elliptical distribution is uniquely determined by  $\boldsymbol{\Sigma}$ , and the particular factorization  $\mathbf{A}$  adds no information.

From  $\mathbf{z} \sim \mathcal{CE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ , it does not follow that  $\mathbf{z}$  has a PDF  $f_{\mathbf{z}}(\cdot)$ . If it exists, it can be related to the density function of the generating variate  $\mathcal{R}$ , provided  $\mathcal{R}$  is absolutely continuous. Then the PDF of  $\mathbf{z}$  has the form:

$$f_{\mathbf{z}}(\mathbf{z}) = c_{m,h} |\boldsymbol{\Sigma}|^{-1} h((\mathbf{z} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \boldsymbol{\mu})) \quad (3.4)$$

where  $h$  is a function such as (3.4) defines a PDF in  $\mathbb{C}^m$ . The function  $h$  is usually called *density generator* and it is assumed to be only approximately known. It satisfies:

$$\gamma_{m,h} = \int_0^\infty t^{m-1} h(t) dt < \infty, \quad (3.5)$$

to ensure the integrability of  $f_{\mathbf{z}}(\cdot)$ ,  $c_{m,h}$  is a normalizing constant which compels that  $f_{\mathbf{z}}(\cdot)$  integrates to 1 and is given by:

$$c_{m,h} = 2(s_m \gamma_{m,h})^{-1}, \quad (3.6)$$

$$s_m = \frac{2\pi^m}{\Gamma(m)}, \quad (3.7)$$

where  $s_m$  corresponds to the surface area of the complex hypersphere ( $m$ -sphere) in  $\mathbb{C}S^m$ . The constant  $c_{m,h}$  could be included into the function  $h$ , but with this notation  $h$  can be independent of the dimension  $m$ . In some occasions for clarity, we write  $h_m$  to stress that  $h$  refers to the density generator of a random vector of dimension  $m$ . In the absolutely continuous case we shall write  $\mathcal{CE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h_m)$  instead of  $\mathcal{CE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ . Remark that the PDF (3.4) depends on  $\mathbf{z}$  only through the quadratic form  $(\mathbf{z} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \boldsymbol{\mu})$ . Thus, the level sets of the density  $f_{\mathbf{z}}(\mathbf{z})$  are ellipsoids in the complex Euclidean  $m$ -space.

The scatter matrix  $\boldsymbol{\Sigma}$  describes the shape and orientation of the elliptical equidensity contours. If the second-order moment exists, then  $\boldsymbol{\Sigma}$  reflects the structure of the covariance matrix  $\mathbf{M}$ , i.e. the covariance matrix is equal to the scatter matrix up to a scalar constant  $\boldsymbol{\Sigma} = k \mathbf{M}$ . Nevertheless, we can always find an appropriate normalization constraint such that  $\text{cov}(\mathbf{z}) = \boldsymbol{\Sigma}$ . This constraint consists in taking  $\mathbb{E}[\mathcal{R}^2] = \text{rank}(\boldsymbol{\Sigma})$ . Note that while the scatter matrix is always defined up to a scalar constant, the covariance matrix does not exist for some CE distributions (e.g. Cauchy distribution). We are interested on the information contained on the structure of the matrix, but not on its scale. In HSI application, this matrix serves to characterize the correlation pattern existing within the spectral bands.

The class of elliptical distributions includes a large number of well-known distributions, as for instance the multivariate Gaussian [Goodman \(1963\)](#), the  $K$ -distribution [Conte et al. \(1991\)](#) or the multivariate  $t$ -distribution [Krishnaiah and Lin \(1986\)](#).

Let us now comment on an important subclass of the elliptical distributions, the Spherically Invariant Random Vectors (SIRV).

### 3.2.2 The Class of Spherically Invariant Random Vectors

A SIRV is a complex compound Gaussian process with random power [Yao \(1973\)](#), [Rangaswamy et al. \(1993\)](#). They represent an important subclass of CE distributions widely used in signal processing applications, e.g. wireless radio propagation problems [Yao et al. \(2004\)](#), radar clutter echoes modeling [Gini \(1997\)](#), hyperspectral background characterization [Manolakis and Marden \(2002\)](#), [Ovarlez et al. \(2011\)](#). The main idea behind compound Gaussian distributions applied to signal processing is to improve the modelization of the background by adding the assumption that the background, locally Gaussian, presents spatially variable power. This intends to represent the information *texture* brought by the high resolution (or any other treatment improvement). A random vector  $\mathbf{z}$  has a compound Gaussian distribution if it can be written as:

$$\mathbf{z} \stackrel{d}{=} \sqrt{\tau} \mathbf{x} + \boldsymbol{\mu}, \quad (3.8)$$

where  $\tau$  is a positive random variable called *texture* and  $\mathbf{x}$  an  $m$ -dimensional independent zero-mean complex Gaussian vector  $CN(\mathbf{0}, \boldsymbol{\Sigma})$  called *speckle*. Given that

$\text{rank}(\mathbf{\Sigma}) = m$ , the PDF of the SIRV always exists and can be expressed as:

$$f_{\mathbf{z}}(\mathbf{z}) = \frac{1}{\pi^m |\mathbf{\Sigma}|} \int_0^{+\infty} \frac{1}{\tau^m} \exp\left(-\frac{(\mathbf{z} - \boldsymbol{\mu})^H \mathbf{\Sigma}^{-1} (\mathbf{z} - \boldsymbol{\mu})}{\tau}\right) f_{\tau}(\tau) d\tau, \quad (3.9)$$

where  $f_{\tau}(\cdot)$  is the texture PDF. The different choice of distribution for  $f_{\tau}(\cdot)$  lead to some well-known examples of SIRV distributions. The complex Normal distribution is obtained when  $\tau$  is Dirac-distributed, i.e.  $f_{\tau}(\tau) = \delta(\tau - \tau_0)$ . If  $\tau$  is Gamma-distributed, the resulting SIRV is a  $K$ -distribution. Its PDF presents long tails depending on a shape parameter  $\nu$  such that, for small values of  $\nu$  the distribution is heavy tailed and, when  $\nu$  tends to infinity, the resulting distribution is the multivariate Normal [Jakeman \(1980\)](#). This will be further detailed in the next subsection. However,  $f_{\tau}(\tau)$  cannot always be written on a closed form expression (e.g. Weibull distribution).

It is worth pointing out that, as the SIRV are a subclass of CE distributions, they admit the stochastic representation of the *Theorem 3.2.1*. Therefore:

$$\mathbf{z} \stackrel{d}{=} \boldsymbol{\mu} + \sqrt{\tau} \mathbf{A} \mathbf{x}, \quad (3.10)$$

where  $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$  and  $\mathbf{\Sigma} = \mathbf{A} \mathbf{A}^H$  any factorization of  $\mathbf{\Sigma}$  with  $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{\Sigma}) = k$ .

SIRV distributions allow to extend several interesting properties from the well-known multivariate Normal distribution. In the case of CE distributions, any marginal distribution of a random vector, whose distribution belongs to a specific elliptical family, also belongs to the same family only if the elliptical distribution belongs to the class of SIRV distributions [Kano \(1994\)](#). Although many of the commonly used CE distributions belong to the class of the SIRV distributions, there are many which do not (e.g. multivariate generalized Gaussian distribution [Pascal et al. \(2013a\)](#)). It is worth pointing out that SIRV distributions constitute a subclass of CE distributions but it does not correspond to the subclass of spherically distributed random vectors discussed above. Thus, they should not be mistaken with spherically distributions.

### 3.2.3 Examples of Complex Elliptical Distributions

Let us now discuss in detail some examples of CE distributions providing their properties such as their PDF, the distribution of the texture variable  $\tau$  in case of SIRV distributions. We consider in this section the absolutely continuous case so that the PDF  $f_{\mathbf{z}}(\cdot)$  exists and  $\text{rank}(\mathbf{\Sigma}) = m$ .

#### 3.2.3.1 Multivariate Normal distribution

Let  $\mathbf{z}$  be an  $m$ -dimensional random vector, such as  $\mathbf{z}$  follows a complex Normal distribution with mean vector  $\boldsymbol{\mu} \in \mathbb{C}^m$  and scatter matrix  $\mathbf{\Sigma}$ . Then, the PDF is of the form detailed in (1.4). Thus, according to the CE framework, the complex Gaussian distribution is obtained with the density generator  $h(t) = \exp(-t)$ , where  $s$  refers

to the quadratic form  $t = (\mathbf{z} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \boldsymbol{\mu})$  and  $c_{m,h} = \pi^{-m}$ . We shall write  $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  for complex Normal model.

### 3.2.3.2 Multivariate $t$ -distribution

A complex random vector  $\mathbf{z}$  of dimension  $m$  has a multivariate  $t$ -distribution with  $\nu > 0$  degrees of freedom if it has the SIRV representation (3.10). Thus, the texture  $\tau \stackrel{d}{=} \nu/x$  with  $x \sim \chi_\nu^2 = \text{Gam}(\nu/2, 2/\nu)$ , i.e. the inverse distribution of the texture is gamma distributed with shape  $\nu/2$  and scale  $2/\nu$ . The density generator is given by:

$$h(t) = \left(1 + \frac{2t}{\nu}\right)^{-\frac{2m+\nu}{2}}, \quad (3.11)$$

and the normalizing constant:

$$c_{m,h} = \frac{2^m \Gamma\left(\frac{2m+\nu}{2}\right)}{(\pi\nu)^m \Gamma\left(\frac{\nu}{2}\right)}. \quad (3.12)$$

The multivariate  $t$ -distribution allows for tails heavier than the Gaussian model. Hence,  $\nu = 1$  depicts the complex Cauchy distribution and when  $\nu$  tends to infinity, the Gaussian distribution is obtained. The covariance matrix exists for  $\nu > 2$ , and in that case,  $\mathbf{M} = \frac{\nu}{\nu-2} \boldsymbol{\Sigma}$ . If we assume finite 2nd-order moments, one can model  $\tau \stackrel{d}{=} (\nu-2)/x$  with  $x \sim \chi_\nu^2$  in which case  $\mathbb{E}[\tau] = 1$  and  $\mathbf{M} = \boldsymbol{\Sigma}$ . This constraint is assumed for identifiability considerations. We shall write  $\mathcal{C}t_\nu(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  to denote this distribution. Complex  $t$ -distribution has been used e.g. for modeling radar clutter [Younsi et al. \(2009\)](#) and is the most widely spread distribution for hyperspectral imaging [Manolakis et al. \(2005\)](#), [Niu et al. \(2010\)](#).

### 3.2.3.3 $K$ -distribution

A complex random vector  $\mathbf{z}$  is said to have a  $m$ -variate  $K$ -distribution with shape  $\nu > 0$  if it has the SIRV representation (3.10) where  $\tau \sim \text{Gam}(\nu, 1/\nu)$ . Thus, one has  $\mathbb{E}[\tau] = 1$  and its PDF takes the form:

$$f_\tau(\tau; \nu) = \frac{\nu^\nu}{\Gamma(\nu)} \tau^{\nu-1} e^{-\nu\tau}. \quad (3.13)$$

And replacing (3.13) into (3.9) yields:

$$h(t) = t^{(\nu-m)/2} K_{\nu-m}(2\sqrt{\nu t}), \quad (3.14)$$

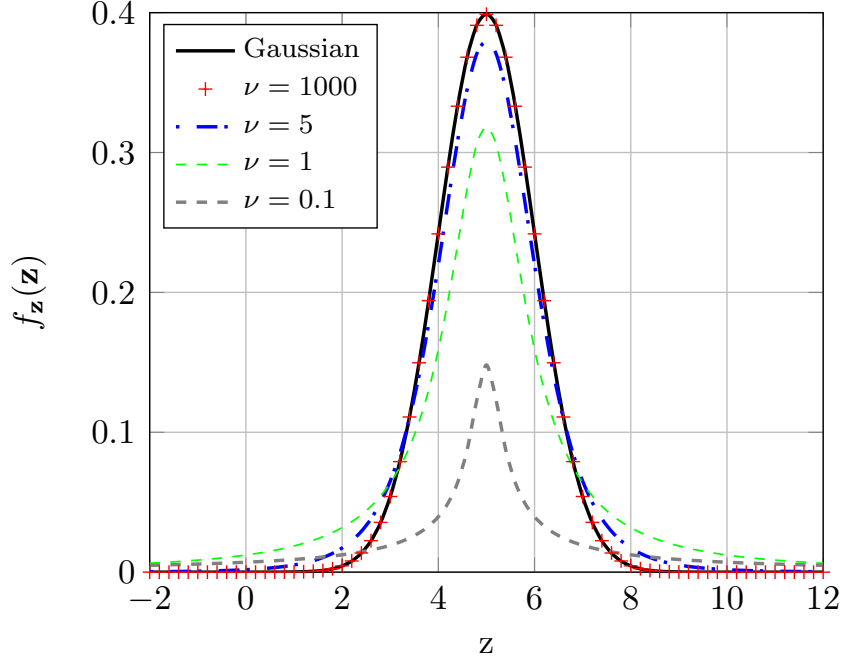


Figure 3.2: Comparison between probability density functions of Gaussian distribution and  $t$ -distribution with different values of  $\nu$ .

as the density generator and  $K_\alpha(\cdot)$  denotes the modified Bessel function of the second kind of order  $\alpha$ . And the normalizing constant is written:

$$c_{m,h} = \frac{2\nu^{(\nu+m)/2}}{\Gamma(\nu)\pi^m}. \quad (3.15)$$

In the limit  $\nu \rightarrow \infty$ , the  $K$ -distribution reduces to the multivariate Normal distribution. The  $K$ -distribution is widely used in radar signal processing (see for e.g. [Gini \(1997\)](#)) and it will be used for simulation in non-Gaussian environments in [Section 3.3.4](#) and [Section 3.4.2](#).

### 3.3 Robust Estimation Procedures

In order to improve the parameter estimation, the objective is to find an appropriate model and to use the corresponding MLEs. This method leads to asymptotically efficient estimators but not necessarily robust. Indeed, the robust estimator is rather one that is still fairly reliable, regardless of the data departure, failing to be optimal in some scenarios. In a real life applications, although elliptical distributions offer a great deal of possible distributions, the risk that the data do not follow the model considered still remains. Thus, the models used always correspond to simplifications of the reality. The fact that a slight deviation between reality and the model assumed has little or no influence on the parameter estimates, is precisely the robustness of the

estimator. Let us now recap the optimal estimators (MLE) in elliptical distributions framework.

### 3.3.1 Maximum Likelihood Estimates

Assume  $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N$  an IID sample from a  $\mathcal{CE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h_m)$  defined with a perfectly known function  $h_m$  and with  $N > m$ . The log-likelihood function for both  $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  can be written as:

$$\Lambda(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = c - N \log |\boldsymbol{\Sigma}| + \sum_{i=1}^N \log (h_m ((\mathbf{z}_i - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{z}_i - \boldsymbol{\mu}))) . \quad (3.16)$$

By applying matrix derivatives with respect to  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  and supposing  $h_m$  real, continuous and differentiable, one obtains:

$$\sum_{i=1}^N u(t_i) \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{z}_i - \hat{\boldsymbol{\mu}}) = \mathbf{0} \quad (3.17)$$

$$\sum_{i=1}^N u(t_i) \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{z}_i - \hat{\boldsymbol{\mu}}) (\mathbf{z}_i - \hat{\boldsymbol{\mu}})^H \hat{\boldsymbol{\Sigma}}^{-1} - N \hat{\boldsymbol{\Sigma}} = \mathbf{0} \quad (3.18)$$

where  $t_i = (\mathbf{z}_i - \hat{\boldsymbol{\mu}})^H \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{z}_i - \hat{\boldsymbol{\mu}})$  and  $u(t) = -h'_m(t)/h(t)$  is a weight function depending on the density generator, and  $h'_m$  is the derivative with respect to  $t = (\mathbf{z} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \boldsymbol{\mu})$ . Thus, the MLE are obtained as fixed point solutions to the following equations :

$$\hat{\boldsymbol{\mu}}_{ML} = \frac{\sum_{i=1}^N u(t_i) \mathbf{z}_i}{\sum_{i=1}^N u(t_i)}, \quad (3.19)$$

$$\hat{\boldsymbol{\Sigma}}_{ML} = \frac{1}{N} \sum_{i=1}^N u(t_i) (\mathbf{z}_i - \hat{\boldsymbol{\mu}}) (\mathbf{z}_i - \hat{\boldsymbol{\mu}})^H \quad (3.20)$$

and an iterative algorithm is required to simultaneously solve both estimating equations. If there is information about the underlying  $CE$ , notably  $h_m$  is perfectly known, then MLEs provide the optimal approach to the parameters estimation problem.

For the Gaussian distribution (i.e.  $h(t) = \exp(-t)$ ), we have  $u(t) = 1$ , which corresponds to the SMV in (1.5) and the SCM in (1.6) as the unique MLE of  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  respectively. In the general case, the function  $u(t)$  is not necessarily a constant. If we consider the  $K$ -distribution detailed in (3.14), one can easily obtain the MLE when taking:  $u(t) = \sqrt{\frac{\nu}{t}} \frac{K_{\nu-m-1}(2\sqrt{\nu t})}{K_{\nu-m}(2\sqrt{\nu t})}$  as weight function.

The MLE for the multivariate  $t$ -distribution will be presented in the next section. Indeed, these estimators belong to the wider class of robust  $M$ -estimators.



### 3.3.2 $M$ -Estimators

A family of estimators, the  $M$ -estimators have been introduced in this context as a generalization of maximum likelihood estimates. They allow to take into account some error in the model. When the density generator  $h_m$  is unknown,  $M$ -estimators provide an alternative approach for robust parameter estimation of elliptical populations. Assume  $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N$  an IID sample from a  $\mathcal{CE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h_m)$  with  $N > m$ . The complex  $M$ -estimators of location and scatter are defined as the joint solutions of:

$$\hat{\boldsymbol{\mu}}_N = \frac{\sum_{i=1}^N u_1(t_i) \mathbf{z}_i}{\sum_{i=1}^N u_1(t_i)}, \quad (3.21)$$

$$\hat{\boldsymbol{\Sigma}}_N = \frac{1}{N} \sum_{i=1}^N u_2(t_i^2) (\mathbf{z}_i - \hat{\boldsymbol{\mu}}) (\mathbf{z}_i - \hat{\boldsymbol{\mu}})^H, \quad (3.22)$$

where  $t_i = \left( (\mathbf{z}_i - \hat{\boldsymbol{\mu}})^H \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{z}_i - \hat{\boldsymbol{\mu}}) \right)^{1/2}$  and  $u_1(\cdot), u_2(\cdot)$  denote any real-valued *weighting functions* on the quadratic form  $t_i$  that satisfy a set of general assumptions detailed below. Remark that  $t_i$  here is nothing but the Mahalanobis distance and the main purpose of  $u_1(\cdot)$  and  $u_2(\cdot)$  is to attenuate high contributions of this distance (outliers rejection).

We have seen that within the framework of elliptical distributions, MLEs verify  $u(t) = -h'_m(t)/h(t)$  for a known density generator. The idea behind  $M$ -estimators is to replace this function  $u(t)$  by another function with particular properties in order to obtain a robust estimator on the whole family of elliptical distributions. The choice of  $u_1(\cdot)$  and  $u_2(\cdot)$  does not need to be related to a particular elliptical distribution and therefore,  $M$ -estimators constitute a wide class of estimates that includes the MLEs for the particular case  $u_1(t) = -h'_m(t)/h_m(t)$  and  $u_2(t^2) = u_1(t)$ .

The  $M$ -estimators were analyzed firstly in the real case. Existence and uniqueness have been proven provided functions  $u_1(\cdot), u_2(\cdot)$  satisfy a set of general assumptions stated by Maronna [Maronna \(1976\)](#). Olilla has shown in [Ollila and Koivunen \(2003a\)](#) that these conditions hold also in the complex case for the scale-only problem. They are recalled in the following:

- $u_1$  and  $u_2$  are non-negative, non increasing, and continuous on  $[0, \infty)$ .
- Define  $\psi_i(t) = tu_i(t), i = 1, 2$ . Then,  $\psi_1$  and  $\psi_2$  are bounded. Let  $K_i = \sup_{t \geq 0} \psi_i(t)$ .
- $\psi_2$  is nondecreasing and is strictly increasing in the interval where  $\psi_2 < K_2$ .
- Let  $P_N(\cdot)$  the empirical distributions of  $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N$ . There exists  $a > 0$  such that for every hyperplane  $H$ ,  $\dim(H) \leq 2m - 1$ ,  $P_N(H) \leq 1 - \frac{m}{K_2} - a$ .

This last condition can be greatly relaxed [Tyler \(1988\)](#), [Kent and Tyler \(1991\)](#). Moreover, it relates the samples  $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N$  with the weight function  $u_2$  and it implies that  $K_2 > m$  for  $a = 0$ .

The solutions  $(\hat{\boldsymbol{\mu}}_N, \hat{\boldsymbol{\Sigma}}_N)$  are estimates for the parameters  $(\boldsymbol{\mu}, \boldsymbol{\Sigma}_0)$  where:

$$\boldsymbol{\Sigma}_0 = \mathbb{E} \left[ u_2 \left[ (\mathbf{z} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}_0^{-1} (\mathbf{z} - \boldsymbol{\mu}) \right] (\mathbf{z} - \boldsymbol{\mu}) (\mathbf{z} - \boldsymbol{\mu})^H \right]. \quad (3.23)$$

For elliptical distributions, the implicit equation (3.23) admits a solution  $\boldsymbol{\Sigma}_0$  and one has:

$$\boldsymbol{\Sigma} = \sigma \boldsymbol{\Sigma}_0 = \frac{1}{c} \text{cov}(\mathbf{z}). \quad (3.24)$$

Hence,  $\sigma$  is obtained solving the following equation given in [Tyler \(1982\)](#) and recapped here. Multiplying (3.23) by  $\boldsymbol{\Sigma}_0^{-1}$  and taking trace yields:

$$\mathbb{E} [\psi_2(\sigma |\mathbf{t}|^2)] = m. \quad (3.25)$$

and  $\mathbf{t} \sim \mathcal{CE}(\mathbf{0}_m, \mathbf{I}_m)$ . This expectation can be evaluated as a simple integral where the density of  $t = |\mathbf{t}|$  can be written as:

$$f(t) = s_m t^{2m-1} h(t^2). \quad (3.26)$$

**Remark 3.3.1.** *A M-estimator of scatter (respectively location) can be interpreted as an implicit weighted covariance matrix (respectively mean vector) estimator. In general, a robust weighting function should decrease to zero. This means that small weights are given to those observations that are highly outlying in terms of the quadratic form  $t_i$ . Observations which stay far-off from the background model will have a smaller contribution to the parameters estimation. For example, the SCM gives unit weight ( $u_2(t) = u_1(t) = 1$ ) to all observations, and hence is naturally non-robust. When dealing with heavy tailed background models, the use of robust estimates decreases the impact of highly impulsive samples and possible outliers in the reference cells.*

### 3.3.2.1 Examples of M-estimators

We present here some examples of  $M$ -estimators of location and scatter.

**3.3.2.1.1 Huber type M-estimators** The *Huber's*  $M$ -estimators proposed in [Huber \(1964\)](#) are defined in the complex case by [Ollila and Koivunen \(2003b\)](#) when taking in (3.21) and (3.22) the following weighting functions:

$$u_1(t) = \min(t, k) \quad (3.27)$$

$$u_2(t^2) = \frac{1}{\beta} \min(t^2, k^2) \quad (3.28)$$

Thus, *Huber's*  $M$ -estimators can be obtained as the solutions of the following implicit equations:

$$\hat{\boldsymbol{\mu}}_{Hub} = \frac{1}{N} \sum_{i=1}^N \left[ \mathbf{z}_i \mathbb{1}_{t_i \leq k} + k \frac{\mathbf{z}_i}{\left( (\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{Hub})^H \hat{\mathbf{M}}_{Hub}^{-1} (\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{Hub}) \right)^{1/2}} \mathbb{1}_{t_i > k} \right], \quad (3.29)$$

$$\begin{aligned} \hat{\mathbf{M}}_{Hub} = \\ \frac{1}{N\beta} \sum_{i=1}^N \left[ (\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{Hub}) (\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{Hub})^H \mathbb{1}_{t_i^2 \leq k^2} + k^2 \frac{(\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{Hub})(\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{Hub})^H}{(\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{Hub})^H \hat{\mathbf{M}}_{Hub}^{-1} (\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{Hub})} \mathbb{1}_{t_i^2 > k^2} \right], \end{aligned} \quad (3.30)$$

where  $t_i^2$  is the quadratic form  $(\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{Hub})^H \hat{\boldsymbol{\Sigma}}_{Hub}^{-1} (\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{Hub})$  and  $\mathbb{1}(\cdot)$  is the indicator function defined as  $\mathbb{1}_A = 1$  if  $A$  and  $\mathbb{1}_A = 0$  otherwise.

The constants  $k$  and  $\beta$  are adjustable parameters that allow to choose the percentage of attenuated data (quadratic form higher than  $t > k$ ) and the coefficient of proportionality between the limit of the estimator and the scatter matrix of the considered elliptical distribution ( $\sigma$  in (3.25)). The constant  $k$  is defined so that:

$$q = F_{\chi_{2m}^2} (2k^2) \quad (3.31)$$

for a chosen value of  $q$  ( $0 < q \leq 1$ ), where  $F_{\chi_{2m}^2}$  denotes the CDF of the  $\chi^2$ -distribution with  $2m$  degrees of freedom. And the scaling parameter  $\beta$  serves to make  $\hat{\mathbf{M}}_{Hub}$  an asymptotically unbiased estimate of the covariance matrix in the Gaussian case, i.e.  $\text{cov}(\mathbf{z}) = \boldsymbol{\Sigma}$ , which yields:

$$\beta = F_{\chi_{2m}^2} (2k^2) + k^2 \frac{1-q}{m} \quad (3.32)$$

Hence,  $\mathbb{E}[u_2(t)t] = m$  is valid when  $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$ . In this case, the quadratic form  $(\mathbf{z} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \boldsymbol{\mu})$  follows a  $(1/2)\chi_{2m}^2$  and  $\sigma = 1$ . The choice  $q = 1$  gives  $u_1 = 1$  and  $u_2 = 1$  and the *Huber* estimators correspond to the classical SMV and SCM respectively, whereas the choice  $q = 0$  yields  $u_1(t) = t^{-1}$  and  $u_2(t^2) = mt^{-2}$  and the *Huber* estimators correspond to Fixed Point Estimators to be defined below. For intermediate value of  $q$ , the *Huber*-estimates might be interpreted as a mixture between the Fixed Point and the classical SCM. The values of the quadratic form below  $k^2$  are kept and treated as in the SMV and SCM estimates (corresponding to the first summation); and the values of  $t_i^2$  higher than  $k^2$  are attenuated by the weighting function similarly to the Fixed Point approach. In a complex Gaussian context, it can be shown that when  $N$  tends to infinity, the proportion of data processed with the SCM is equal to the parameter  $q$ . Note that the conditions specified above are satisfied and  $K_2 = k^2/\beta > m$ .

**3.3.2.1.2  $t$ -distribution  $M$ -estimators** The  $t$ -distribution or Student  $M$ -estimator corresponds to the MLE for the complex  $t$ -distributions presented above. The weighting function takes the form:

$$u(t) = \frac{\nu + 2m}{\nu + 2t} \quad (3.33)$$

and  $u_1(t) = u_2(t^2)$ . Note that the resulting estimator depend on the particular shape of the  $t$ -distribution through the parameter  $\nu$ . If the underlying distribution is perfectly known, the value of the shape parameter  $\nu$  is available and one can use it in Eq. (3.19) and Eq. (3.20) leading to the MLEs for the  $t$ -distribution. If there is no reliable information on the real value of  $\nu$ , we rather choose a small value to obtain an estimate robust to spiky distributions. Remark that taking  $\nu = 1$  leads to the most pessimistic case corresponding to Cauchy MLE.

**3.3.2.1.3 Fixed Point Estimators** The Fixed Point Estimators (FPE) firstly introduced in Tyler (1987), satisfy the following implicit equations:

$$\hat{\boldsymbol{\mu}}_{FP} = \frac{\sum_{i=1}^N \frac{\mathbf{z}_i}{\left( (\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{FP})^H \hat{\mathbf{M}}_{FP}^{-1} (\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{FP}) \right)^{1/2}}}{\sum_{i=1}^N \frac{1}{\left( (\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{FP})^H \hat{\mathbf{M}}_{FP}^{-1} (\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{FP}) \right)^{1/2}}}, \quad (3.34)$$

$$\hat{\mathbf{M}}_{FP} = \frac{m}{N} \sum_{i=1}^N \frac{(\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{FP}) (\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{FP})^H}{\left( (\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{FP})^H \hat{\mathbf{M}}_{FP}^{-1} (\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{FP}) \right)}, \quad (3.35)$$

which are the particular cases of (3.21) and (3.22) for  $u_1(t) = t^{-1}$  and  $u_2(t^2) = m t^{-2}$ . They are specified by fixed point equations and can be easily computed using a recursive algorithm. If the limit of the algorithm exists, it must be a solution. Although, the theoretical convergence of the procedure has not been proven, the empirical behavior is suitable. The major difficulty is found when there is at least one sample belonging to the secondary data which has the same value as the mean, or some samples with values close to the mean. In that case, the algorithm will probably diverge. Huber  $M$ -estimator described above could be used instead, as it overcomes this possible algorithm divergence.

For the scale-only problem, the covariance matrix FPE has been widely investigated in statistics and signal processing literature Pascal et al. (2008a), Gini and Greco (2002), Soloveychik and Wiesel (2014). We refer to Pascal et al. (2008b) for a detailed performance analysis.

The main results on the statistical properties of  $\hat{\mathbf{M}}_{FP}$  are recalled for elliptical distribution framework (and  $\boldsymbol{\mu}$  assumed to be known):

- $\hat{\mathbf{M}}_{FP}$  is a consistent and unbiased estimate of  $\boldsymbol{\Sigma}$ ;

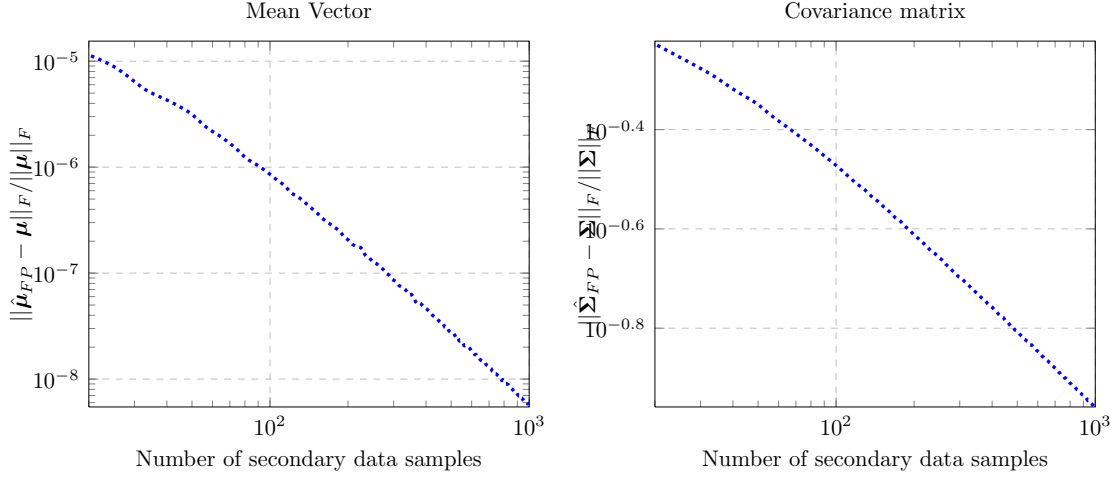


Figure 3.3: Empirical consistency of the FPE for the joint estimation problem.

- its asymptotic distribution is Gaussian and its covariance matrix is fully characterized in [Pascal et al. \(2005\)](#);
- for  $N$  sufficiently large  $\hat{\mathbf{M}}_{FP}$  behaves as a Wishart matrix with  $\frac{m}{m+1}N$  degrees of freedom.

The asymptotic behavior for the whole family of  $M$ -estimators is further discussed below. Remark that the distribution of  $\hat{\mathbf{M}}_{FP}$  does not depend on the specific elliptical distribution. In order to establish consistency and asymptotic normality, the population distribution cannot be too heavily concentrated around the center. Consistency and asymptotic distribution of  $\hat{\mathbf{M}}_{FP}$  are demonstrated for the joint location-scatter estimation in the real case in [Tyler \(1987\)](#). For identification purposes, one may define a normalization constraint on the matrix estimate, e.g.  $\text{Tr}(\hat{\mathbf{M}}_{FP}) = m$ . In the SIRV framework, the covariance matrix FPE has been obtained according to the MLE approach in [Gini and Greco \(2002\)](#). Hence, considering the texture  $\tau$  as a deterministic and unknown parameter, the scatter matrix estimate in (3.35) is the true MLE but it corresponds to the approximate MLE when  $\tau$  is a random variable.

The problem of both location and scatter estimation leads to consistent estimates. Although we do not discuss here the proof for consistency, the empirical behavior claims for consistent assumption. The empirical consistency has been analyzed over  $K$ -distributed simulated data of dimension  $m = 10$  and shape parameter  $\nu = 0.1$ . On Fig. 3.3, the  $x$ -axis represents the number of  $N$  secondary data beginning at  $2m$ ; and on the  $y$ -axis the relative error of the estimator  $\|\hat{\boldsymbol{\mu}}_{FP} - \boldsymbol{\mu}\|_F / \|\boldsymbol{\mu}\|_F$  and  $\|\hat{\boldsymbol{\Sigma}}_{FP} - \boldsymbol{\Sigma}\|_F / \|\boldsymbol{\Sigma}\|_F$ .

Although FPE have the general form of  $M$ -estimators, it is a limit case as they do not satisfy the conditions defined by Maronna. Indeed, their weighting function  $u_i$  is not defined in 0. Similarly, the SMV-SCM are not  $M$ -estimator in the sense of

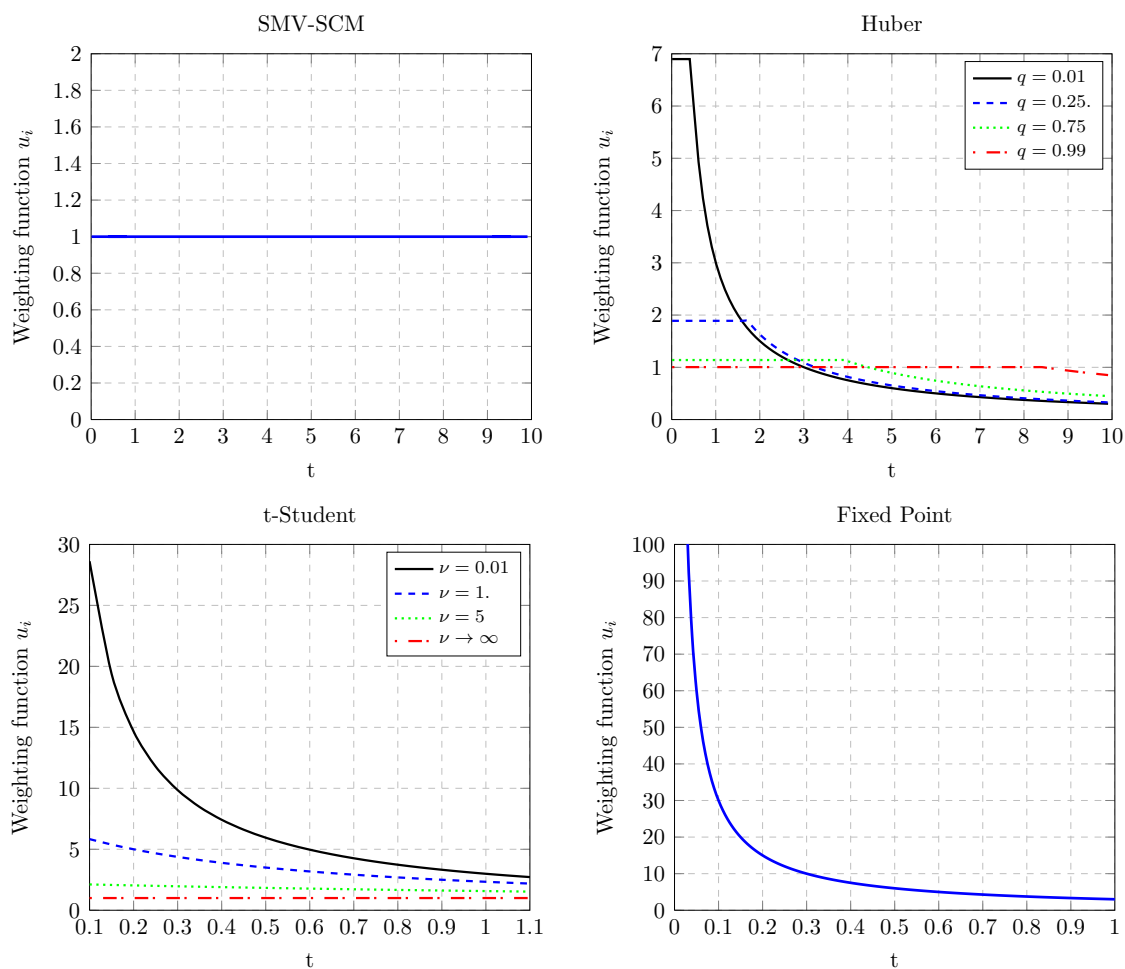


Figure 3.4: Comparison of weighting functions for the SMV-SCM, Huber,  $t$ -distribution and FPE with  $m = 3$ .

Maronna as the upper limit of the function  $\psi_i = tu_i(t)$  is infinite. Figure 3.4 compares the weighting functions of the four estimators detailed above. Note that the SCM and the FPE are the limit cases of *Huber* and *t*-distribution estimators.

### 3.3.2.2 Asymptotic distribution of the $M$ -estimators

Let us specify the asymptotic distribution of the  $M$ -estimators firstly in the real case. Assume  $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N$  an IID sample from a  $\mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h_m)$ . We shall denote  $\hat{\boldsymbol{\Sigma}}_M$  and  $\hat{\boldsymbol{\mu}}_M$  the real  $M$ -estimator that verify the equations (3.21) and (3.22) respectively and the conditions from Maronna [Maronna \(1976\)](#) detailed above,  $h$  is decreasing and  $\mathbb{E}[\psi'_1(\sigma^{1/2}x)] > 0$ . We denote  $\boldsymbol{\Sigma}_M$  the solution of equation (3.23). Then,

$$\sqrt{N} \left( \hat{\boldsymbol{\Sigma}}_M - \boldsymbol{\Sigma}_M, \hat{\boldsymbol{\mu}}_M - \boldsymbol{\mu} \right) \xrightarrow{d} (\mathbf{N}, \mathbf{n}), \quad (3.36)$$

where  $\mathbf{N}$  and  $\mathbf{n}$  are independent and:

$$\mathbf{n} \sim \mathcal{N}(\mathbf{0}, (\alpha/\beta^2)\boldsymbol{\Sigma}_M), \quad (3.37)$$

$$\mathbf{N} \sim \mathcal{N}(\mathbf{0}, \sigma_1 (\mathbf{I} + \mathbf{K}_m)(\boldsymbol{\Sigma}_M \otimes \boldsymbol{\Sigma}_M) + \sigma_2 \text{vec}(\boldsymbol{\Sigma}_M)\text{vec}(\boldsymbol{\Sigma}_M)^T). \quad (3.38)$$

with  $\sigma$  being the solution of  $\mathbb{E}[\psi_2(\sigma t^2)] = m$ , where:

$$\alpha = m^{-1} \mathbb{E}[\psi_1^2(\sigma^{1/2}t)], \quad (3.39)$$

$$\beta = \mathbb{E}[(1 - m^{-1})u_1(\sigma^{1/2}t) + m^{-1}\psi'_1(\sigma^{1/2}t)], \quad (3.40)$$

$$\sigma_1 = a_1 (m + 2)^2 (2a_2 + m)^{-2}, \quad (3.41)$$

$$\sigma_2 = a_2^{-2} \{ (a_1 - 1) - 2a_1 (a_2 - 1) [m + (m + 4)a_2] (2a_2 + m)^{-2} \}, \quad (3.42)$$

and:

$$a_1 = [m(m + 2)]^{-1} \mathbb{E}[\psi_2^2(\sigma t^2)], \quad (3.43)$$

$$a_2 = m^{-1} \mathbb{E}[\sigma t^2 \psi'_2(\sigma t^2)]. \quad (3.44)$$

These results are due to [Maronna \(1976\)](#), but Tyler in [Tyler \(1982\)](#) obtained the asymptotic variance parameters  $\sigma_1$  and  $\sigma_2$ . Moreover, this results have been extended for the covariance matrix estimators to the complex case in [Ollila et al. \(2012\)](#), [Mahot et al. \(2013\)](#). Assume  $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N$  an IID sample from a  $\mathcal{CE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h_m)$ . We shall denote  $\hat{\boldsymbol{\Sigma}}_N$  and  $\hat{\boldsymbol{\mu}}_N$  the complex  $M$ -estimator. We denote  $\boldsymbol{\Sigma}_0$  the solution of equation (3.23). Then, one has:

$$\sqrt{N}(\hat{\boldsymbol{\Sigma}}_N - \boldsymbol{\Sigma}_0) \xrightarrow{d} \mathcal{CN}(\mathbf{0}, \nu_1 (\boldsymbol{\Sigma}_0^T \otimes \boldsymbol{\Sigma}_0) + \nu_2 \text{vec}(\boldsymbol{\Sigma}_0)\text{vec}(\boldsymbol{\Sigma}_0)^H), \quad (3.45)$$

with:

$$\nu_1 = \frac{\mathbb{E}[\psi_2^2(\sigma t^2)]}{m(m + 1)(1 + [m(m + 1)]^{-1} \mathbb{E}[\sigma t^2 \psi'_2(\sigma t^2)])^2}, \quad (3.46)$$

$$\nu_2 = \frac{\mathbb{E}[(\psi_2(\sigma t) - m\sigma)^2]}{(m + \mathbb{E}[\sigma t^2 \psi'_2(\sigma t^2)])^2} - \frac{\nu_1}{m}, \quad (3.47)$$

where  $\nu_1 > 0$  and  $\nu_2 \geq -\frac{\nu_1}{m}$  and  $\sigma$  solves Eq. (3.25).

Remark that the classical SCM verifies the previous conditions under Gaussian assumption taking  $\nu_1 = 1$  and  $\nu_2 = 0$ .

The problems analyzed in this work stand for continuous and differentiable functions  $H$ , which are invariant with respect to the scale of the matrix (homogeneity of degree 0), i.e. for any positive definite Hermitian matrix  $\Sigma$  and any positive scalar  $c$  they satisfy  $H(\Sigma) = H(c\Sigma)$ . It means that for any homogeneous of degree 0 function  $H(\cdot)$ :

$$H(\Sigma) = H(c\Sigma) = H(\mathbf{M}) = H(c\mathbf{M}), \quad (3.48)$$

Moreover, one can define the processor as  $H(\mathbf{V})$ , where  $\mathbf{V}$  denotes the shape matrix, defined as  $\mathbf{V} = m\Sigma/\text{Tr}(\Sigma)$ . Thus, let us now recall an important property of  $M$ -estimators. Let  $H(\mathbf{V})$  be a function on the set of complex positive definite Hermitian  $m \times m$  matrices that satisfies  $H(\mathbf{V}) = H(c\mathbf{V})$  for any positive scalar  $c$ . Moreover, let us assume that all the partial derivatives are continuous. Then, one has:

$$\sqrt{N} \left( H(\hat{\Sigma}) - H(\Sigma) \right) \xrightarrow{d} \mathcal{CN}(\mathbf{0}, \vartheta_1 H'(\Sigma)(\Sigma^T \otimes \Sigma)H'(\Sigma)^H), \quad (3.49)$$

where  $H'(\Sigma) = \frac{dH(\Sigma)}{d\text{vec}(\Sigma)}$ .

In addition, this property applied to Wishart matrices, i.e. the SCM under Gaussian distribution and more generally the  $M$ -estimators in elliptical distribution framework, Tyler's theorem allows to show that  $\sqrt{N} \left( H(\hat{\Sigma}_{SCM}) - H(\Sigma) \right)$  and  $\sqrt{\sigma_1 N} \left( H(\hat{\Sigma}_N) - H(\Sigma) \right)$  share the same distribution. In practice,  $H(\cdot)$  corresponds to a function that given a covariance matrix associates a parameter. For instance, in target detection, the parameter of interest is the outcome of the detector. Generally, it could be used with any function for which multiplying the covariance matrix by an scalar does not change the result.

### 3.3.3 Robust shrinkage estimators

We address here high dimensional covariance estimation for elliptical distributed samples. HSI target detection methods usually involve the estimation of high-dimension covariance matrix and mean vector. We present shrinkage methods that are suitable for high dimensional problems with small number of samples (large  $m$  small  $N$ ). In these “large  $m$  small  $N$ ” problems, classical estimators suffer from a distorted eigenstructure and improved estimators are required. A common regularization approach has been widely studied, the diagonally loaded approach introduced in Abramovich (1981), Carlson (1988). Thus, in Gaussian context the regularized SCM takes the form:

$$\hat{\mathbf{M}}_{SCM-DL}(\beta) = \frac{1-\beta}{N} \sum_{i=1}^N (\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{SMV})(\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{SMV})^H + \beta \mathbf{I}_m. \quad (3.50)$$



In presence of non-Gaussian, impulsive background the estimate in Eq. (3.50) suffers from the same drawbacks than the SCM and the class of robust estimates are more appropriate. Yet, the class of  $M$ -estimators described above exhibit important shortcomings in high dimensional context and they cannot be computed for the undersampling case when  $m > N$ . This section deals with shrinkage FPE.

### 3.3.3.1 Shrinkage Fixed Point Estimates

We extend here covariance matrix FPE to the high dimensional setting using shrinkage regularization. Many estimation methods were proposed in the literature for the shrinkage covariance matrix FPE. An original approach that introduces a normalization constraint in the algorithm for the shrinkage FPE is found in [Chen et al. \(2011b\)](#). Moreover, in [Abramovich and Besson \(2013b,a\)](#), [Besson and Abramovich \(2013\)](#), this estimator has been used within the Expected Likelihood framework. The optimization of the shrinkage parameter  $\beta$  has been discussed in [Couillet and McKay \(2014\)](#).

Let us consider now the shrinkage FPE introduced in [Pascal et al. \(2013b\)](#) and defined as the solution of the following fixed point equation:

$$\hat{\mathbf{M}}_{FP}(\beta) = (1 - \beta) \frac{m}{N} \sum_{i=1}^N \frac{(\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{FP})(\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{FP})^H}{((\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{FP})^H \hat{\mathbf{M}}_{FP}^{-1}(\beta) (\mathbf{z}_i - \hat{\boldsymbol{\mu}}_{FP}))} + \beta \mathbf{I}_m, \quad (3.51)$$

for  $\beta \in (0, 1]$  and  $\hat{\boldsymbol{\mu}}_{FP}$  the same as given in Eq. (3.34).

The mean vector has to be included in the construction of the shrinkage FPE. One solution would be to develop a joint algorithm that regularizes also the mean vector to a chosen steering vector. However, this method would entail a decentralization of the samples bringing them closer to the chosen direction. If the intended direction is perfectly known, it would correspond to the real value of the mean parameter and it would not be necessary to estimate it. Therefore, this regularization strategy does not make sense. We propose to use the mean FPE given in Eq. (3.34). The choice of the mean vector estimate is not likely to have a significant impact on the the shrinkage covariance estimate. We select a robust solution as the mean vector plays a key role in the target detection schemes analyzed in this work. On the other hand, notice that the mean vector estimate depends on the regularization parameter  $\beta$  through the covariance matrix estimator. This fact can lead to unexpected behavior for the shrinkage FPE when mean vector is unknown and has to be jointly estimated. Statistical properties of these estimates should be further investigated for the joint estimation problem.

Thus, it was shown in [Pascal et al. \(2013b\)](#), when mean vector is known, that Eq. (3.51) admits a solution  $\mathbf{M}$  if and only if  $\beta \in (\beta', 1]$  where  $\beta' := \max(0, 1 - N/m)$ . Moreover,  $\mathbf{M}$  satisfies:

$$\text{Tr}(\mathbf{M}^{-1}) = m. \quad (3.52)$$

Although this constraint on the trace was originally proven for a known mean vector, it is easy to shown that it holds for the non-central case. When  $\beta$  tends to 0, the proposed shrinkage FPE in Eq. (3.51) tends intuitively to the conventional FPE in Eq. (3.35) whose inverse has its trace equal to  $m$ . Note that the normalization constraint on the FPE in this case is different to the one proposed in Section 3.3.2.1 for identification purposes ( $\text{Tr}(\hat{\mathbf{M}}_{FP}) = m$ ). The basis of the proposed method are the Fixed Point Estimators. However, the approach presented here could be extended to other  $M$ -estimators.

### 3.3.4 Simulations

This section is devoted to the different examples of  $M$ -estimators detailed above comparing the estimation performances for elliptical distribution framework. We compare here the performance of the different  $M$ -estimators for different scenarios and show the advantages of the  $M$ -estimators from both robustness and performance loss in non-optimal cases. The error is measured according to the criterion  $C_1$ , namely the Normalized Mean Square Error (NMSE) resulting from:

$$\text{NMSE}(\hat{\boldsymbol{\mu}}) = \frac{\mathbb{E}[\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|_F^2]}{\|\boldsymbol{\mu}\|_F^2} \quad (3.53)$$

$$\text{NMSE}(\hat{\boldsymbol{\Sigma}}) = \frac{\mathbb{E}[\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_F^2]}{\|\boldsymbol{\Sigma}\|_F^2}. \quad (3.54)$$

For the criteria  $C_1$ , we consider the overall error, i.e.  $C_1 = \text{NMSE}(\hat{\boldsymbol{\mu}}) + \text{NMSE}(\hat{\boldsymbol{\Sigma}})$ . The expected error is approximated by 200 Monte-Carlo trials.

#### 3.3.4.1 Gaussian analysis

The experiments have been conducted on Gaussian vectors for different dimensions  $m$  and for different number of secondary data  $N$ . The true covariance is chosen as a Toeplitz matrix whose entries are  $\Sigma_{i,j} = \rho^{|i-j|}$  and for different values of  $\rho$  and the mean vector is set to have all entries equal to  $(3 + 4j)$ .

Fig. 3.5 depicts the evolution of the error with the number of secondary data. In this case,  $m = 10$  and the true covariance matrix is a Toeplitz matrix with  $\rho = 0.4$ . Notice that the error decreases as the number of secondary data increases. The estimators that provide better performance in terms of NMSE are the SMV-SCM as expected, as they are the MLE for Gaussian case. Note that the error is slightly higher and the performances slightly worse for the considered  $M$ -estimators. One can observe that robust  $M$ -estimators represent a good trade-off between efficiency and robustness which is a general result widely known in estimation theory. The most robust estimators are the FPE, then *Huber's*  $M$ -estimators and Student  $M$ -estimators, and the classical SMV-SCM are the less robust solutions. The order in robustness correspond to the inverse order for efficiency. Finally, note that *Huber's*  $M$ -estimators and Student  $M$ -estimators perform similarly in Gaussian environment.

Fig. 3.6 illustrates the NMSE as a function of the dimension  $m$ . In this experiments

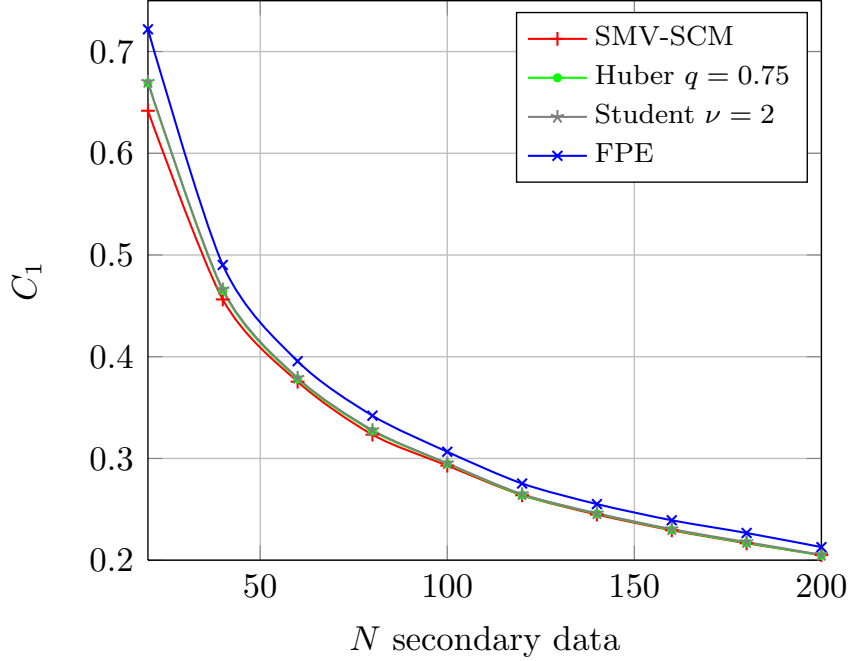


Figure 3.5: Performance comparison for different estimators for samples drawn from Gaussian distribution for  $m = 10$  and  $\rho = 0.4$ .

we have  $N = 200$  and the true covariance matrix is a Toeplitz matrix with  $\rho = 0.4$ . It is interesting to notice that for a fixed value  $N$ , when the dimension  $m$  grows the error increases. Fig. 3.6 reveals that the estimation is deteriorated for high values of  $m$  and it is essential to take into account this high-dimensionality in estimation problems. For this reason, we have presented Shrinkage estimators (brown, purple and orange curves). The results are shown for three values of the parameter  $\beta$ ,  $\beta = 0.2$ ,  $\beta = 0.5$  and  $\beta = 0.8$ . The choice of this value will impact on the resulting NMSE and the evolution of the error with  $m$  will change for different values of the  $\beta$ . The optimal  $\beta$  depends on the underlying true covariance matrix and theoretical methods for optimizing over  $\beta$  are detailed in [Couillet and McKay \(2014\)](#), [Chen et al. \(2011b\)](#). Note that the error remains almost the same for the Shrinkage FPE with  $\beta = 0.8$  among the different values of  $m$ . This enforces the fact that Shrinkage FPE are more suitable in high-dimensional problems. Finally, the main comments on the previous figure concerning robustness and efficiency are still valid.

We analyze in Fig. 3.7 the NMSE for different covariance matrices. In this experiments  $m = 10$  and  $N = 200$ . The coefficient  $\rho$  is related to the covariance correlation. For  $\rho = 0.4$  the same errors than in the previous figures are found. It is worth pointing out that the error drops when the covariance matrices structure becomes more correlated (high values of  $\rho$ ). Thereupon,  $M$ -estimators are robust for badly conditioned matrices outperforming the SMV-SCM estimators, even in Gaussian case.

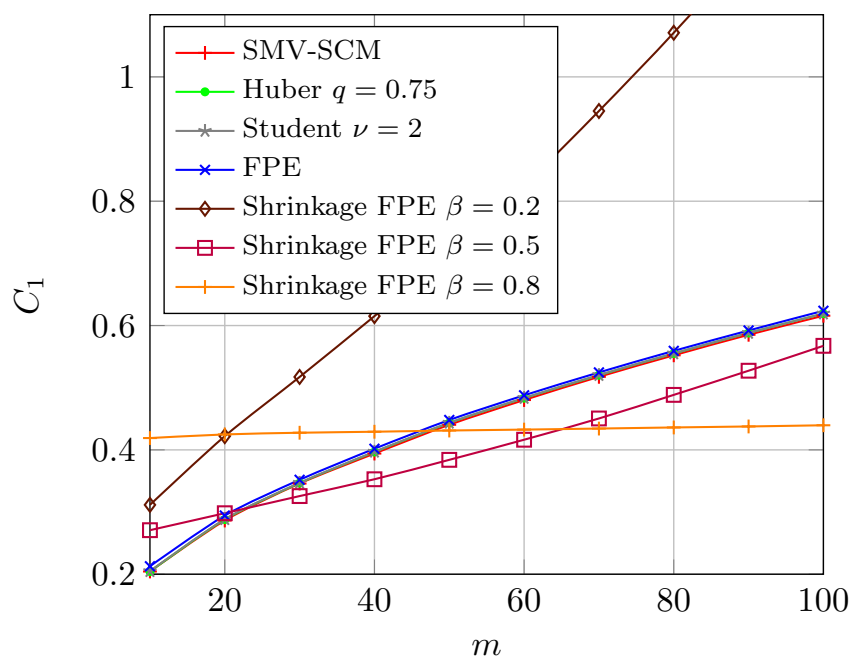


Figure 3.6: Performance comparison for different estimators for samples drawn from Gaussian distribution for  $N = 200$  and  $\rho = 0.4$ .

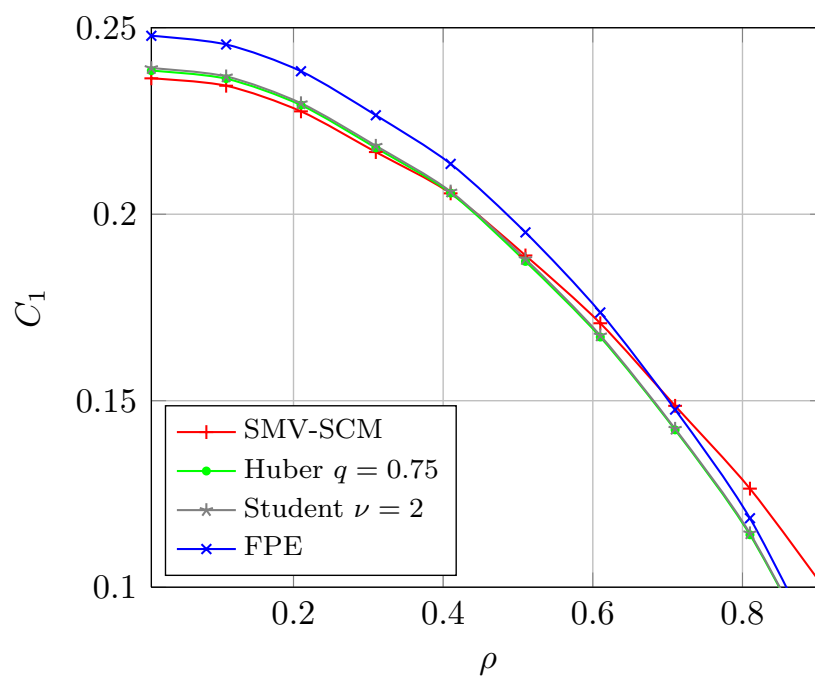


Figure 3.7: Performance comparison for different estimators for samples drawn from Gaussian distribution for  $m = 10$  and  $N = 200$ .

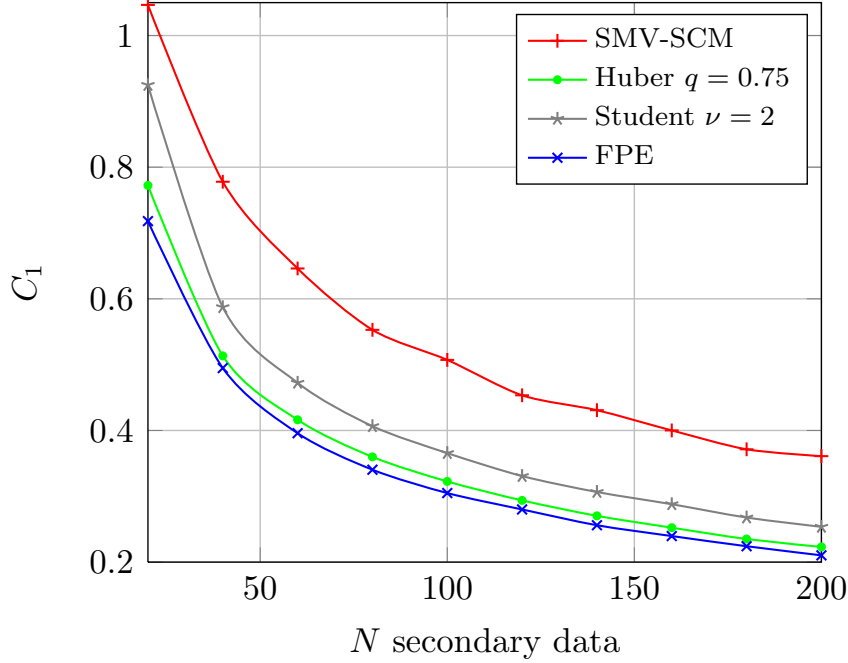


Figure 3.8: Performance comparison for different estimators for samples drawn from  $K$ -distribution with shape parameter  $\nu = 0.5$  for  $m = 10$  and  $\rho = 0.4$ .

### 3.3.4.2 Non-Gaussian analysis

This section compares the different estimators in non-Gaussian environment. The experiments have been conducted on a  $K$ -distribution with shape parameter  $\nu = 0.5$  (see Section 3.2.3.3) for different dimensions  $m$  and for different number of secondary data  $N$ . The true covariance is chosen as a Toeplitz matrix whose entries are  $\Sigma_{i,j} = \rho^{|i-j|}$  and for different values of  $\rho$  and the mean vector is set to have all entries equal to  $(3 + 4j)$ .

Fig. 3.8 is the analogous to Fig. 3.5 and presents the NMSE variation with the number of secondary data. The NMSE decreases as the number of secondary data increases. The estimators providing the best performance are the FPE. It is worth pointing out that the results with the FPE are the same under Gaussian distribution and under  $K$ -distribution, i.e. their performance do not depend on the real distribution (distribution-free). *Huber's*  $M$ -estimators perform slightly worse than the FPE, but better than Student  $M$ -estimators. *Huber's*  $M$ -estimators can be interpreted as a mixture between the FPE and the SMV-SCM. Thus, they deliver very good results in both Gaussian and non-Gaussian environments. Notice that the error obtained with classical SMV-SCM has considerably increased under non-Gaussian assumption.

Fig. 3.9 details the NMSE for different values of the dimension  $m$ . It is important to notice that the error is considerably higher for the classical SMV-SCM estimates. The FPE performance drop for high values of  $m$  while *Huber's*  $M$ -estimators hold the smaller error in terms of NMSE. This can be explained through the robustness analysis of the  $M$ -estimators studied in Mahot (2012). On the other hand, the Shrinkage FPE

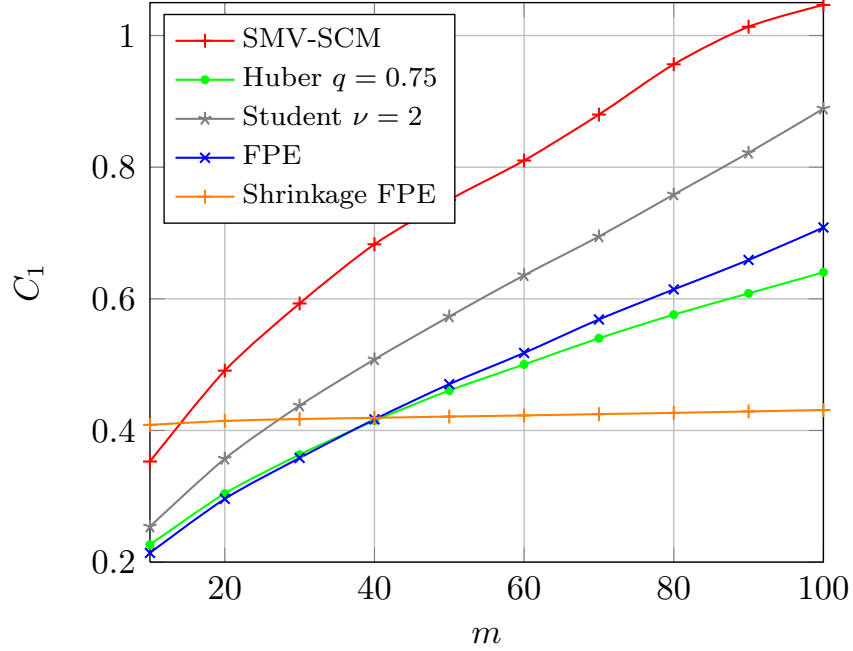


Figure 3.9: Performance comparison for different estimators for samples drawn from  $K$ -distribution with shape parameter  $\nu = 0.5$  for  $N = 200$  and  $\rho = 0.4$ .

presents almost the same error through the different values of  $m$ . The regularization parameter  $\beta$  has been fixed here to  $\beta = 0.8$ . Notice that the results with the Shrinkage FPE are the same than in Gaussian case (see Fig. 3.6). Thus, the Shrinkage FPE is also distribution-free.

Fig. 3.10 depicts the NMSE as a function of the true covariance matrix. Remark that the FPE present a smaller value for the NMSE for the different values of the correlation coefficient  $\rho$ . Once again robust  $M$ -estimators allow for a smaller error than classical Gaussian-based SMV-SCM.

We conclude that the robust methods presented in this chapter outperform significantly the classical Gaussian based SMV-SCM. Therefore,  $M$ -estimators offer a versatile alternative to Gaussian estimates. They allow to obtain better performances in impulsive environments while keeping good results in Gaussian background.

### 3.4 Adaptive non-Gaussian Detection

Through this non-Gaussian assumption, this section highlights the fact that robust estimation procedures presented in Section 3 are an interesting alternative to classical methods and can bring some great improvement to the detection process. The goal of this Section is to extend classical Gaussian based detectors to non-Gaussian framework. Furthermore, theoretical closed-form expression for false-alarm regulation is derived and the CFAR property is pursued to allow the detector to be independent of

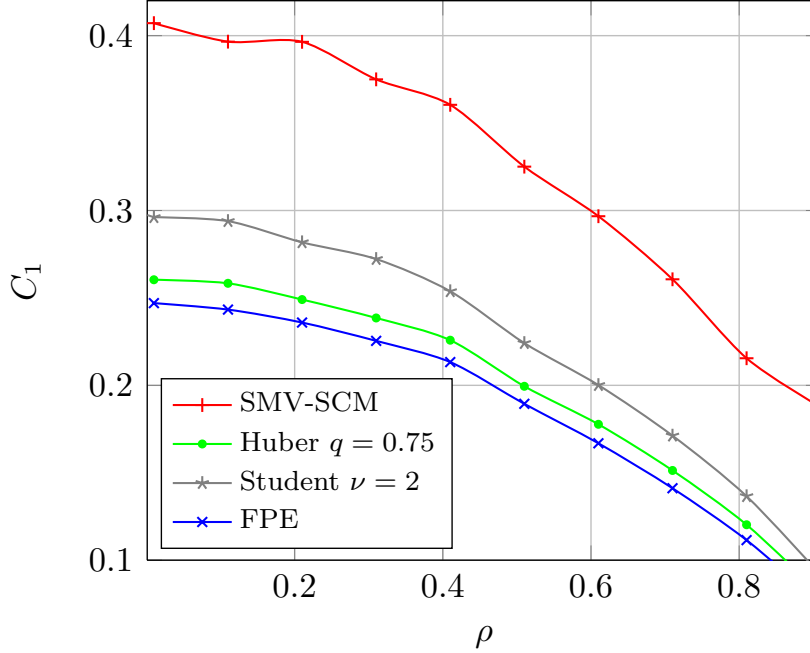


Figure 3.10: Performance comparison for different estimators for samples drawn from  $K$ -distribution with shape parameter  $\nu = 0.5$  for  $m = 10$  and  $N = 200$ .

nuisance parameters. This section presents the most important contribution of this work.

### 3.4.1 Adaptive Normalized Matched Filter

In non-Gaussian context, the ANMF detector takes advantage of its invariance properties and delivers better results than the other Gaussian based detectors, the AMF and the Kelly test, [Gini and Greco \(2002\)](#). If the background does not fulfill the Gaussian hypothesis, the detector performance can be deteriorated, increasing the false-alarm rate. To take into account the heterogeneity and non-Gaussianity for background modeling, a possible way is to use of the ANMF test built with robust estimates.

If some *a priori* knowledge of the noise statistics (e.g.,  $K$ -distribution,  $t$ -distribution, etc.) is available, then  $\Sigma$  and  $\mu$  can be estimated by the MLEs  $\hat{\Sigma}$  and  $\hat{\mu}$  of the covariance matrix and the mean vector of the assumed elliptical model. When there is no reliable statistical information on secondary data, they are assumed to be IID random samples from an unknown CE distribution. Then, practically any robust  $M$ -estimator could be used in the detector scheme. For heavy-tailed non-Gaussian background, robustness of the selected  $M$ -estimator is perhaps the most important design criterion.

We replace the covariance matrix and the mean vector by a robust MLE or  $M$ -estimator of scatter and location as they are consistent estimators of the covariance matrix up to a positive scalar and the mean vector within the class of CE distributions

(two-step GLRT). Thus, the ANMF for both the mean vector and the covariance matrix estimation takes the form (see Eq. (2.35)):

$$\Lambda_{ANMF, \hat{\Sigma}, \hat{\mu}} = \frac{|\mathbf{p}^H \hat{\Sigma}_N^{-1} (\mathbf{x} - \hat{\mu}_N)|^2}{(\mathbf{p}^H \hat{\Sigma}_N^{-1} \mathbf{p}) \left( (\mathbf{x} - \hat{\mu}_N)^H \hat{\Sigma}_N^{-1} (\mathbf{x} - \hat{\mu}_N) \right)} \stackrel{\mathcal{H}_1}{\geq} \lambda. \quad (3.55)$$

where  $\hat{\mu}_N$  and  $\hat{\Sigma}_N$  stand for any couple of  $M$ -estimators and where  $N$  stresses the dependency with the number of secondary data. Note that the ANMF falls into the class of homogeneous functions  $H(\cdot)$  of degree 0, i.e. the resulting detector does not depend on the scale factor of the matrix. When robust  $M$ -estimators are used jointly with the ANMF, the false-alarm can be regulated according to the following proposition.

**Proposition 3.4.1.** *The theoretical relationship between the PFA and the threshold for the ANMF, built with  $M$ -estimators of location and scatter  $\hat{\mu}_N$  and  $\hat{\Sigma}_N$ , is given by:*

$$PFA_{ANMF, \hat{\Sigma}, \hat{\mu}} = (1 - \lambda)^{a-1} {}_2F_1(a, a-1; b-1; \lambda), \quad (3.56)$$

with  $a = \sigma_1(N-1) - m + 2$  and  $b = \sigma_1(N-1) + 2$ , where  $N$  is the number of secondary data and  $m$  the dimension of the vectors.  $\sigma_1$  is related to the particular choice of  $M$ -estimators and is obtained according to:

$$\sigma_1 = \frac{\mathbb{E}[\psi_2^2(\sigma^2)]}{m(m+1)(1 + [m(m+1)]^{-1} \mathbb{E}[\sigma^2 \psi_2'(\sigma^2)])^2}. \quad (3.57)$$

*Proof.* The “PFA-threshold” relationship for the ANMF detector is perfectly known in Gaussian context and when the used estimators are the SMV and a Wishart matrix obtained with the SCM. The “PFA-threshold” is derived in Section 2.2.3 and recalled here :

$$PFA_{ANMF, \hat{\Sigma}, \hat{\mu}} = (1 - \lambda)^{a-1} {}_2F_1(a, a-1; b-1; \lambda), \quad (3.58)$$

where  $a = (N-1) - m + 2$  and  $b = (N-1) + 2$  and  ${}_2F_1(\cdot)$  is the hypergeometric function detailed in Eq. (1.23).

The statistical behavior of the  $M$ -estimators has been analyzed in Section 2.2.3. It has been shown that for  $N$  large enough all  $M$ -estimators are Wishart distributed. Therefore, their distribution rely on the asymptotic variance of the considered  $M$ -estimators,  $\sigma_1$ , detailed above. Compared with the classical SCM-SMV, the only change is the correction factor  $\sigma_1$  acting on  $(N-1)$ . For the general case of  $M$ -estimators, the relationship in Eq. (3.58) is verified for  $N$  large enough replacing  $N-1$  by  $(N-1)/\sigma_1$ .

This allows to give an approximated “PFA-threshold” relationship for the  $M$ -estimators



and for functions in the class of homogeneous functions of degree 0 (as it is the case for the ANMF). Indeed, we note that the test statistic in Eq. (3.55) stays the same if one substitutes  $\hat{\Sigma}$  by  $\hat{\Sigma}/\text{Tr}(\hat{\Sigma})$ . Thus, for  $N$  sufficiently large, the “PFA-threshold” relationship is given by:

$$PFA_{ANMF\hat{\Sigma},\hat{\mu}} = (1 - \lambda)^{a-1} {}_2F_1(a, a - 1; b - 1; \lambda), \quad (3.59)$$

with  $a = \sigma_1(N - 1) - m + 2$  and  $b = \sigma_1(N - 1) + 2$ . This function depends only on the size  $m$  of the vectors and on the number  $N$  of secondary data used for the estimation stage as well as the asymptotic variance  $\sigma_1$  of the considered  $M$ -estimators.  $\square$

**Remark 3.4.1.** *Note that the variance of the mean estimator will not affect the distribution as it appears both at the numerator and the denominator and subsequently, it disappears.*

Although FPE do not belong to the class of  $M$ -estimators (as they do not satisfy the conditions of Maronna), these results can also be extended to the FPE. The approximated “PFA-threshold” is obtained replacing in Eq. (2.36)  $N - 1$  by  $\frac{m}{m+1}(N - 1)$  as  $\sigma_1 = \frac{m+1}{m}$  which is an extension of [Pascal et al. \(2006\)](#) for unknown mean vector.

The CFAR property of this detector in any heterogeneous and/or non Gaussian background is reached when FPE are used. On the other hand, as the background is non-Gaussian and/or heterogeneous, statistical distribution of the ANMF built with the SCM estimate cannot be predicted but will surely vary with the background. The ANMF built with  $M$ -estimators (and particularly FPE) does overcome the non-Gaussianity and/or heterogeneity of the data. This implies, thanks to the properties of the CE, that the detector behaves according to the same distribution regardless of the true CE, i.e., it is distribution-free (see [Ollila and Tyler \(2012\)](#)). In addition, the asymptotic variance  $\sigma_1$ , which is always greater than 1, quantifies the loss of performance for the detector in Gaussian distributed background. Despite this small loss in Gaussian case,  $M$ -estimators bring robustness to the detection scheme and allow for perfect false-alarm regulation within the class of CE distributions. The improvement pointed for false alarm regulation leads to a better performance in probability detection terms. Notably, the SNR required to detect a target can be considerably decreased.

### 3.4.2 Simulations

In this section, we illustrate the properties of the selected detection test on simulated data. Firstly, false alarm regulation is analyzed for different scenarios and the proposed robust estimation procedures. Then, the performance of the different detection schemes is compared in terms of probability of detection.

#### 3.4.2.1 False Alarm Regulation

Firstly, the FA regulation is presented for the ANMF detection test.

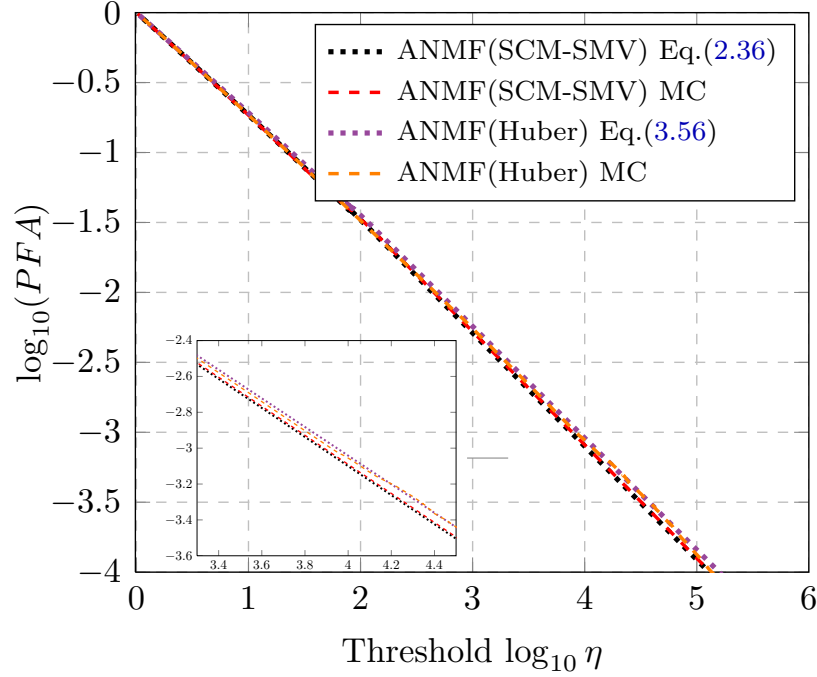


Figure 3.11: PFA-threshold for the ANMF under Gaussian distribution for  $m = 10$  and  $N = 50$  when (1) the SCM-SMV are used (red and black curves) (2) Proposition 3.4.1: the Huber estimates with  $q = 0.75$  are used (purple and orange curves)

#### 3.4.2.1.1 Gaussian analysis

The experiments were conducted on  $m = 10$  dimensional Gaussian vectors for  $N = 50$  secondary data and the computations have been made through  $10^6$  Monte-Carlo trials. The true covariance is chosen as a Toeplitz matrix whose entries are  $\Sigma_{i,j} = \rho^{|i-j|}$  and where  $\rho = 0.4$ . The mean vector is arbitrarily set to have all entries equal to  $(3 + 4j)$ . For clarity purposes, the results are displayed in terms of the threshold  $\eta$  which corresponds to take  $\eta = (1 - \lambda)^{-m}$ .

Fig. 3.11 presents the FA regulation for the ANMF under Gaussian assumption, for both the *Huber's*  $M$ -estimators and the SCM-SMV. The perfect agreement between theoretical relationship (purple curve) and simulated data (orange curve) validates the results of Proposition 3.4.1 and it shows that the correcting factor used for *Huber's*  $M$ -estimators (Eq. (3.57)) allows to perfectly regulate the FA, even in Gaussian context. Notice that the value of  $\sigma_1$  is close to 1 and the gap between the ANMF (SMV-SCM) and the ANMF (Huber) is very small. In order to better appreciate the separation between the curves, a zoom on a part of the figure is displayed at the bottom-left.

Fig. 3.12 and Fig. 3.13 present the results for the Student  $M$ -estimators and the FPE respectively analogous to those obtained with the *Huber's*  $M$ -estimators. The

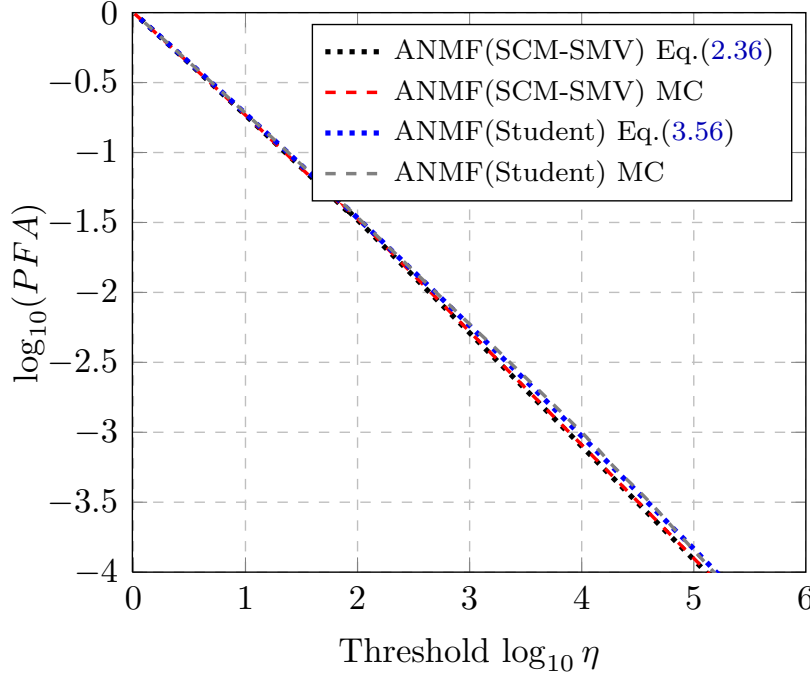


Figure 3.12: PFA versus threshold for the ANMF under a Gaussian distribution for  $m = 10$  and  $N = 50$  when (1) the SCM-SMV are used (red and black curves) (2) Proposition 3.4.1: the Student  $M$ -esimtors are used (blue and gray curves)

value of  $\sigma_1$  is greater for the FPE compared to the other estimators. Thus, the loss in performance is larger and the gap between ANMF (SMV-SCM) and the ANMF (FPE) is larger. However, this reduction in efficiency is offset by a greater robustness as it has been discussed above.

Further, Fig. 3.14 illustrates the fact that ANMF (FPE) built with  $(m+1)/m N$  secondary data (dark green curve) reaches the same performance than the ANMF built with SMV-SCM estimates (red curve) under Gaussian distribution. Regular FPE built with  $N$  secondary data is also plotted for comparison (yellow curve). It is worth pointing out that the error compared to the optimal case can be estimated as  $(m+1)/m$  and this correction factor allows us to regain the optimal solution. Moreover, the results for the ANMF (FPE) built with  $(m+1)/m N$  secondary data are very accurate even for small values of  $N$  (here  $N = 21$ ) and the approximation is adequate.

#### 3.4.2.1.2 Non-Gaussian analysis

In this section, the experiments have been realized over a  $K$ -distribution with shape parameter  $\nu = 0.5$  (see Section 3.2.3.3) for  $m = 10$  dimensional vectors,  $N = 50$  secondary data and the computations have been made through  $10^6$  Monte-Carlo trials. Under a  $K$ -distribution, as shown on Fig. 3.15, Fig. 3.16 and Fig. 3.17, the theoretical “PFA-threshold” relationship in Eq. (3.56) is in perfect agreement with

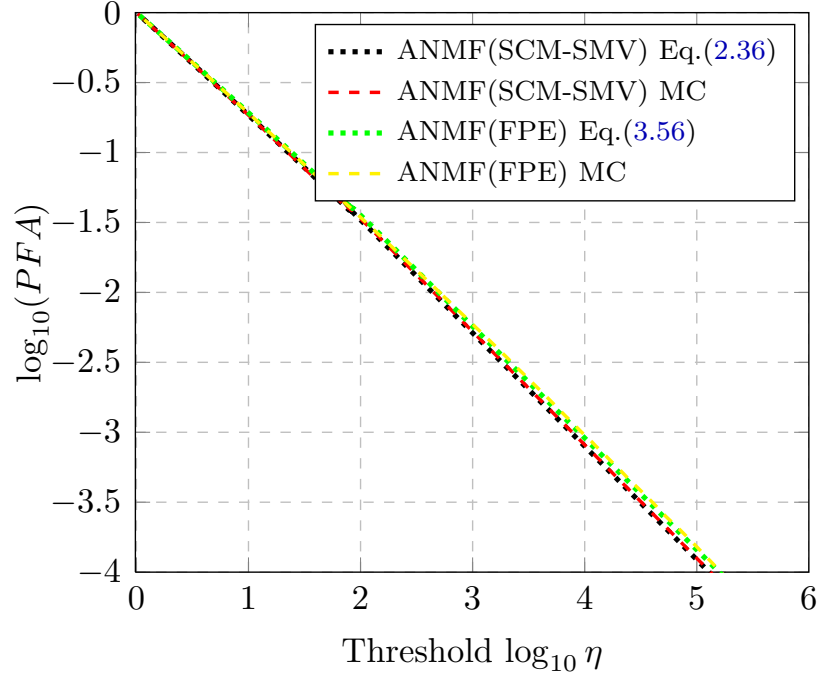


Figure 3.13: PFA versus threshold for the ANMF under a Gaussian distribution for  $m = 10$  and  $N = 50$  when (1) the SCM-SMV are used (red and black curves) (2) Proposition 3.4.1: the FPE are used (yellow and green curves)

the Monte-Carlo simulations for the *Huber's M*-estimates, Student *M*-estimators and FPE respectively, while for the SCM-SMV, the theoretical “PFA-threshold” relationship obtained in Section 2.2.3 is not valid anymore (since the Gaussian assumption is not respected anymore). We have left the theoretical “PFA-threshold” relationship for Gaussian estimators (black curve) for information.

Notice that on both Gaussian and *K*-distribution contexts, the FA regulation for the FPE leads to the same results. Thus, the curve just depends on the size of the vector  $m$  and on the number of secondary data  $N$ . This fact emphasizes the maximal invariance obtained with the ANMF built with the FPE, i.e. the distribution of the detector under only background hypothesis remains the same for all different impulsive distributions within the class of CE distributions. If one considers the subclass of SIRV distributions, the secondary data can be written according to the following representation:  $\mathbf{x} = \boldsymbol{\mu} + \sqrt{\tau} \mathbf{z}$  where  $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$ . It is easy to show that the texture  $\tau$ , that brings the spatial variation in power and determines the underlying SIRV distribution, disappears if we replace  $\mathbf{x}$  on the ANMF detector in Eq. 3.55 by its SIRV representation. However, the detector's distribution will not depend on the particular SIRV only if the covariance matrix estimator does not depend on the underlying distribution. This is the case only for the FPE, which are distribution-free, when used with an homogenous function, here the ANMF detection test. This has been referred as CFAR property and will be further detailed below.

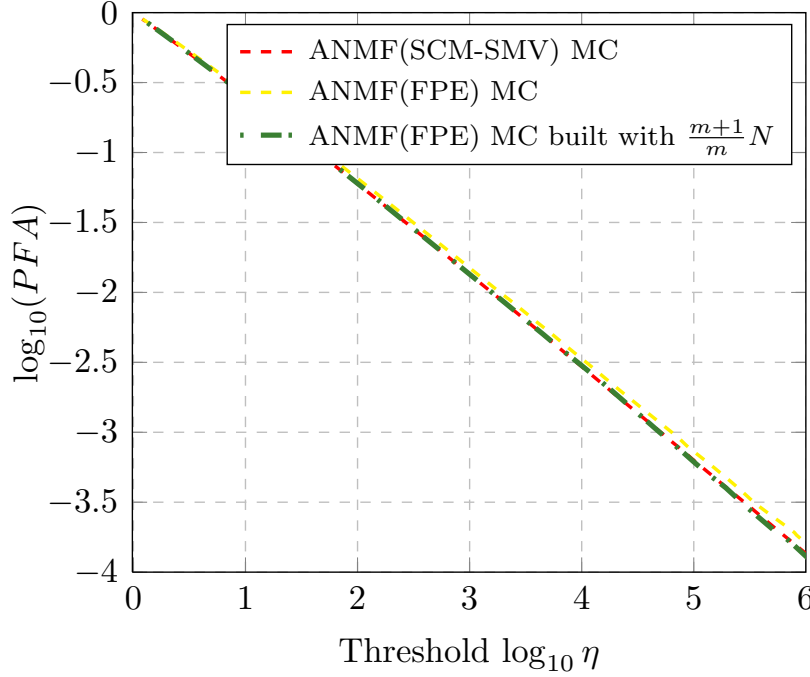


Figure 3.14: PFA versus threshold for the ANMF under a Gaussian distribution for  $m = 3$  and  $N = 21$  for (1) the SCM-SMV estimates (red) (2) the FPE (yellow) and (3) the FPE built  $m + 1/mN$  (dark green).

#### 3.4.2.1.3 Dependency on $m$ and $N$

As pointed out previously, the theoretical “PFA-threshold” relationship for the ANMF detection scheme only depends on the dimension of the vector  $m$  and on the number of secondary data  $N$  used to perform the parameters estimation. Hence, Fig. 3.18 shows the behavior of the ANMF detector built with the FPE as a function of the dimension  $m$ . The experiments were conducted on Gaussian distribution data for different values of  $m$ , with  $N = 50$  secondary data and the computations have been made through  $10^6$  Monte-Carlo trials. The true covariance is chosen as a Toeplitz matrix whose entries are  $\Sigma_{i,j} = \rho^{|i-j|}$  and where  $\rho = 0.4$ . The mean vector is arbitrarily set to have all entries equal to  $(3 + 4j)$ . In this figure, the results are displayed in terms of the threshold  $\eta = (1 - \lambda)$  which also does not depend on  $m$ . It is worth pointing out that for  $m = 25$  and  $N = 50$ , the FPE does not yield in asymptotic regime and the conditions for the Proposition 3.4.1 are not satisfied. Further, Fig 3.19 exemplifies the dependency of the detector’s distribution through the number of secondary data  $N$ . As  $N$  grows, the curves get closer to each other and tend to the known covariance matrix and known mean vector NMF. This is due to the fact that when  $N \rightarrow \infty$  the detector converges in distribution to the NMF.

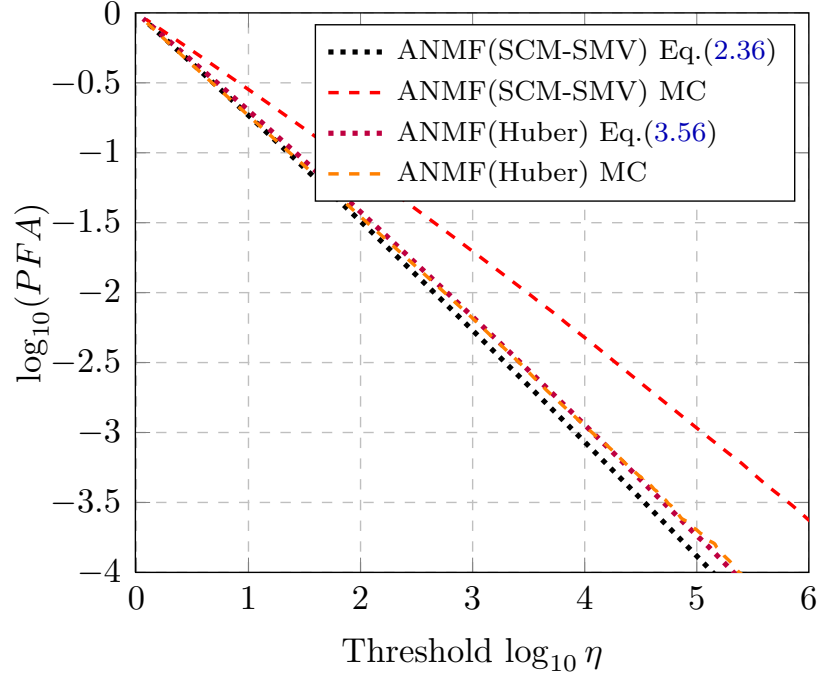


Figure 3.15: PFA-threshold under  $K$ -distribution with shape parameter  $\nu = 0.3$  for  $m = 10$  and  $N = 50$  when (1) the SCM-SMV are used (red and black) (2) Proposition 3.4.1: the Huber estimators with  $q = 0.75$  are used (orange and purple)

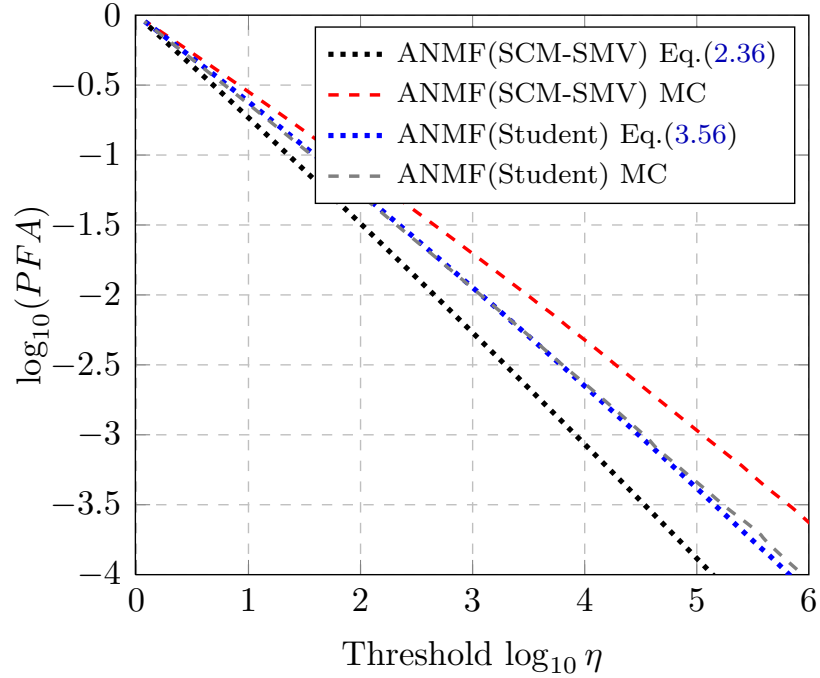


Figure 3.16: PFA-threshold under  $K$ -distribution with shape parameter  $\nu = 0.3$  for  $m = 10$  and  $N = 50$  when (1) the SCM-SMV are used (red and black) (2) Proposition 3.4.1: the Student  $M$ -estimators are used (gray and blue curves)

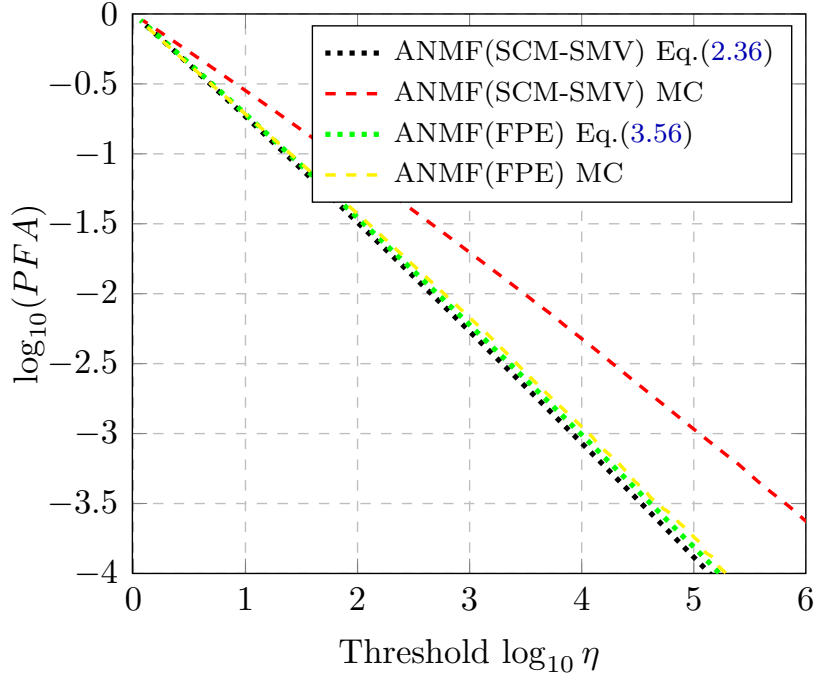


Figure 3.17: PFA versus threshold for the ANMF under a  $K$ -distribution with shape parameter  $\nu = 0.3$  for  $m = 10$  and  $N = 50$  when (1) the SCM-SMV are used (red and black curves) (2) Proposition 3.4.1: the FPE are used (yellow and green curves)

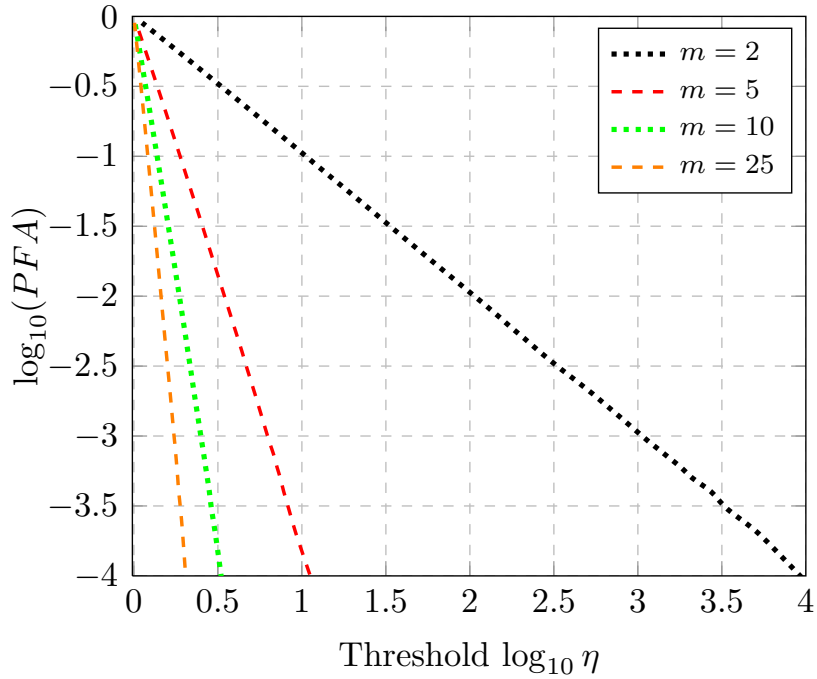


Figure 3.18: PFA versus threshold for the ANMF under a Gaussian distribution for the FPE obtained with  $N = 50$  secondary data and for different values of  $m$ .

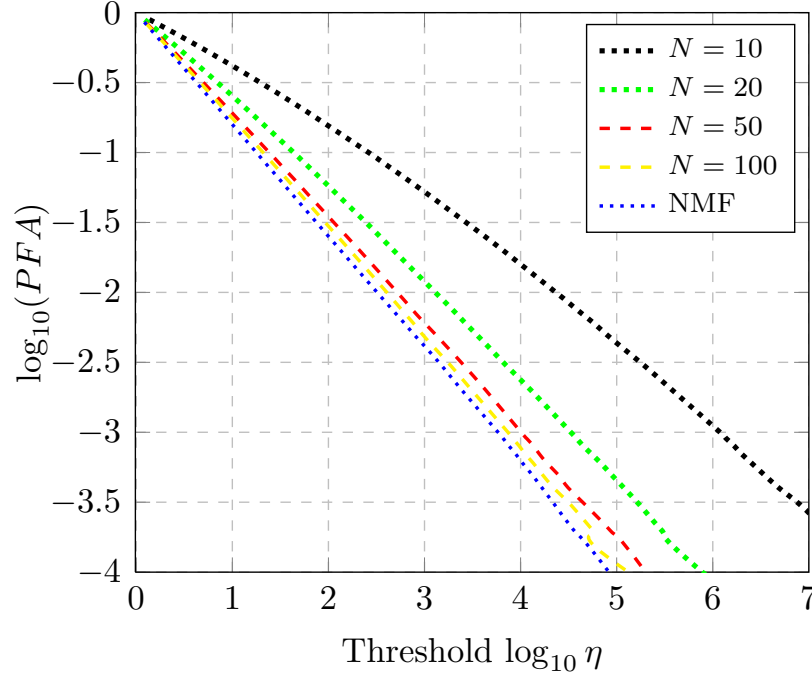


Figure 3.19: PFA versus threshold for the ANMF under a Gaussian distribution with  $m = 5$  for the FPE and for different values of  $N$  secondary data.

#### 3.4.2.1.4 CFAR properties

One of the most attractive properties of the ANMF (FPE) detector is its distribution invariance to the true matrix (CFAR-matrix), to the true mean vector (CFAR-mean) and to the underlying distribution itself (CFAR texture), i.e. the distribution of the detector remains the same even for impulsive distributions and for different parameters of the corresponding distributions. This fact is highlighted by the next figures. The experiments were conducted for  $m = 3$  with  $N = 21$  secondary data and the computations have been made through  $10^6$  Monte-Carlo trials. Fig. 3.20 depicts the detector for different values of the true covariance matrix chosen as a Toeplitz matrix and different values of the correlation coefficient,  $\rho = 0.01, 0.25, 0.5, 0.75$  and  $0.99$ . Is it clear that the resulting distribution does not depend on the true value of the covariance matrix. Moreover, the same experiments were realized for different values of the mean vector in Fig. 3.21 leading to the same distribution. Thus, the ANMF (FPE) detector holds the CFAR property with respect to the distribution parameters, the covariance matrix and the mean vector.

The most interesting property stands from the fact that the detector behaves according to the same distribution regardless of the true elliptical distribution, i.e. it is distribution-free. This is evidenced in Fig. 3.22 for different impulsive distributions. However, this is the case only for the FPE and not for the other  $M$ -estimators.



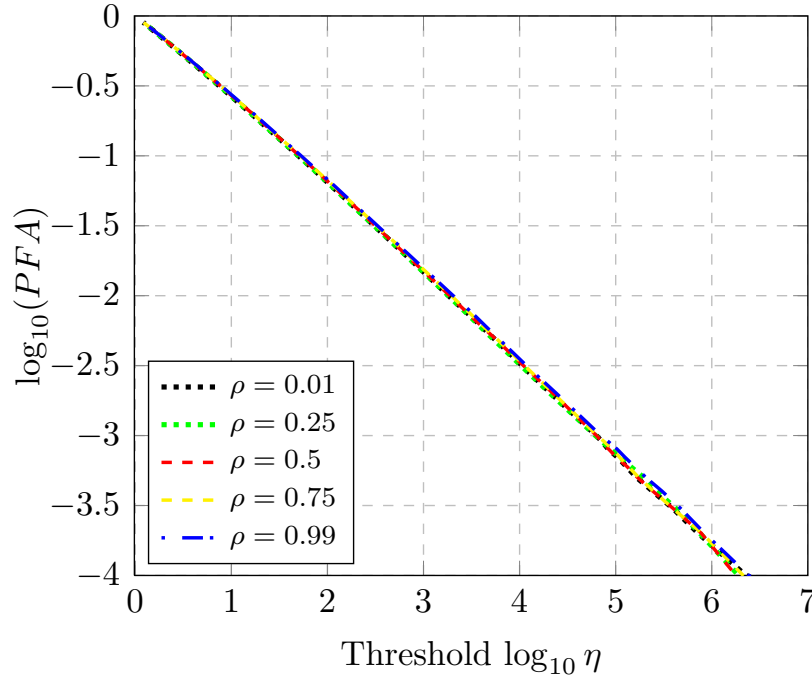


Figure 3.20: PFA versus threshold for the ANMF under a Gaussian distribution with  $m = 3, N = 21$  for the FPE and for different values of  $\rho$  secondary data CFAR MATRIX.

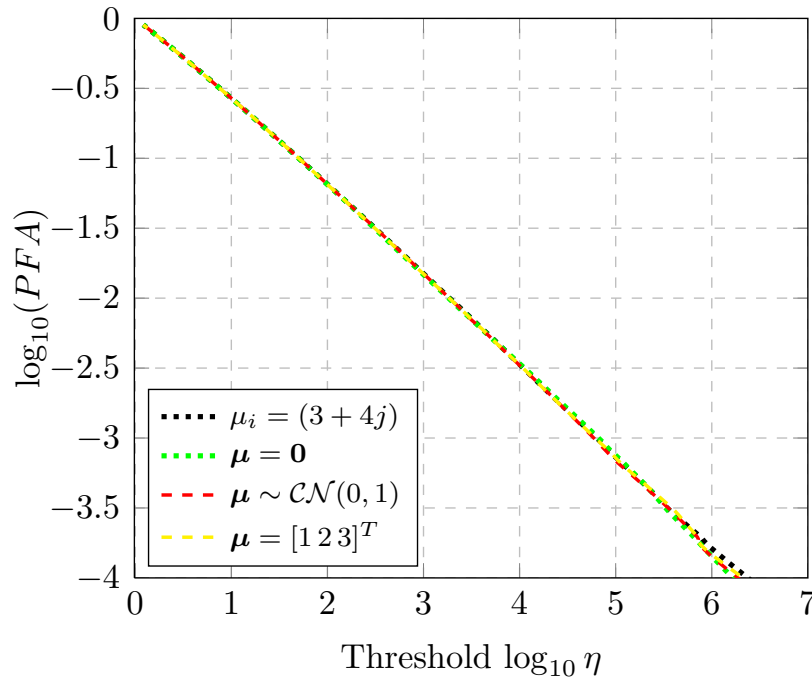


Figure 3.21: PFA versus threshold for the ANMF under a Gaussian distribution with  $m = 3, N = 21$  for the FPE and for different values of  $\mu$  secondary data.

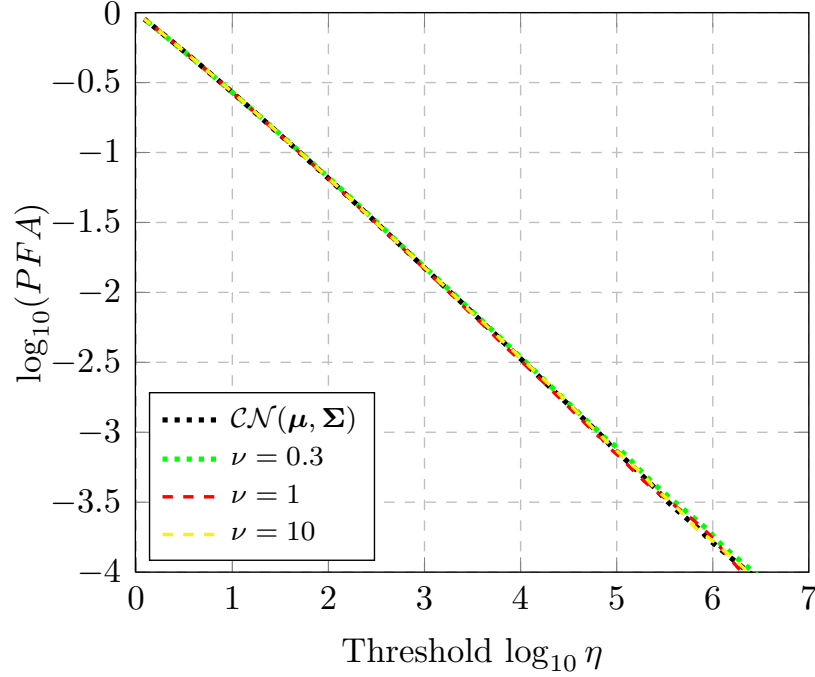


Figure 3.22: PFA versus threshold for the ANMF under a Gaussian distribution with  $m = 3$  and  $N=21$  for the FPE for different underlying distributions Gaussian and  $K$ -distribution with shape parameter  $\nu$ .

### 3.4.2.2 Performance evaluation

The classical SCM-SMV are compared to the robust  $M$ -estimators (*Huber's*  $M$ -estimators, Student  $M$ -estimators and FPE) in terms of probability of detection for the ANMF detection test. In addition, we plot the results for the classical Gaussian detector studied in Chapter 2. The experiments were conducted on  $K$ -distributed vectors with shape parameter  $\nu = 0.1$ , dimension  $m = 10$  and  $N = 50$ . The true covariance is chosen as a Toeplitz matrix whose entries are  $\Sigma_{i,j} = \rho^{|i-j|}$  and where  $\rho = 0.4$ . The mean vector is arbitrarily set to have all entries equal to  $(3+4j)$ . Firstly, one sets the probability of false alarm to a specific value,  $PFA = 10^{-3}$ . Then, the threshold is adjusted to reach the desired PFA, according to the false alarm regulation curves described above. The SNR is obtained according to:  $SNR = |\alpha^2| \mathbf{p}^H \Sigma^{-1} \mathbf{p}$ . Hence, one can add artificial targets with steering vector  $\mathbf{p}$  and the variation on the amplitude  $\alpha$  will lead to the different SNR values. The artificial targets signature used for the simulations is the unity vector  $\mathbf{p} = [1, \dots, 1]^T$  which corresponds to the non-prior approach that does not introduce any information into the detector. When data follow a multivariate  $K$ -distribution, the detector delivering the best performance results is the ANMF built with the robust  $M$ -estimators, notably with the FPE see Fig. 3.23. The performances obtained with classical SMV-SCM estimates are deteriorated for small SNR values. The target may be blurred in highly impulsive background. The required SNR is slightly smaller for the ANMF (FPE) detector for a given  $PD$  compared to the other  $M$ -estimators. Note that the classical Gaussian-based de-

tectors, the AMF and the Kelly test, are not suited for non-Gaussian background and need for a higher SNR. For instance, for  $PD = 0.5$ , the required SNR for the ANMF(FPE) is 8 dBs lower than for the ANMF (SMV-SCM) and more than 20 dBs smaller compared to the classical AMF and Kelly detectors.

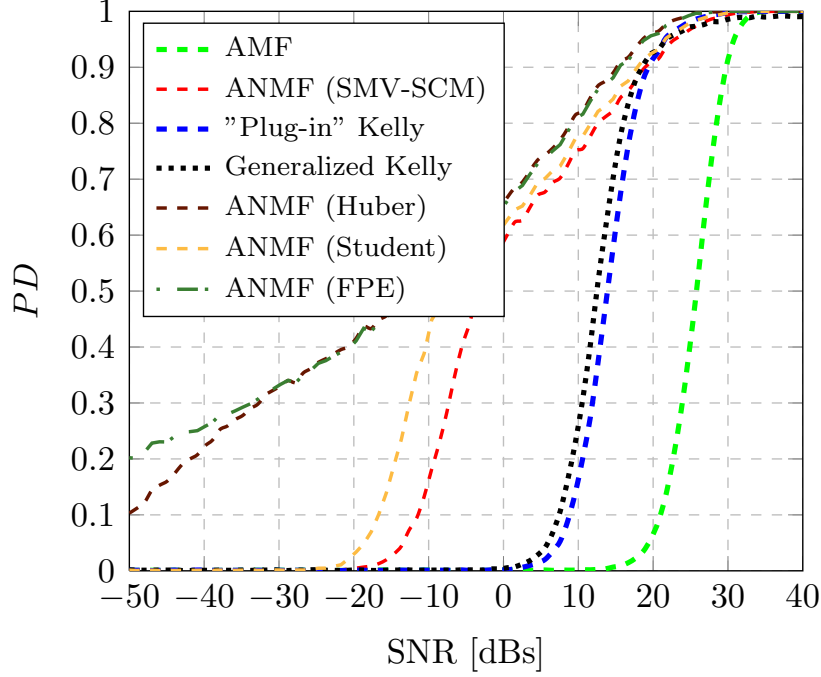


Figure 3.23: Probability of detection for different SNR values and  $PFA = 10^{-3}$  in non-Gaussian environment.

## 3.5 Anomaly Detection

All the detection schemes explained in Chapter 2 are derived under Gaussian assumption. In this section, we explore the use of robust estimation methods presented above for anomaly detection. These can then be used as plug-in estimators in place of the unknown mean vector and/or of the covariance matrix in the detection scheme. This is a simple but often efficient method to obtain robust properties for signal processors derived under the Gaussian assumption. As mentioned previously, the anomaly detectors allowing for better results are the Mahalanobis based anomaly detectors.

### 3.5.1 Robust Kelly Anomaly Detector

The Kelly anomaly detector has the advantage that the mean vector and the covariance matrix are independent to each other and to the observation vector, which is not the case for the Generalized Kelly or the classical RXD. This allows replacing the unknown parameters by the MLE or robust  $M$ -estimators and the detector can be

written as:

$$\Lambda_{KellyAD, \hat{\Sigma}, \hat{\mu}} = (\mathbf{x} - \hat{\mu}_N)^T \hat{\Sigma}_N^{-1} (\mathbf{x} - \hat{\mu}_N) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda, \quad (3.60)$$

and  $\hat{\mu}_N$  and  $\hat{\Sigma}_N^{-1}$  are the same as in Eq. (3.21) and Eq. (3.22). It is important to highlight that the distribution of this detector is still an open question, as far as the authors are aware. In fact, it will surely depend on the underlying particular CE distribution, i.e. the distribution will change with the choice of  $h_m(\cdot)$ .

### 3.5.2 Binary Partition Trees

In this section we propose to use a binary partition tree (BPT)-based approach to define the secondary data. The BPT is a hierarchical tree representation of the data that exploits the spatial and spectral information contained in the image as detailed in [Salembier and Garrido \(2000\)](#). This BPT-based definition of the secondary data allows to overcome the limitations of the conventional sliding window approach. On one hand it helps to define more homogeneous background regions and on the other hand it is more flexible in order to fit adequately the geometry of the background.

In the BPT representation, the leaf nodes correspond to an initial partition of the image, which can be the individual pixels, or a coarser segmentation map. From this initial partition, an iterative bottom-up region merging algorithm is applied until only one region remains. This last region represents the whole image and corresponds to the root node. All the nodes between the leaves and the root result from merging two adjacent children regions. An example of BPT is displayed in Fig. 3.24. If the initial partition contains  $n$  leaf nodes, the final BPT contains  $2n - 1$  nodes.

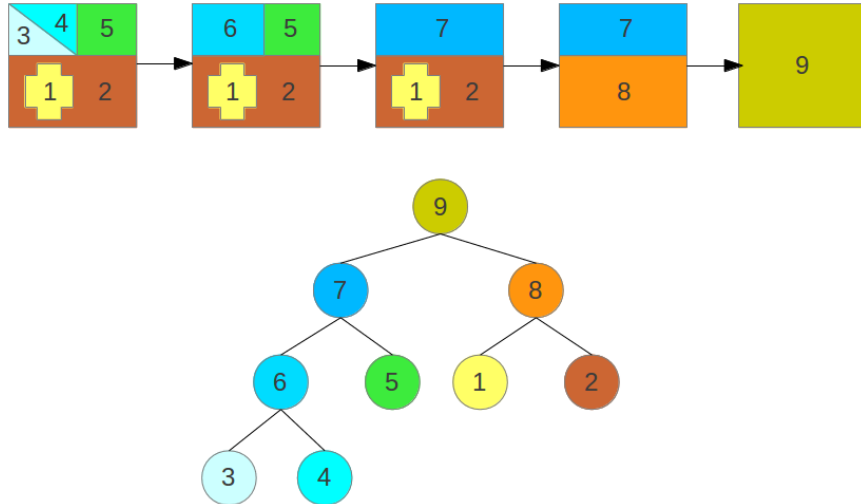


Figure 3.24: Construction of the Binary Partition Tree (BPT).

Two notions are of prime importance when defining a BPT:

1. The *region model*  $\mathcal{M}_{\mathcal{R}}$  which specifies how a region  $\mathcal{R}$  is modeled,
2. The *merging criterion*  $\mathcal{O}(\mathcal{M}_{\mathcal{R}_\alpha}, \mathcal{M}_{\mathcal{R}_\beta})$ , which is a distance measure between the region models of any two regions  $\mathcal{R}_\alpha$  and  $\mathcal{R}_\beta$ .

Each merging iteration involves the search of the two adjacent regions which achieve the lowest pair-wise similarity among all the pairs of adjacent regions in the current segmentation map. Those two regions are consequently merged.

Given a hyperspectral region  $\mathcal{R}$ , with  $N_{\mathcal{R}}$  hyperspectral samples  $\mathbf{x}_i \in \mathbb{R}^m$ ,  $i \in 1, \dots, N_{\mathcal{R}}$ , the first-order parametric model  $\mathcal{M}_{\mathcal{R}}$  is defined by the sample mean vector of the hyperspectral samples  $\hat{\boldsymbol{\mu}}_{\mathcal{R}}$ :

$$\mathcal{M}_{\mathcal{R}} : \hat{\boldsymbol{\mu}}_{\mathcal{R}} = \frac{1}{N_{\mathcal{R}}} \sum_{i=1}^{N_{\mathcal{R}}} \mathbf{x}_i. \quad (3.61)$$

Using the first-order parametric model (3.61), a merging criterion is defined as the spectral angle distance,  $d_{\text{SAM}}$ , between the sample mean vectors of any two adjacent regions [Veganzones et al. \(2013\)](#):

$$\mathcal{O}(\mathcal{M}_{\mathcal{R}_\alpha}, \mathcal{M}_{\mathcal{R}_\beta}) : d_{\text{SAM}}(\hat{\boldsymbol{\mu}}_{\mathcal{R}_\alpha}, \hat{\boldsymbol{\mu}}_{\mathcal{R}_\beta}), \quad (3.62)$$

where  $d_{\text{SAM}}(\mathbf{a}, \mathbf{b}) = \arccos\left(\frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|}\right)$ .

The building of a BPT may suffer from small and meaningless regions resulting in a spatially unbalanced tree. To overcome this limitation, a priority term is included in the merging criterion that forces those regions smaller than a given percentage of the average region size to be merged first as explained in [Tochon et al. \(2012\)](#).

Let a branch of the observation test,  $\mathbf{x}$ , denoted as  $\mathcal{B}(\mathbf{x})$ , define the sequence of nodes ascending on the BPT representation from the leaf containing the observation test up to the root node. For instance, given the BPT representation depicted in Fig. 3.24, the branch of an observation vector in the leaf number 5 is:  $\mathcal{B}(\mathbf{x}) = \{5 \rightarrow 7 \rightarrow 9\}$ . This branch definition is a sorted list of nodes starting in the leaf node and ending in the root node. Then, the guard and outer nodes as defined as follows:

- The *guard node* of an observation vector,  $G(\mathbf{x})$ , is the first node in the observation vector's branch,  $\mathcal{B}(\mathbf{x})$ , containing the observation vector and at least, a given number of guard pixels that will be dismissed in the estimation stage.
- The *outer node* of an observation vector,  $O(\mathbf{x})$ , is the first node in the observation vector's branch,  $\mathcal{B}(\mathbf{x})$ , containing the observation vector, the guard pixels and at least, a given number of secondary pixels used to perform the parameters estimation.

Thus, in order to define the guard and outer nodes it is necessary to set the number of guard and secondary pixels, which works in a similar manner to the guard and outer windows size respectively. The secondary data,  $S(\mathbf{x})$ , is then defined by the pixels contained in the outer node once the pixels in the guard node has been removed:

$$S(\mathbf{x}) = O(\mathbf{x}) \setminus G(\mathbf{x}), \quad (3.63)$$

where  $A \setminus B$  denotes the complement operation between sets  $A$  and  $B$ .

## 3.6 Summary

We have detailed the class of complex elliptical distributions as a general model for background characterization in Hyperspectral Imaging. Elliptical distributions account for heterogeneity and long tail distributions present in real hyperspectral data. Once established that hyperspectral data cannot fit a multivariate Normal distribution, the use of the Gaussian MLE (SCM and SMV) does not provide the optimal parameter estimation. We propose the use of robust estimates for the mean vector and the covariance matrix. We have described the  $M$ -estimators, notably the *Huber* type estimators,  $t$ -distribution estimators and the Fixed Point approach. Further, we have described robust shrinkage estimators suited for high-dimensional estimation problems. The joint estimation of both covariance matrix and mean vector is a new challenging problem that opens many unknowns. We have introduced the use of these estimates on classical detection methods. We conclude that the robust estimation tools presented in this chapter offer a versatile alternative to Gaussian estimates. We remark that proposed  $M$ -estimators in Gaussian environment are capable of reaching the same results as the SCM and SMV. On the other hand, they outperform the classical estimation methods in case of non-Gaussian impulsive noise. This adaptability and their robustness make them suitable estimates in most scenarios.



# 4

## Application to Hyperspectral Imaging

### 4.1 Introduction

As mentioned previously, Hyperspectral data represent the radiation at a large number of wavelength for each position in an image. Therefore, they are real and positive and the classical zero-mean (or known mean) Gaussian hypothesis is not valid anymore. The mean vector has to be jointly estimated with the covariance matrix and included in the adaptive detection scheme. For this purpose, Chapter 2 establishes the theoretical framework for adaptive non-zero mean Gaussian target detection. However, the performance of these detection methods strongly relies on the statistical parameters estimation. Accordingly, when the background is non-homogeneous or the noise independence assumption is not fulfilled, the detector performance can be deteriorated. Chapter 3 details the class of elliptical distributions for background characterization as a generalization of Gaussian model. In this context, robust estimation methods are considered for joint mean vector and covariance matrix estimation. These methods are introduced for the corresponding adaptive detectors. The aim of this work is to highlight the improvement arising from the robustness of the  $M$ -estimators described in the previous chapter. Although  $M$ -estimators are not adapted to high dimensional problems, we have explored some dimensionality reduction techniques in order to decrease the number of bands in real hyperspectral data. However, shrinkage approach detailed in the previous chapter or the use of projection methods could be more suitable solutions.

The theoretical proposals have been analyzed in the previous chapters through simulations. Here, we discuss the results obtained in real hyperspectral images. The detection schemes are evaluated for four data sets: the NASA Hyperion Image, the Blind Test HYMAP Image, the Rural scene and synthetic galaxies MUSE project.



## 4.2 NASA Hyperion Image

The scene analyzed is the NASA Hyperion sensor dataset displayed in Fig. 4.1. The image is constituted of  $798 \times 253$  pixels and 116 spectral bands after water absorption bands have been removed.

In order to satisfy Gaussian distribution assumption, the analysis has been done on a homogenous part of the image corresponding to the water region on the top-right of the image. The extracted part consists in  $60 \times 20$  pixels. To ensure the validity of the proposed methods, we show in Fig. 4.2 the outcome of a classical Gaussianity test “Q-Q plot” for the selected region over the band 42. In addition, Fig. 4.3 depicts the histogram for the reflectance values of the pixels belonging to the chosen area and the approximative fit with a Gaussian distribution. Even if this allows to “validate” the Gaussianity of each band, it cannot ensure the Gaussianity of the corresponding multivariate vector.

Since hyperspectral data are real and positive, we propose to use a Hilbert filter in order to render them complex to fit the derived expressions for target detection techniques. A downsampling taking one over two consecutive bands is required to avoid redundant information that can reduce the covariance matrix rank. However, it is important to note that the real component after Hilbert transform is still the original signal. To avoid the well-known problem due to high dimensionality, we have chosen sequentially six bands in the complex representation. In this approach, both covariance matrix and mean vector are estimated using a sliding window of size  $5 \times 5$ , having  $N = 24$  secondary data.

### 4.2.1 False Alarm Regulation

The outcome of the AMF, Kelly test and ANMF detection schemes under null hypothesis for this image are shown on the Fig. 4.4, Fig. 4.5 and Fig. 4.6 respectively. The axis in Fig. 4.4 represent the threshold  $\lambda$  of the detector for the AMF test in Eq. (2.1). While the results in Fig. 4.5 and Fig. 4.6 are displayed in terms of the threshold  $\eta$  from Eq. (1.29),  $\eta = (1 - \lambda)^{-(N+1)}$  for Kelly’s test, and Eq. (1.34),  $\eta = (1 - \lambda)^{-m}$  for the ANMF detector, respectively and a logarithmic scale is used. The results obtained on real HSI data on a Gaussian distributed region agree with the theoretical relationships presented above. Remark that the false-alarm rate that can be achieved depends on the number of points on which the detector is calculated (in a similar manner to the Monte-Carlo trials). As the homogenous area is bounded and the data set is small, the distribution of the detectors may divert for small values of the PFA directly related to the size of the region.

Depending on the underlying material, the distribution of the detector might divert from the expected behavior when Gaussian distribution is assumed. This is the case on these real data since the extracted area is not perfectly Gaussian. Moreover, we have studied other regions of this image for which we have observed a strong non-Gaussianity. Fig 4.7 (a) depicts the corresponding “Q-Q plot” of an



Figure 4.1: True color composition of the Hyperion scene.

extracted region on the top-left part of the image (in orange in the true color image) over the band 8 and Fig. 4.7 (b) shows the results of the Gaussianity test on the cloudy region on the bottom-left over band 20. Notice that the tails deviate from the Gaussian distribution. This suggests the use of non-Gaussian distributions to model the background for hyperspectral imaging.

#### 4.2.2 Performance Evaluation

We present here the results in terms of probability of detection in a real hyperspectral scene. Foremost, the threshold is determined to ensure a  $PFA = 10^{-2}$  according to the “PFA-threshold” relationships detailed in Section 4.2.1. In contrast, the threshold for the generalized Kelly test is obtained empirically from the data. As pointed above, we analyze here an homogenous Gaussian distributed region. The SNR is

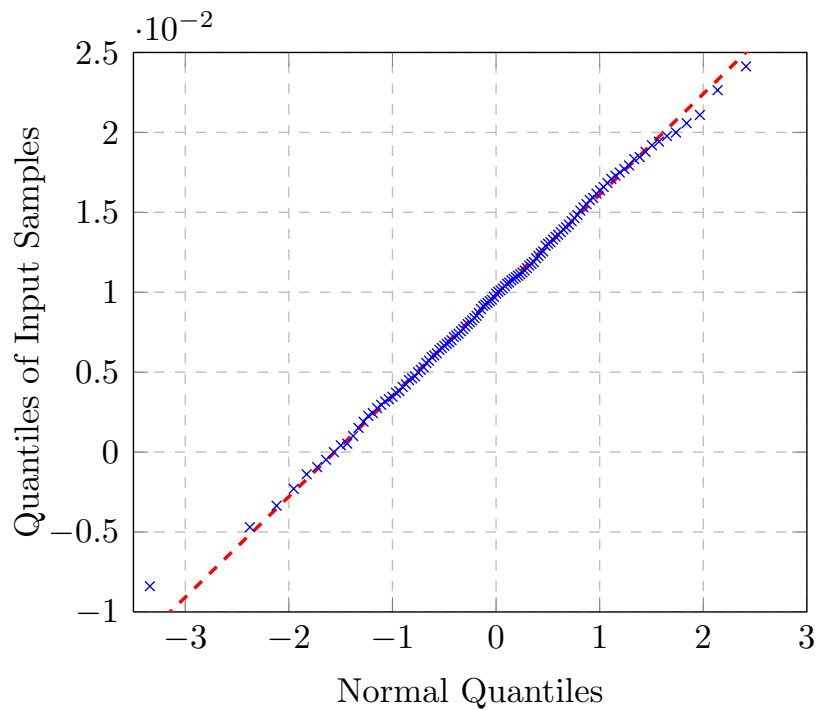


Figure 4.2: Q-Q Plot of the data sample versus the Normal theoretical distribution.

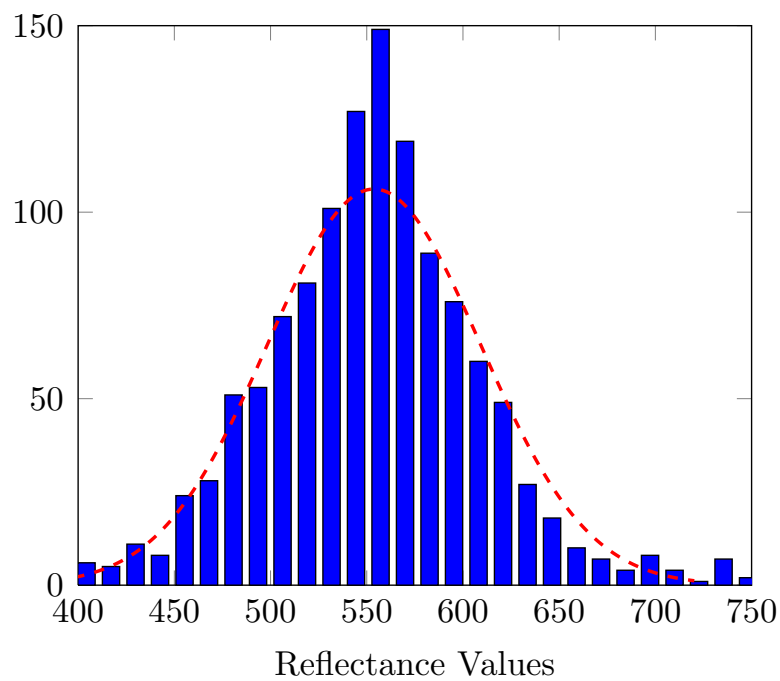


Figure 4.3: Histogram of the data sample and comparison with the Gaussian probability density function.

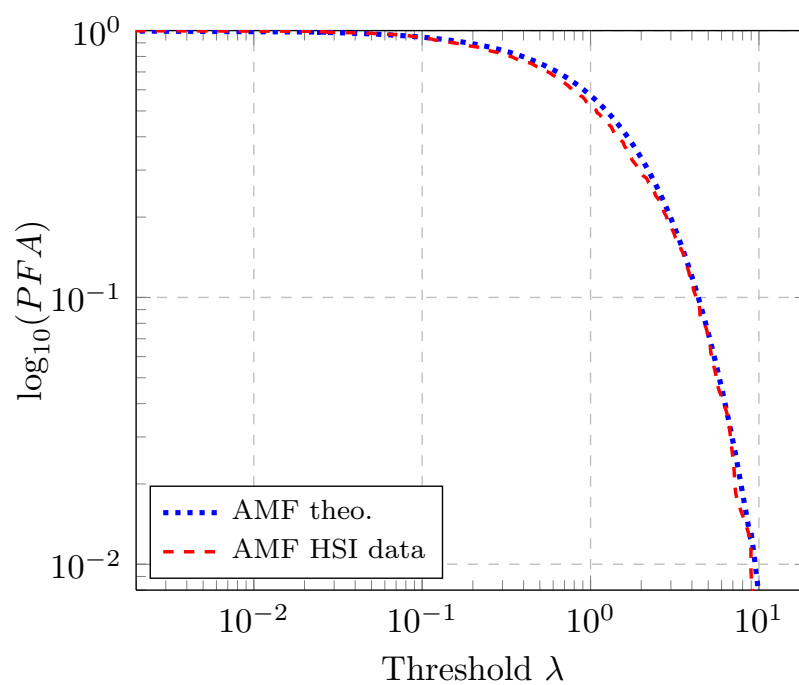


Figure 4.4: AMF false-alarm regulation for a real HSI image

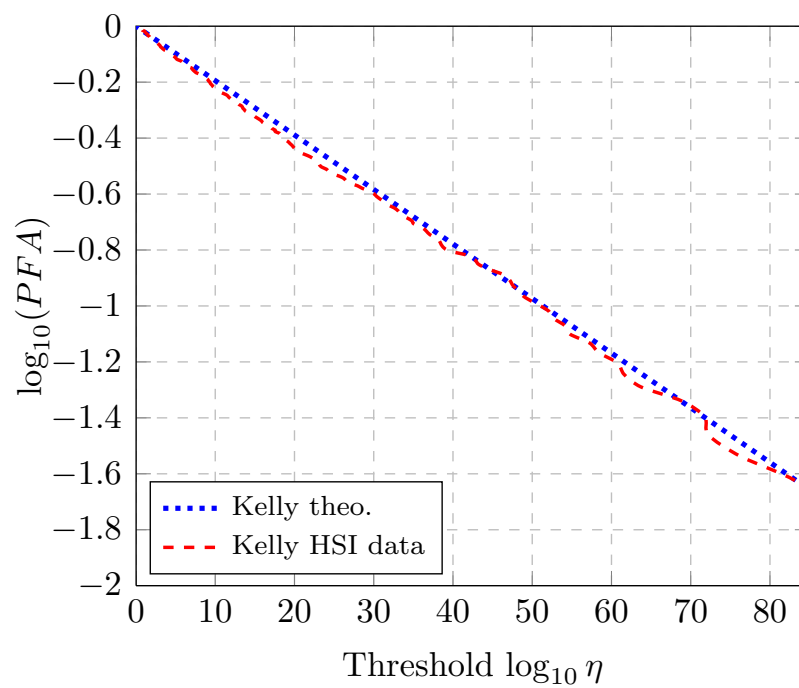


Figure 4.5: Kelly false-alarm regulation for a real HSI image

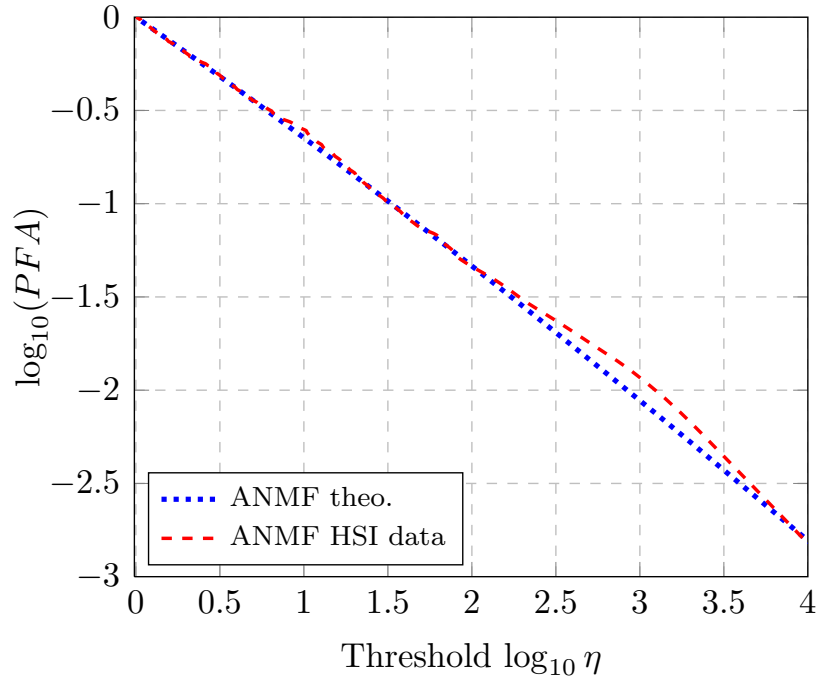


Figure 4.6: ANMF false-alarm regulation for a real HSI image

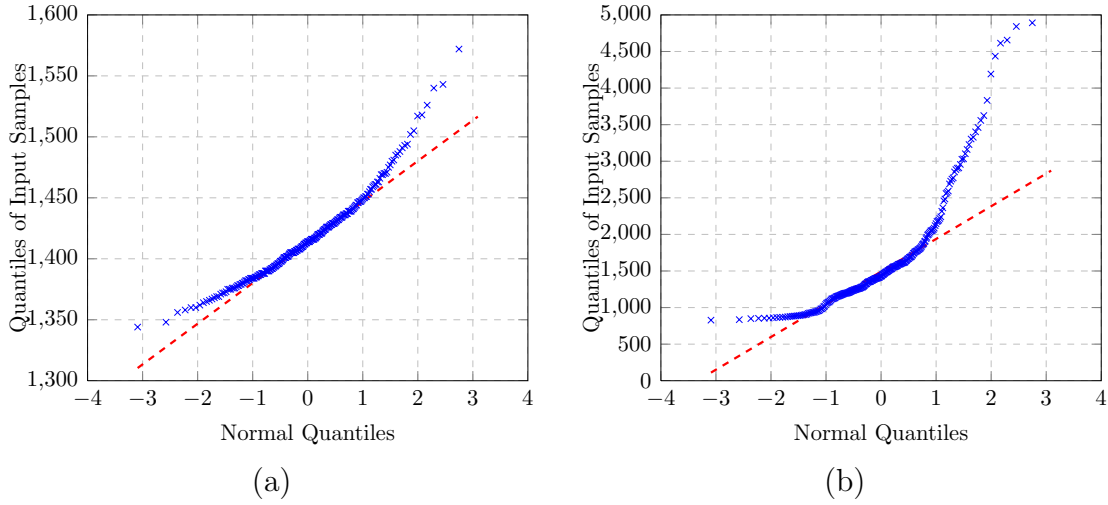


Figure 4.7: Q-Q Plot of data sample versus the Normal theoretical distribution.

obtained according to:  $SNR = |\alpha^2| \mathbf{p}^H \mathbf{\Sigma}^{-1} \mathbf{p}$ . Hence, one can add artificial targets with steering vector  $\mathbf{p}$  and the variation on the amplitude  $\alpha$  will lead to the different SNR values. The artificial targets signature used for the simulations is the unity vector  $\mathbf{p} = [1, \dots, 1]^T$  which corresponds to the non-prior approach that does not introduce any information into the detector. The goal of this experiment is to retrieve the same behavior and the same performance as in Fig. 2.4. Thus, the proposed generalized Kelly performs similarly to the classical Kelly test. Moreover, Kelly detection test allows for a lower SNR requirements compared to the AMF and the ANMF detectors, as expected.

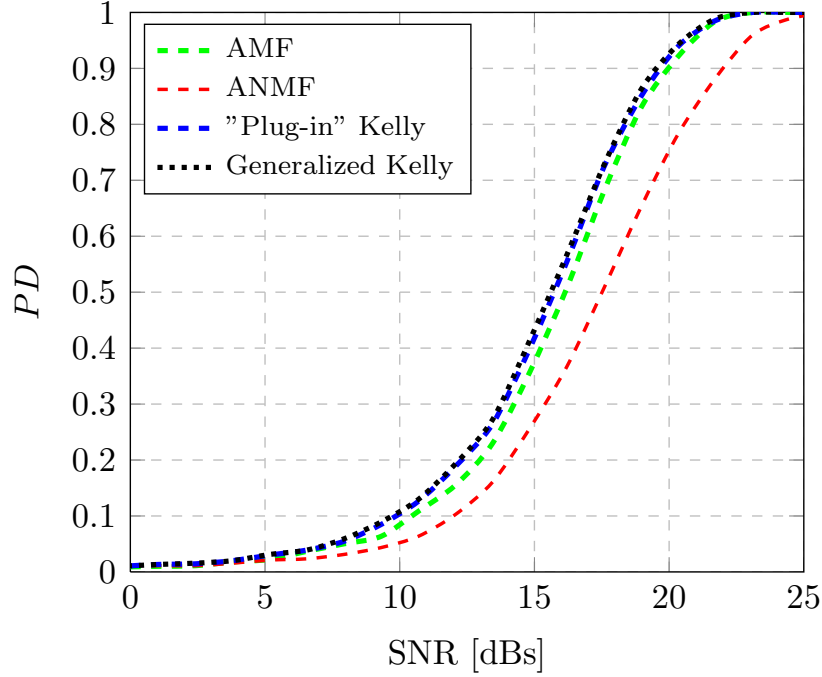


Figure 4.8: Probability of detection for different SNR values and  $PFA = 10^{-2}$  for real HSI data.

### 4.2.3 Anomaly Detection

Fig. 4.9 shows the resulting distribution of Kelly anomaly detector on real data. The distribution of the detector is plotted in red. We also depict the theoretical relationship defined by Eq. (2.64). Results obtained on real HSI data on a Gaussian distributed region agree with the theoretical relationships presented above.

Finally, we illustrate the detection capability of the proposed methods when artificial anomalies with known spectral signature are inserted on the real hyperspectral image. For this purpose, we extract the spectral signature from ground materials in Fig. 4.1 and the anomaly spectra used as steering vector is depicted in Fig. 4.10. Fig. 4.11 (a) details the position and the shape of the targets. For a fixed value of

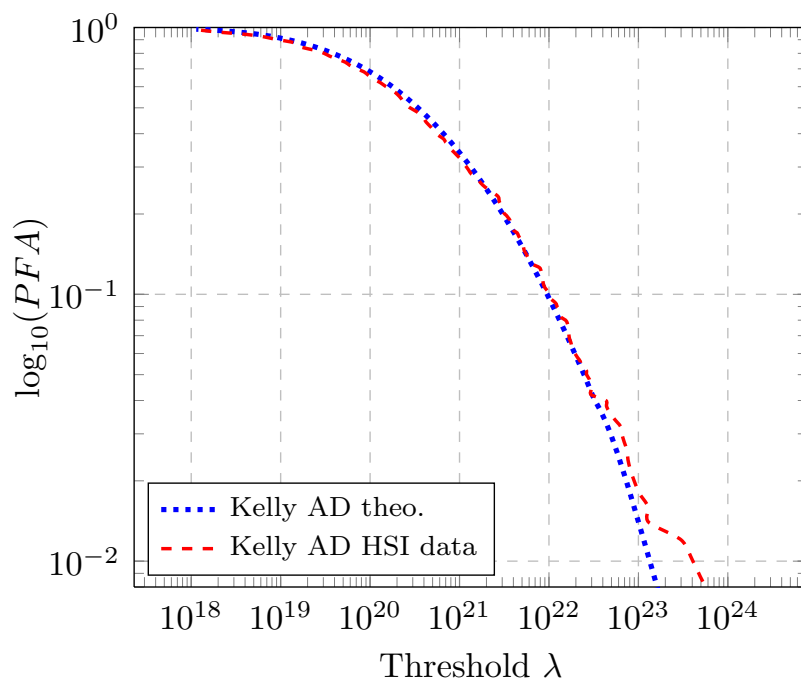


Figure 4.9: Kelly AD complementary CDF of the Mahalanobis distance for a real HSI image

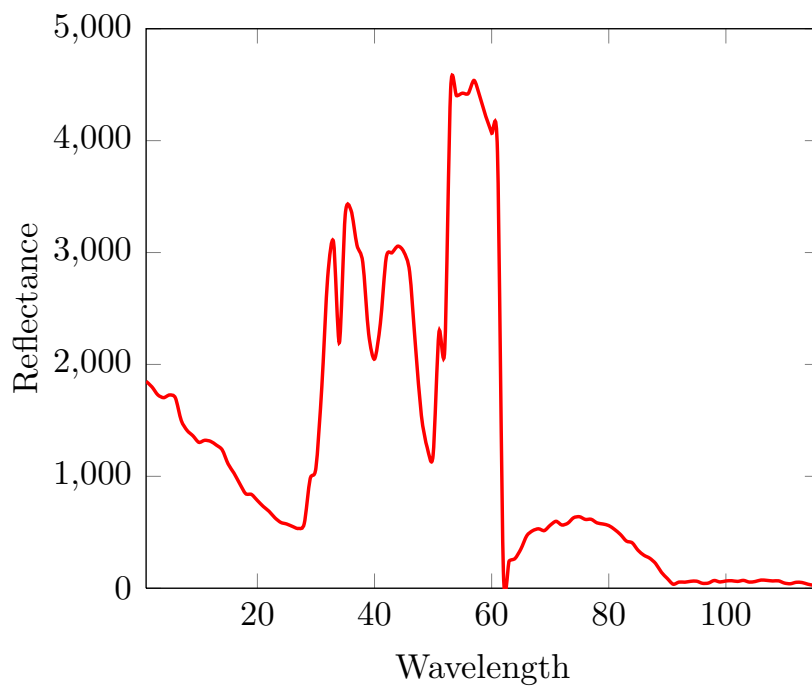


Figure 4.10: Endmember used in the experiment which corresponds to land materials.

FA,  $PFA = 0.1$ , we present in Fig. 4.11 the outcome of the different detectors. All the results are displayed for a same PFA value, regardless of the real image scale, i.e. the color level is obtained with respect to the background contrast. For that purpose, we compare only PFA and PD and not the absolute scale of the image in terms of power. The detectors based on the Mahalanobis distance, the RXD, the Kelly AD and the Generalized Kelly, deliver best results for detection purposes as expected and the matched filter based detectors, the N-RXD and the UTD, do not detect properly the artificial targets. These detection maps are in agreement with the SNR figure detailed above. Moreover, Kelly AD and Generalized Kelly detector provide the best detection results while keeping the false alarm lower. Classical RXD allows for detection of the one-pixel targets, while UTD detection results locate only some of the targets. N-RXD results are very poor, strong residual background samples are detected instead of the targets. Its use should be avoided for anomaly detection in Gaussian environment. Remark that the two-pixel targets are not detected by any of the detection schemes. This problem is due to the presence of a strong target in the secondary data that pollutes the covariance matrix estimation. Its occurrence has a significant impact on the detection process and it degrades the detectors' performances.

Fig. 4.12 shows the results of the  $\Lambda_{Kelly AD}$  built now with FPE and with shrinkage estimators, both SCM and FPE. The same value of FA,  $PFA = 10^{-1}$  is considered. For the shrinkage estimators, the regularization parameter  $\beta$  has been fixed to  $\beta = 0.8$ . We have empirically optimized the results over  $\beta$  leading to the chosen value. Note that there are theoretical techniques for optimizing over  $\beta$ , detailed in Couillet and McKay (2014), Chen et al. (2011b), that could be used instead. Remark that all the anomalies of interest are now detected even those bigger than one pixel. FPE and shrinkage FPE take advantage of their robustness to outliers and in this case, the outliers are the strong targets present in the secondary data. Thereafter, the proposed estimation methods allow for better detection results even in Gaussian case.



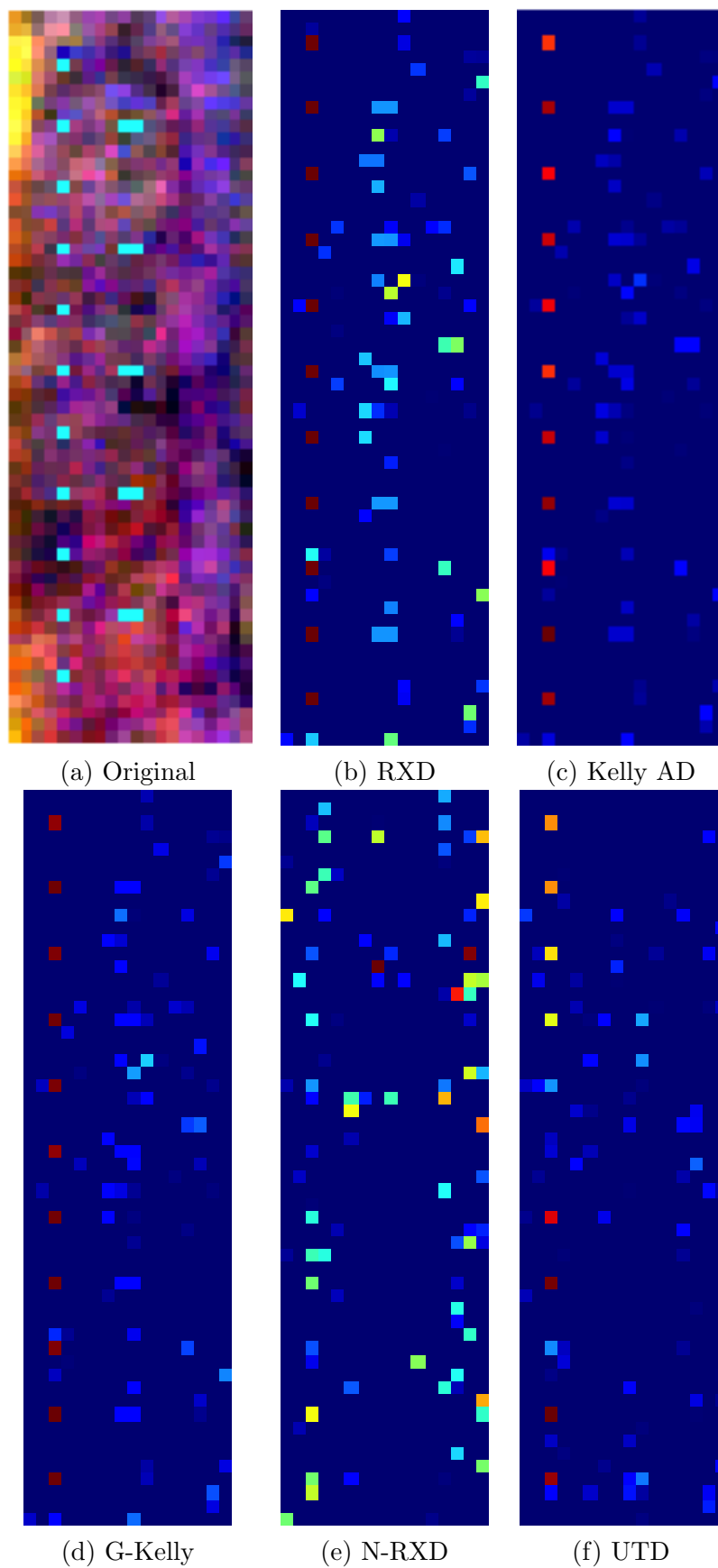


Figure 4.11: Anomaly detection for artificial targets in real HSI.

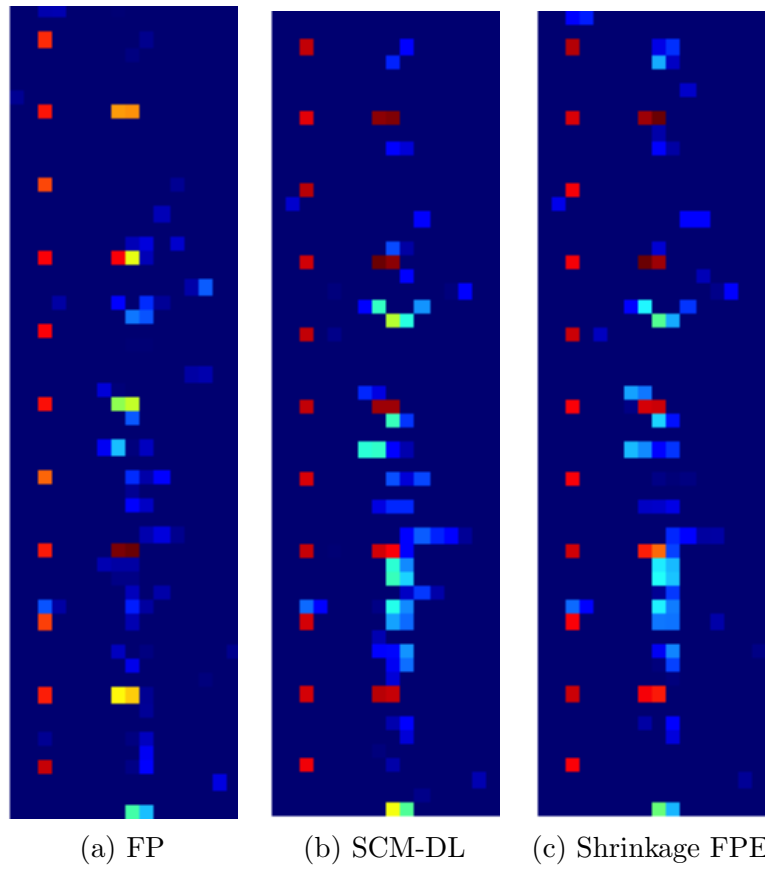


Figure 4.12: Kelly AD built with robust estimates for artificial targets in real HSI.

### 4.3 Blind Test HYMAP Image

Experimental hyperspectral scenes cannot be characterized using only classical Gaussian models. As it was mentioned in [Manolakis and Marden \(2002\)](#), HSI data often present tails heavier than the Normal distribution. Hence, the class of elliptical distributions is chosen as a more general assumption.

We illustrate here the non-Gaussianity of a real hyperspectral image. The scene analyzed is the Target Detection blind test provided in [Snyder et al. \(2008\)](#). The image was collected in July 2006 around the small town of Cooke City, Montana, USA. The hyperspectral imagery was collected by the HyMap sensor operated by HyVista. The HyMap imagery has approximately 3 meter ground resolution. Civilian vehicles and small fabric panels were used as targets. The image is constituted of  $280 \times 800$  pixels and 126 spectral bands. The true color composition of the scene is depicted on Fig. 4.13.



Figure 4.13: True color composition of the HyMap scene.

Fig. 4.14 exhibits the “Q-Q plot” of the data contained in the bands 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup> and 100<sup>th</sup>. On this examples, we observe that the tails are different from Gaussian distribution. Moreover, Fig. 4.15 depicts the histogram for the reflectance values and the approximative fit with a Gaussian distribution, notice that the Gaussian distribution does not characterize properly the empirical distribution. As the marginal distributions cannot fit a Normal univariate distribution, neither will the joint distribution. This suggests the need of non-Gaussian distributions to model the background for hyperspectral imaging.

Exceedance plots, as mentioned in the previous chapter, are essentially cumulative histograms of Mahalanobis distance values, and they provide a useful way to visualize the fat tails of multivariate distributions. Fig. 4.16 shows the theoretical results when Gaussian distribution is assumed and experimental hyperspectral data distribution which diverts from the expected behavior.

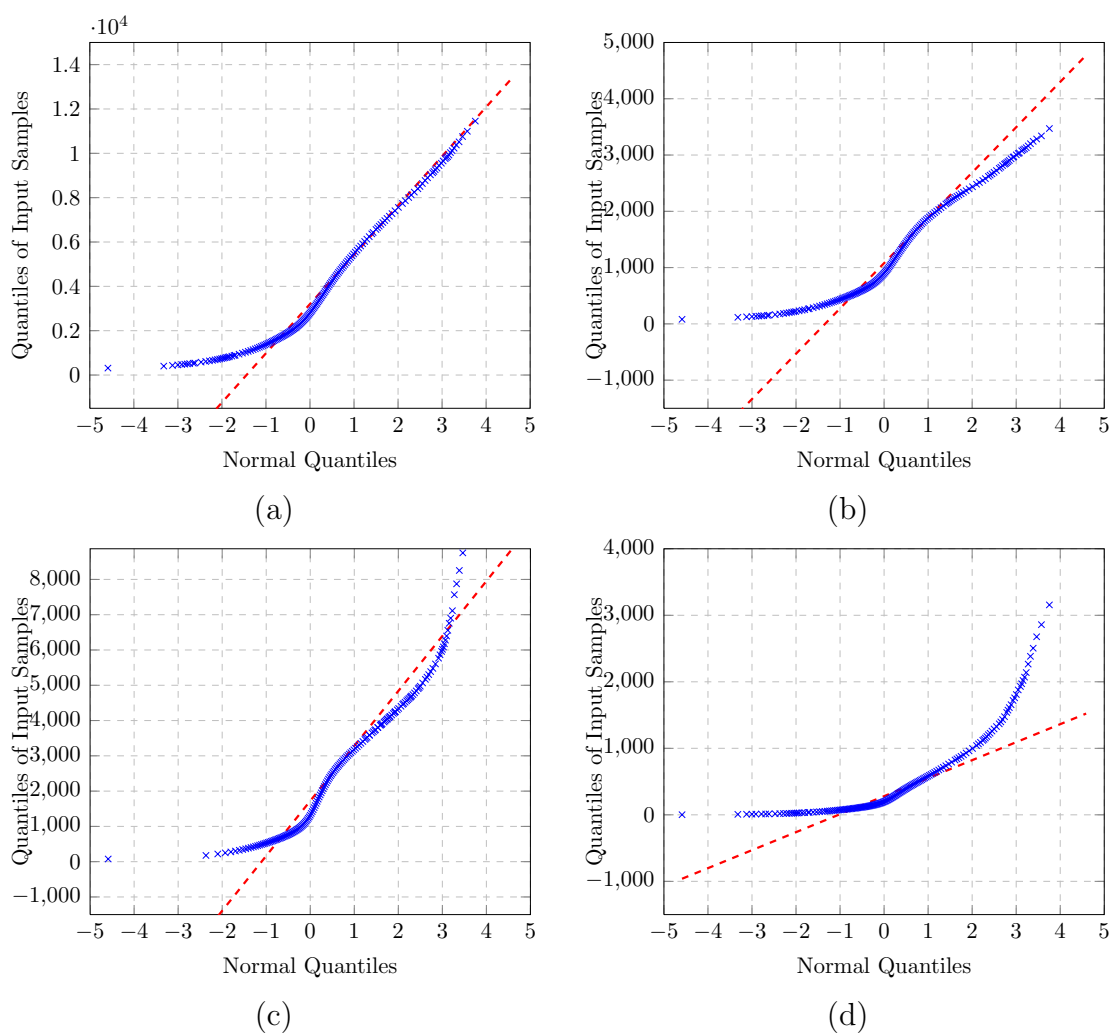


Figure 4.14: Q-Q Plot of data sample versus the Normal theoretical distribution.

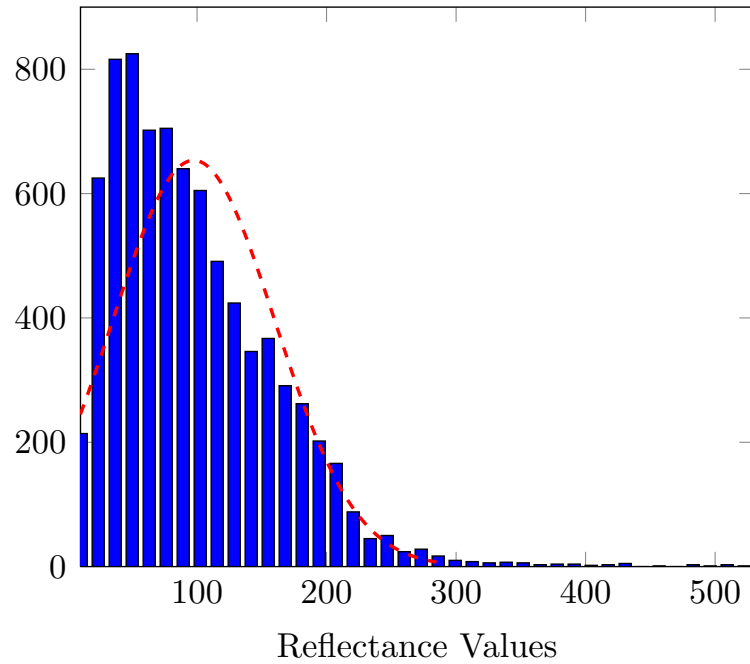


Figure 4.15: Histogram of the data sample and comparison with the Gaussian probability density function.

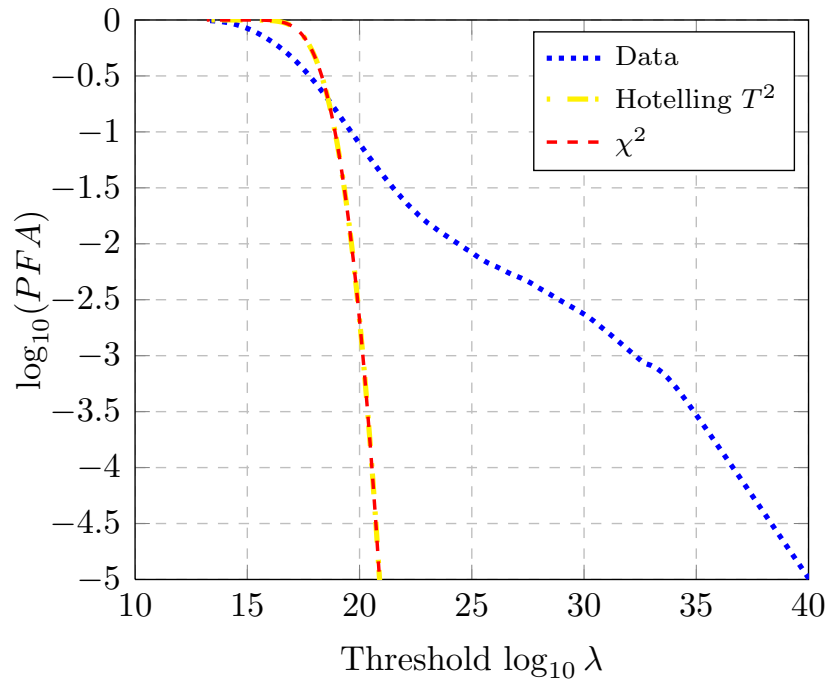


Figure 4.16: Complementary CDF of the Mahalanobis distance  $\mathbb{P}(T^2 > \lambda)$  for the Hymap image (blue) and the theoretical relationship for multivariate Normal distribution (yellow and red).

### 4.3.1 False Alarm Regulation

Again, since hyperspectral data are real and positive, we propose to use a Hilbert filter in order to render them complex to match the target detector distribution. To avoid the well-known problem due to high dimensionality, we have chosen sequentially eight bands in the complex representation. In this approach, both covariance matrix and mean vector are estimated using a sliding window of size  $11 \times 9$ , having  $N = 98$  secondary data.

We show in Fig. 4.17 the outcome of the ANMF detection scheme built with classical SMV-SCM (red curve) and with robust FPE (yellow curve). Moreover, the theoretical “PFA-threshold” relationships are depicted for the SMV-SCM estimates (black curve) and the FPE (green curve). The results obtained on real HSI data on a non-Gaussian distributed region agree with the theoretical relationship derived for the FPE, while the distribution of the detector built with the SMV-SCM do not follow anymore the theoretical “PFA-threshold” relationship in non-Gaussian distributed background. Remark that the distribution of the detector under the null hypothesis does not depend on the underlying material, for the ANMF-FPE. Moreover, the distribution of the detector might divert from the expected behavior when Gaussian estimators are used. This fact emphasizes the interest of robust estimators in target detection for HSI and suggests its use in other HSI applications where covariance matrix and mean vector are unknown and have to be estimated from the background (e.g. classification, unmixing, etc.).

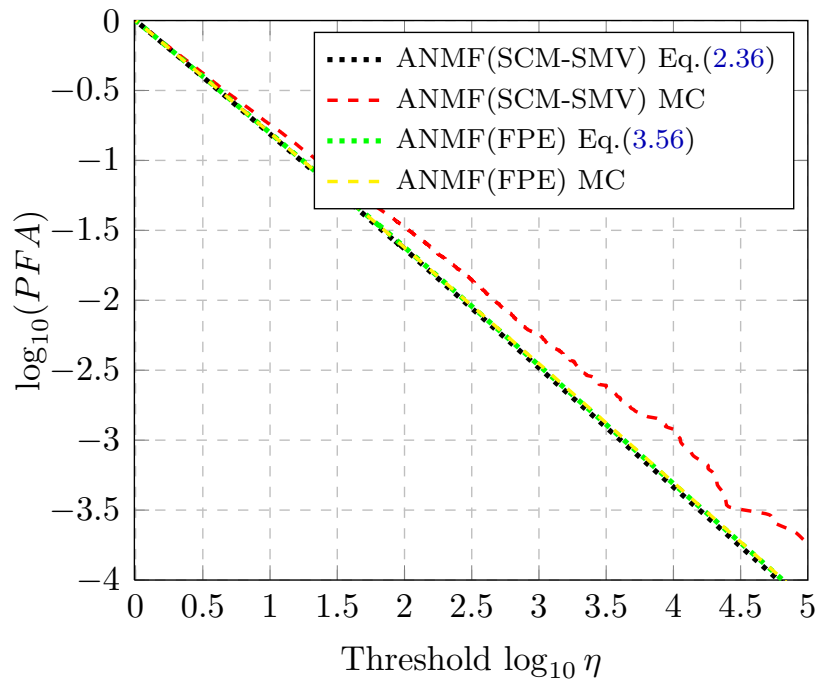


Figure 4.17: PFA versus threshold for the ANMF for real HSI scene and  $m = 8$  and  $N = 98$  when (1) the SCM-SMV are used (red and black curves) (2) the FP estimates are used (yellow and green curves).

### 4.3.2 Anomaly Detection

Let us now present some results for anomaly detection on a real hyperspectral image in which the background cannot be characterized with Gaussian distribution and artificial targets were introduced as anomalies.

#### 4.3.2.1 Robust Kelly Anomaly Detector

The original data set consists on  $50 \times 50$  pixels with 126 bands. This set corresponds to the top-left region from the scene in Fig. 4.13 from which we have chosen sequentially nine bands. For the artificial targets with known spectral signature introduced as anomalies in the background we used the spectra of Fig. 4.19 (a) as endmember. It characterizes the tissue material from Fig. 4.19 (b) measured in a laboratory and available in the Target Detection Blind test project. Fig. 4.18 shows the spatial position and the shape of the considered targets. For this example, both covariance matrix and mean vector are estimated using a sliding window of size  $9 \times 9$  having  $N = 80$  secondary data.

The results for the  $\Lambda_{Kelly AD}$  built with classical SMV-SCM, FPE and shrinkage estimators are shown in Fig 4.20, the FA is fixed at a value of  $PFA = 0.03$ . Once again, the regularization parameter  $\beta$  has been fixed to  $\beta = 0.8$  for the shrinkage estimators. We have empirically optimized the results over  $\beta$  leading to the chosen value. In this case, FPE and notably shrinkage FPE are capable of locating all the artificial targets and exhibit a lower number of false alarms. This improvement is due to the fact that FPE treat the outliers and impulsive samples in order for them to have a smaller contribution to the background characterization process, while the SMV-SCM estimators (and its respective diagonal loaded version) suffer from the presence of strong reflectance pixels in the secondary data. Remark that the shrinkage FPE allow for a better detection compared to FPE.

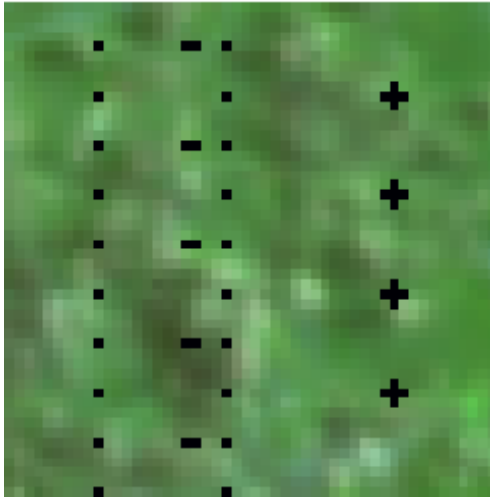
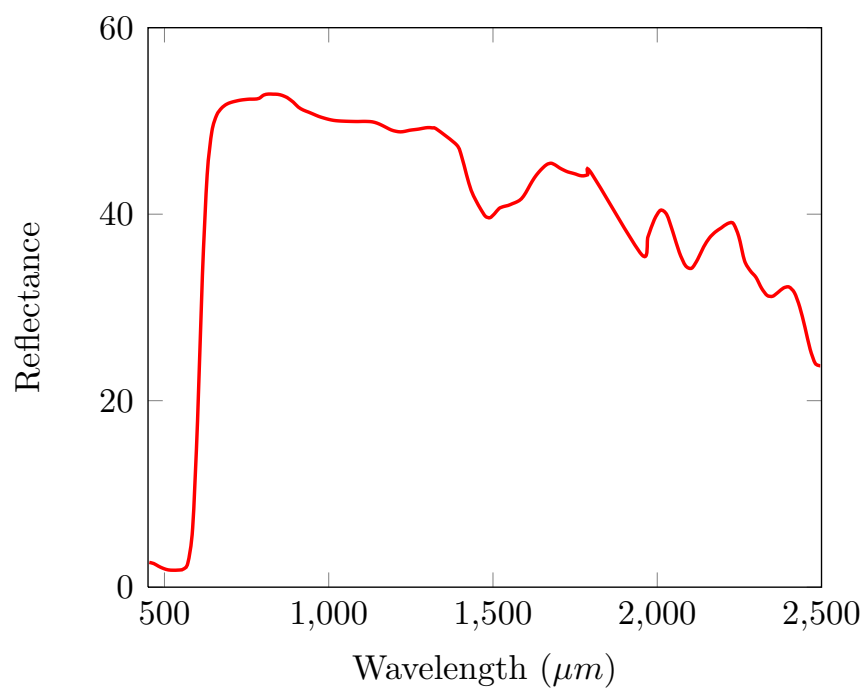
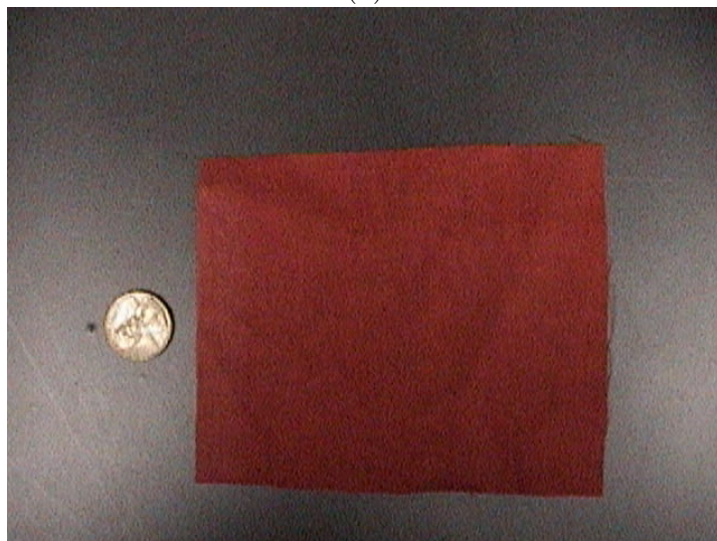


Figure 4.18: Original data set and position of the artificial anomalies.





(a)



(b)

Figure 4.19: (a) Endmember used in the experiment which corresponds to (b) a fabric panel.

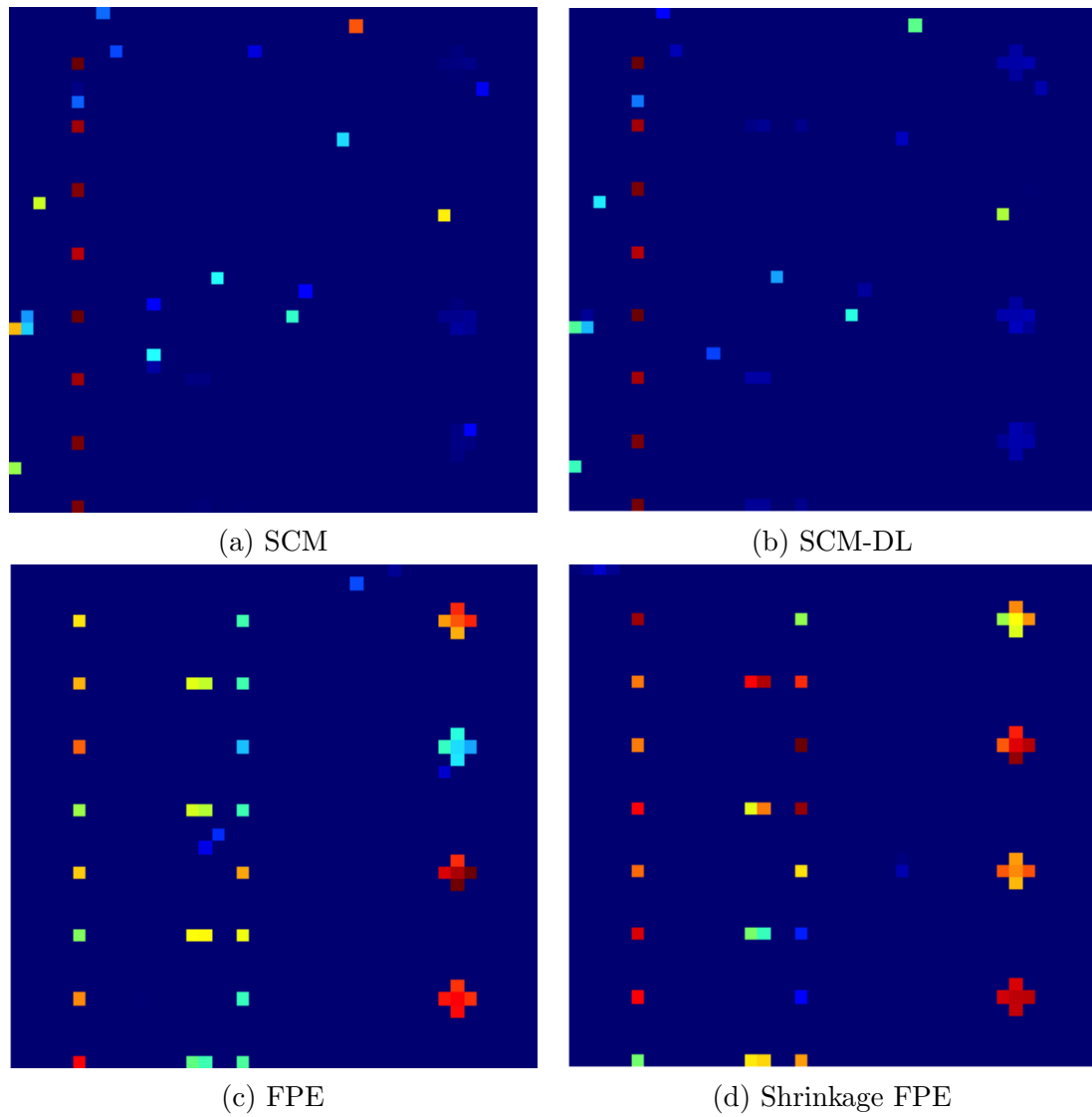


Figure 4.20: Kelly AD built with robust estimates for artificial targets in real HSI.

### 4.3.2.2 Binary Partition Trees

Let us now consider all the real targets embedded in the image. Three civilian vehicles and four small (1m-3m) fabric panels were used as targets, for a total of 129 target pixels in the image. Fig. 4.21 shows the location of the targets in the map.

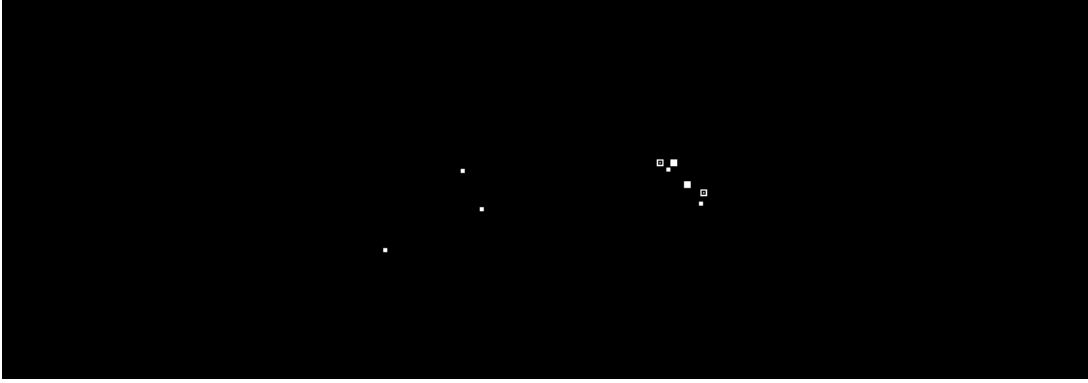


Figure 4.21: Location of the targets in the image.

The output of the Kelly anomaly detector, either using the sliding window or the BPT-based approaches, has been calculated using Eq. (3.60). Moreover, for the sliding window approach, we have studied both SMV-SCM estimates and FPE. For the sliding window approach, the guard window is only the observation vector which is not involved in the parameters estimation and the number of secondary data is  $N = 15 \times 15$ . For the BPT-based approach, the guard node should include the observation vector while the outer node should have at least 196 pixels.

Performances are often compared and summarized by the Receiver Operating Characteristics (ROC) curves, widely used in detection theory. ROC curves plot PD against PFA for different values of threshold Kay (1998). Experimental performance can be evaluated for a specific operating scenario by knowing the locations of the targets of interest within the image. These ratios are computed by averaging the detection threshold  $\lambda$  and counting the number of targets properly detected and the corresponding number of false alarms.

Fig. 4.22 shows the ROC curve comparing the sliding window strategy built with SMV-SCM estimates and the FPE against the BPT-based adaptive Kelly anomaly detector over the dataset. The plot shows that the BPT-based approach outperforms the SMV-SCM sliding window approach in almost all the PFA range. This improvement is due to the selection of more homogeneous regions to perform the estimation stage. Moreover, the results obtained with the FPE are comparable to the BPT approach. However, the computational cost of the BPT strategy is significantly higher than the sliding window approach. Thus, robust estimation procedures allow for good performance while keeping the computational time reduced. It is worth pointing out that the comparative performance gain increases as the number of secondary data (i.e. the size of the window) grows. The fact that all curves are close to the diagonal (bad

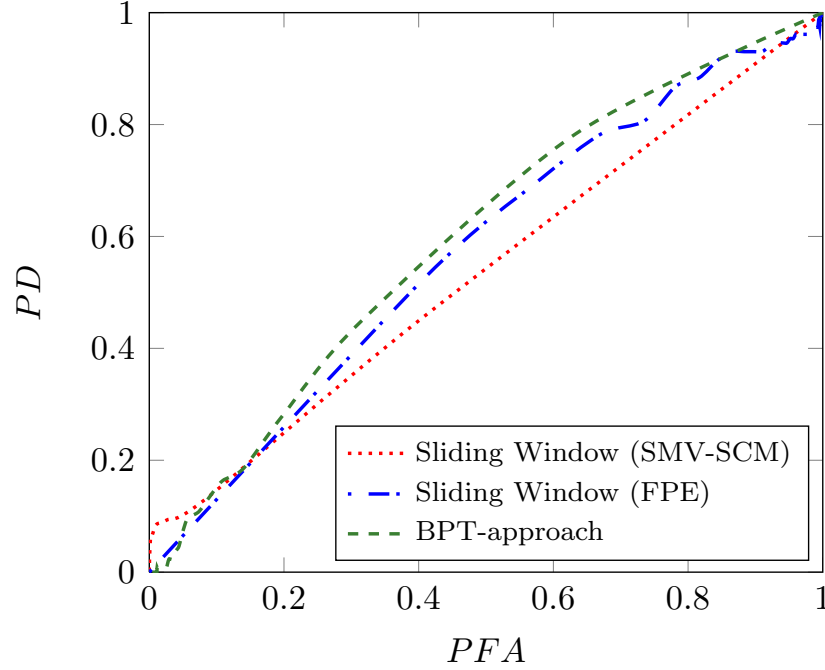
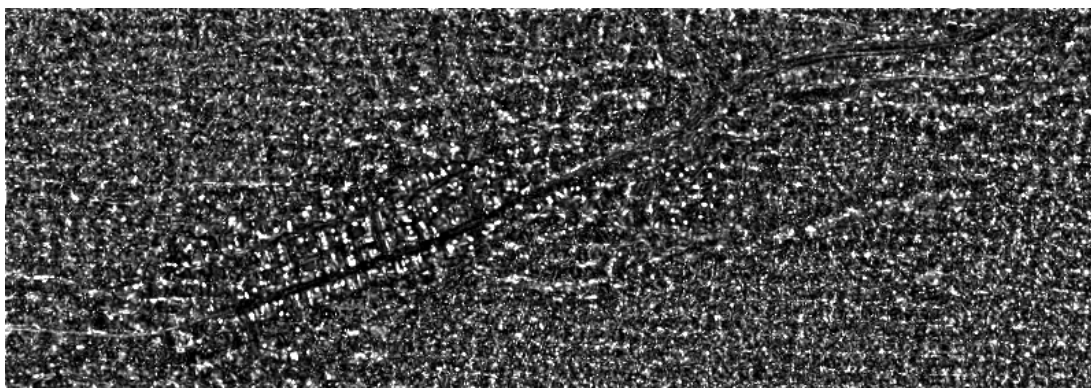


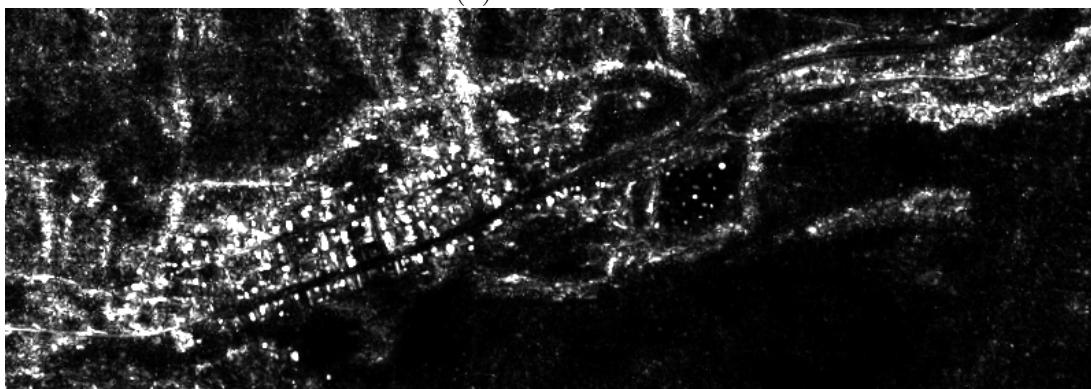
Figure 4.22: PD-PFA plot comparing classical sliding window against BPT strategy

performances, high PFA for small values of the PD) is due to the size of the image. As we have  $280 \times 800$  pixels for only 129 target pixels, it is difficult to keep a small false alarm rate. One could just analyze a smaller region containing the targets, in order to obtain ROC curves closer to the top-left corner (small PFA for high PD values).

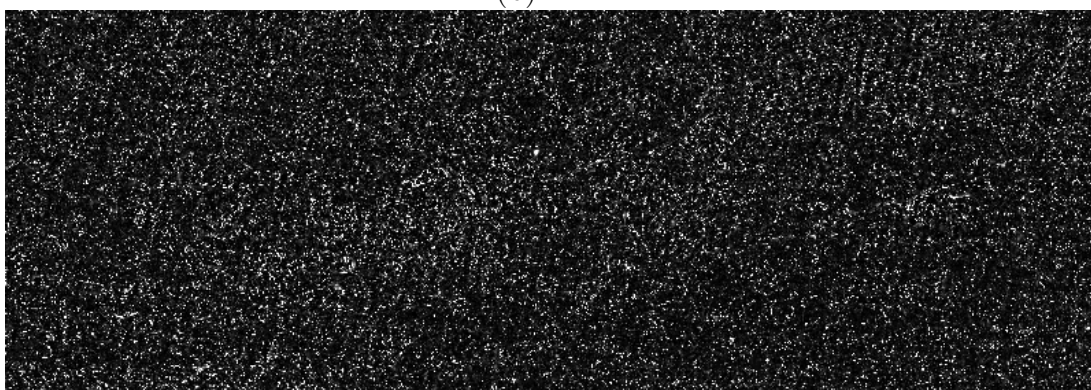
Fig. 4.23 show the detection maps obtained by the Kelly anomaly detector for fixed values of the FA,  $PFA = 0.1$  using the conventional windowing approach, SMV-SCM and FPE, and the proposed BPT-based approach. Remark that for a given PFA, the SMV-SCM solution in Fig. 4.23 (a) provides false detection all over the image. The result is very noisy and the only shape easily observed is the road crossing from the bottom-left of the image to the top-right of the image. This can be due to the use of the sliding window with Gaussian estimators in a non-Gaussian environment. Fig 4.23 (b) illustrates the improvement for real target detection purposes compared to the classical SMV-SCM and to the BPT proposal. In this case, one can locate the intended targets by visual inspection on the center of the image as they are found in an homogenous zone. The buildings corresponding to the urban area are detected as anomalies, as they are different from their background, while the forest on the bottom-right of the image yields an homogenous region. Finally, Fig 4.23 (c) show the results for the BPT-approach. Although, the number of FA is considerable smaller than with SMV-SCM estimators (the white pixels are less concentrated), it is difficult to find an interpretation related to the selection of homogenous regions used to perform the estimation.



(a) SMV-SCM



(b) FPE



(c) BPT

Figure 4.23: Detection maps for sliding window using (a) Gaussian estimates, (b) FPE and (c) BPT-approach

## 4.4 Rural scene

The hyperspectral image analyzed is a Rural scene displayed in Fig. 4.24. The image is constituted of  $742 \times 226$  pixels and 180 spectral bands.

### 4.4.1 False Alarm Regulation

Again, since hyperspectral data are real and positive, we use the Hilbert filter to render data complex for target detector distribution. To reduce dimensionality, and because many of the bands are corrupted by noise, we have chosen sequentially five bands in the complex representation. In this approach, both covariance matrix and mean vector are estimated using a sliding window of size  $19 \times 19$ , having  $N = 360$  secondary data.

Fig. 4.25 depicts the results of the ANMF detector built with classical SMV-SCM (red curve) and with robust FPE (yellow curve). Moreover, the theoretical “PFA-threshold” relationship are plotted for the SMV-SCM estimates (black curve) and the FPE (green curve). The results obtained on real HSI data on a non-Gaussian distributed region agree with the theoretical relationships derived for the FPE, while the distribution of the detector built with the SMV-SCM divert greatly from the expected behavior.

### 4.4.2 Anomaly Detection

We present now some results on anomaly detection. The results for the  $\Lambda_{Kelly AD}$  built with classical SMV-SCM and FPE are shown in Fig 4.26, the FA is fixed at a value of  $PFA = 0.01$ . For the outcome of the detector built with the SMV-SCM, firstly note that, detected pixels are scattered over the whole image. The edges of the corresponding roads can be appreciated. It is worth point out that the vertical lines on the image correspond to misalignments originated in the spectrometer. On the other hand, the results obtained with the FPE allow to distinguish more easily different parts of the images. The line on the top of the image is sharply delineated and the edges of the roads and even of the trees are enhanced. Moreover, on the center-left of the image, one can observe a homogenous part which corresponds to the smooth green part in the true color image. The presence of a strong anomaly on the bottom-left of the image is more clearly detected with the FPE. Finally, as the misalignment lines are found regularly over the image, the detection test built with the FPE only detects one of the lines which is stronger in power to the others. Interestingly these lines cannot be visually identified in the true color image.



Figure 4.24: True color composition of the Rural scene.

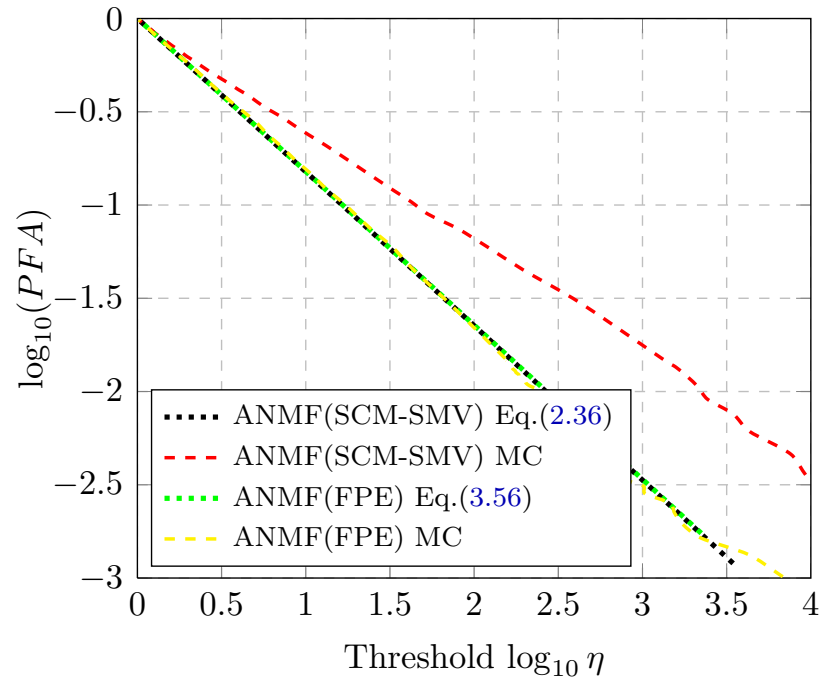


Figure 4.25: PFA versus threshold for the ANMF for real HSI scene and  $m = 8$  and  $N = 98$  when (1) the SCM-SMV are used (red and black curves) (2) the FP estimates are used (yellow and green curves).



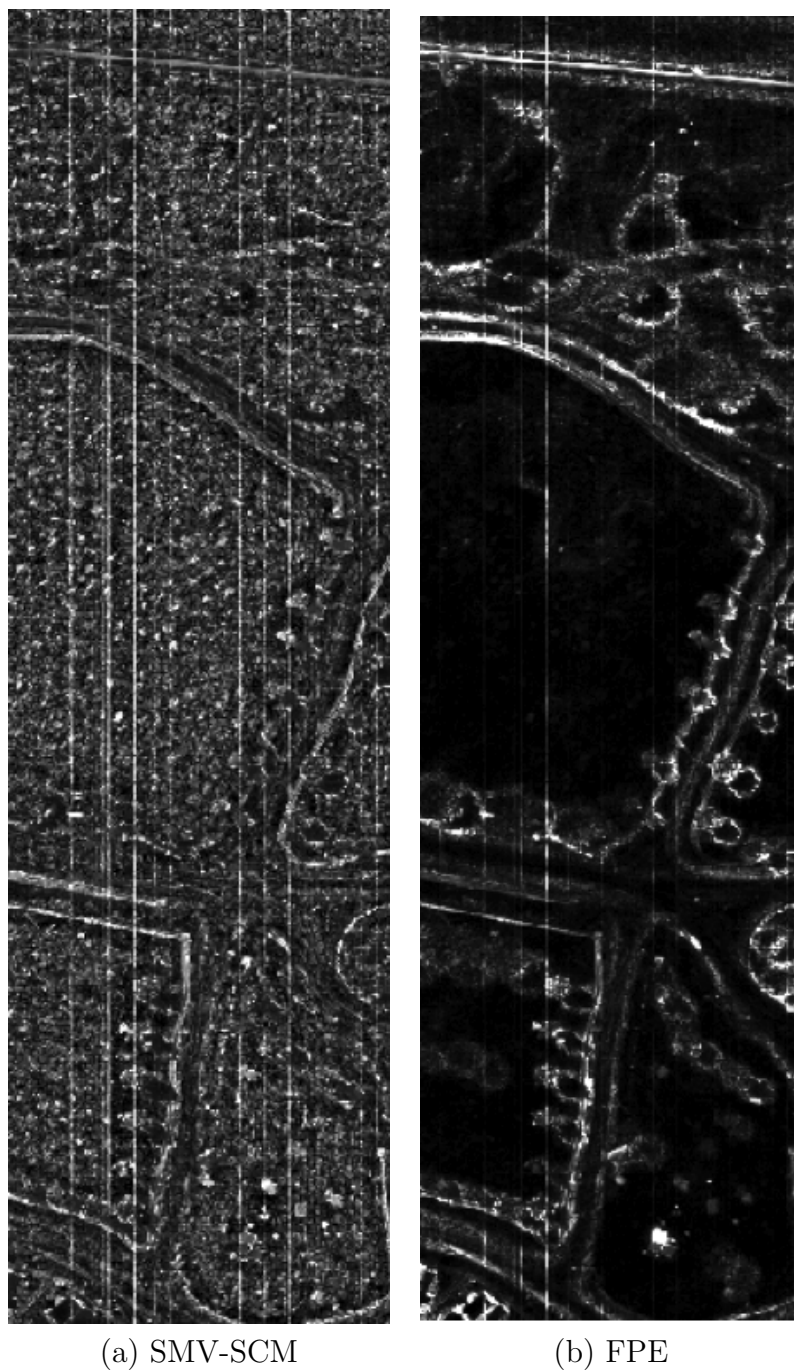


Figure 4.26: Kelly AD built with (a) Gaussian SMV-SCM and (b) robust FPE.

## 4.5 MUSE project

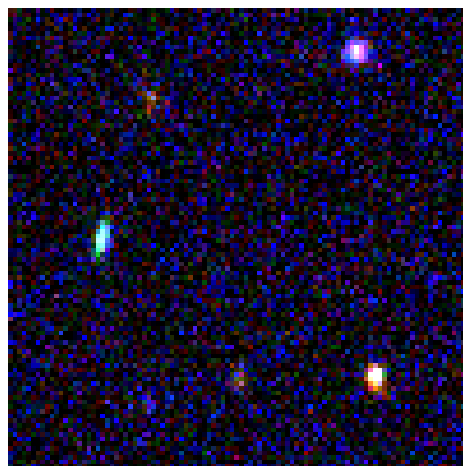
The robust anomaly detector has also been applied for galaxy detection on the MUSE data cube. The Multi Unit Spectroscopic Explorer (MUSE) project <sup>1</sup> aims to provide astronomers with a new generation of optical instrument, capable of simultaneously imaging the sky (in 2D) and measuring the optical spectra of the light received at a given position on the sky. MUSE was installed on the VLT telescope and operational in 2013, and its performances are expected to allow observation of far galaxies up to 100 times fainter than those presently detectable. MUSE will deliver a 3D data-cube made of a stack of images recorded at 3578 different wavelengths over the range 465-930 nm. Each monochromatic image represents a field of view of  $60 \times 60$  arcsec, recorded with a spatial sampling of 0.2 arcsec. Each record results in a data cube of size 1570 MB encoding 3578 images of  $300 \times 300$  pixels, possibly containing thousands of objects (galaxies) existing over different subsets of wavelengths.

An example of MUSE data cube image is displayed in Fig. 4.27 (a), from the 3578 available bands, we have chosen one band of each 100. The results for anomaly detection are presented in Fig. 4.27 for a fixed value FA  $PFA = 10^{-3}$ . Note that detection with FPE (c) provides results with lower false alarm rate than classical ones (b).

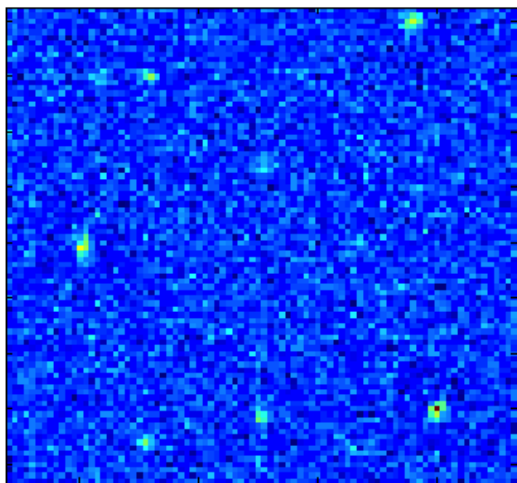
These examples illustrate the robust behavior of FPE in non-Gaussian environments or for close targets detection problems.

---

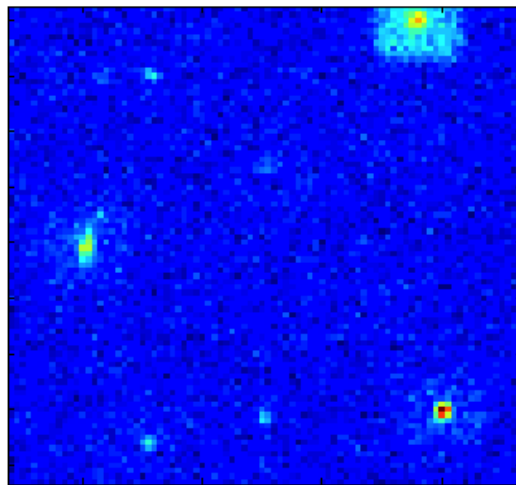
<sup>1</sup><http://muse.univ-lyon1.fr/>



(a) MUSE data cube



(b) SMV-SCM



(c) FP estimates

Figure 4.27: Classical and Fixed-point anomaly detection in a hyperspectral image of  $300 \times 300$  in 3578 channels.

## 4.6 Summary

Finally, the theoretical framework described in Gaussian and non-Gaussian environments have been investigated for real hyperspectral data. The detection schemes and the closed-form expressions derived for false-alarm regulation have been validated through several hyperspectral images. Moreover, the improvement brought by robust estimation methods is exemplified in terms of probability of detection over real target detection problems. Ultimately false-alarm regulation is reached in most real scenarios due to the robustness and invariance of the proposed techniques.



# Conclusion

## Summary

Target detection and anomaly detection of multidimensional signals have proved to be valuable techniques in a wide range of applications, including search-and-rescue, surveillance, rare mineral and land mines detection, etc. Target detection aims to discover the presence of a specific signal of interest among a set of signals. Statistical target detection is based on the Neyman-Pearson criterion, which maximizes the probability of detection for a given probability of false alarm. Anomaly detection is a special case of target detection in which no a-priori target is provided. Hence, the goal of anomaly detection is to detect signals that are anomalous respect to the background. However, the performance of the classical detection methods strongly relies on the statistical parameters estimation. Accordingly, when the background is non-homogeneous or the noise independence assumption is not fulfilled, the detector performance can be deteriorated. This work intends to provide a unified framework for different target detection problems in most scenarios.

Classical target detection schemes are usually obtained deriving the likelihood ratio under Gaussian hypothesis and replacing the unknown background parameters by their estimates. In most applications, interference signals are assumed to be Gaussian with zero mean or with a known mean vector that can be removed and with unknown covariance matrix. When the mean vector is unknown, it has to be jointly estimated with the covariance matrix, as it is the case for instance in hyperspectral imaging. In this work, the adaptive versions of the classical Matched Filter and the Normalized Matched Filter, as well as two versions of the Kelly detector are first derived and then are analyzed for the case when the mean vector of the background is unknown. More precisely, theoretical closed-form expressions for false-alarm regulation are derived and the CFAR property is pursued to allow the detector to be independent of nuisance parameters. It is worth pointing out that closed-form expressions are essential to automatically set the detector's threshold. Otherwise, numerical methods or Monte-Carlo simulations have to be used leading to less accurate results.

This thesis also proposes a new kind of robust detectors allowing to get over non-Gaussianity and heterogeneity of real Hyperspectral data. When accounting for heterogeneity and non-Gaussianity, elliptically symmetric distributions provide reliable models for background characterization. Through this non-Gaussian assumption,

this work highlights the fact that robust estimation procedures are an interesting alternative to classical Gaussian estimators.  $M$ -estimators for the mean vector and the scatter matrix are described. In particular, *Huber's*  $M$ -estimators, Student  $M$ -estimators and FPE are detailed. The performances of robust estimators have been studied. This analysis reveals the superiority of  $M$ -estimators in non-Gaussian environment. Moreover, the loss in performance is quantified for Gaussian distributed background. The goal of this thesis is then not only to recall well-known methodologies for target detection but also to propose ways to extend them to non-Gaussian framework. Furthermore, theoretical closed-form expressions for false-alarm regulation are again derived and CFAR property reached within the class of elliptical distributions.

Anomaly detection methods are used for target detection in which no a priori information about the spectra of the targets of interest is available. Some classical anomaly detection schemes are reviewed such as the widely spread Reed-Xiaoli Detector, the Kelly-based anomaly detector and some of its variations. Moreover, the Mahalanobis distance based detector, rigorously derived from a Kelly's test-based approach, is analyzed and its exact distribution is derived when both mean vector and covariance matrix are unknown and have to be estimated. Although, most of these techniques are based on Gaussian distribution, we have also propose ways to extend them to non-Gaussian framework. For this purpose, we show that using robust estimators as plug-in estimators in anomaly detectors leads to some great improvement in the detection process. Notably, the use of the FPE allow for a higher probability of detection for a fixed value of the  $PFA$ .

Once the theoretical framework has been established for both target detection and anomaly detection, the proposed methods have been investigated for four real hyperspectral images. The false alarm regulation expressions derived under Gaussian assumption are validated in experimental data, once the Gaussianity of an extracted region has been accepted. In addition, detection schemes have been evaluated in terms of probability of detection for target detection methods and detection maps are provided for anomaly detection purposes. In most scenarios, HSI cannot be characterized only through Gaussian distribution. Thereupon, robust  $M$ -estimators are considered for parameters estimation and the improvement brought by robust estimation methods is illustrated over real target detection problems. Ultimately false-alarm regulation is reached in most real scenarios due to the robustness and invariance of the proposed techniques.

Generally, this work finds its purpose in signal processing methods for which both mean vector and covariance matrix are unknown.

## Perspectives

The analysis detailed for false-alarm regulation could be extended to the target present hypothesis. When including in the derivation of the detector's distribution the  $SNR$ , theoretically, one could obtain the corresponding expressions for the probability of detection. However, such techniques rarely lead to closed-form equations and the resulting threshold is usually computed numerically.

One of the majors challenges encountered in hyperspectral imaging is the high-dimensionality of data. In addition, the proposed estimation methods for the covariance matrix and the mean vector suffer for some limitations for high-dimensional vectors. Indeed, the robustness of the  $M$ -estimators decreases with the number of bands. One solution considered in this work is the use of Shrinkage FPE more appropriate than conventional FPE. However, other solution can be further explored.

One could use low rank techniques for adaptive target detection [Kirsteins and Tufts \(1994\)](#). Studies carried on hyperspectral data show that the background information often lies on subspace lower than the number of spectral bands. Thus, instead of a covariance matrix estimator, it would be appropriate to use an estimate of the projector for the orthogonal to the background space [Rangaswamy et al. \(2004\)](#). To follow this idea of low rank, using a projector presents the drawback of renouncing to much of the information contained in the covariance matrix. Thus, only part of the eigenvectors are used, excluding any information provided by the small eigenvalues. This treatment reduces the number of samples needed for the estimation, it also implies a loss of performance. Therefore, it may be advantageous to study the resulting projectors built with FPE and  $M$ -estimators, in order to obtain robust projectors within the class of elliptical distribution. These ideas are being currently explored in the work of A. Breloy [Breloy et al. \(2014\)](#) applied to radar target detection.

Related to the high-dimensionality problem, authors have proposed many dimensionality reduction methods. One of the most popular is the Principal Component Analysis (PCA) which involves the singular value decomposition of the covariance matrix. This operation can be interpreted as the opposite process to anomaly detectors, in particular, the studied Mahalanobis based detector. In anomaly detection techniques, one is interested in rare objects usually lying in the space associated to small eigenvalues, while principal component analysis tries to identify the dominant directions and dismiss those directions associated to small eigenvalues. Thus, it would be interesting to explore the use of robust  $M$ -estimators for principal component analysis in dimensionality reduction. This techniques will probably lead to robust methods for feature extraction. Any other techniques found in HSI that requires the estimation of the mean vector or covariance matrix could be revisited through elliptical distributions and robust estimation procedures.

Further, methods of random matrix theory could be investigated. These are suitable for analyzing high-dimensional covariance matrix estimates, i.e. given a small



sample size compared to the number of dimensions. Classical results of random matrix theory [Edelman and Rao \(2005\)](#) fail if the sample covariance matrix is used in the context of elliptically distributions and heavy tailed data. The Marčenko-Pastur law can be used to analyze the eigenspectrum of high-dimensional distributions [Marčenko and Pastur \(1967\)](#). The key idea is that this can be done even if the sample size is small compared to the number of dimensions whereas classical statistical analysis fails in that case. Moreover, the eigenspectrum can be analyzed even if the number of data points falls below the number of dimensions. Thereupon, the use of robust estimators in classical random matrix theory investigated in [Couillet et al. \(2012, 2011a,b,c\)](#) could be explored for hyperspectral data processing.



A class of robust estimates for detection in  
hyperspectral images using elliptical  
distributions background.



# A CLASS OF ROBUST ESTIMATES FOR DETECTION IN HYPERSPECTRAL IMAGES USING ELLIPTICAL DISTRIBUTIONS BACKGROUND

*J. Frontera-Pons<sup>1</sup>, M. Mahot<sup>1</sup>, J.P. Ovarlez<sup>1,3</sup>, F. Pascal<sup>1</sup>, S.K. Pang<sup>2</sup>, J. Chanussot<sup>4</sup>*

## ABSTRACT

When dealing with impulsive background echoes, Gaussian model is no longer pertinent. We study in this paper the class of elliptically contoured (EC) distributions. They provide a multivariate location-scatter family of distributions that primarily serve as long tailed alternatives to the multivariate normal model. They are proven to represent a more accurate characterization of HSI data than models based on the multivariate Gaussian assumption. For data in  $\mathbb{R}^k$ , robust proposals for the sample covariance estimate are the  $M$ -estimators. We have also analyzed the performance of an adaptive non-Gaussian detector built with these improved estimators. Constant False Alarm Rate (CFAR) is pursued to allow the detector independence of nuisance parameters and false alarm regulation.

**Index Terms**— hyperspectral imaging, target detection, elliptical distributions,  $M$ -estimators

## 1. INTRODUCTION

Anomaly detection and detection of targets or activity such as chemical plumes, aerosols, vehicles, anomalous targets, arise in many military and civilian applications [1]. Hyperspectral imaging sensors provide 2D spatial image data containing spectral information. This information can be used to address such detection tasks. Hyperspectral imaging sensors measure the radiance or reflectance of the materials within each pixel area at a very large number spectral wavelength bands.

It is often assumed that signals, interferences, noises, background are Gaussian stochastic processes. Indeed, this assumption makes sense in many applications. In these contexts, Gaussian models have been widely investigated in the framework of Statistical Estimation and Detection Theory. They have led to appealing and well known algorithms as for instance the Matched Filter and its adaptive variants in radar detection [2, 3]. The mathematical framework for the design and evaluation of detection algorithms is provided by the well known binary hypothesis testing procedure. The basic problem of detecting a complex signal corrupted by an additive noise  $\mathbf{c}$  in a  $m$ -dimensional complex vector  $\mathbf{y}$  can be stated as

a binary hypothesis test with two competing hypotheses  $H_0$  and  $H_1$ . In practice the  $\mathbf{c}$  background statistics are unknown and have to be estimated from  $K$  secondary data  $\mathbf{c}_i$ 's. Under hypothesis  $H_1$ , it is assumed that the observed data  $\mathbf{y}$  consists in the sum of a signal  $\mathbf{s} = \alpha \mathbf{p}$  and background noise  $\mathbf{c}$ , where  $\mathbf{p}$  is a perfectly known complex steering vector (characterizing for example the spectral material to detect) and  $\alpha$  is the signal amplitude.

Generally, the statistical parameters (covariance matrix  $\mathbf{M}$ , mean  $\boldsymbol{\mu}$ , ...) of the background can be estimated by using all pixels within an area of interest. The size of the area has to be chosen large enough to ensure the invertibility of the covariance matrix and small enough to justify both spectral homogeneity (stationarity) and spatial homogeneity. In hyperspectral imaging, the actual response of a detector to the background pixels differs from the theoretically predicted distribution for Gaussian backgrounds. In fact, as stated in [5], the empirical distribution usually has heavier tails compared to the theoretical distribution, and these tails strongly influence the observed false-alarm rate of the detector. Since the two hypotheses contain unknown parameters (for example, the covariance matrix of the background) that have to be estimated from the data, the detector has to be adaptive, and it is usually designed by using the Generalized-Likelihood-Ratio Test (GLRT) approach.

One of the most general and elegant impulsive noise model often used in radar detection schemes is provided by the so-called *Spherically Invariant Random Vectors* (SIRV). A SIRV  $\mathbf{y}$  is a compound process, it is the product of a Gaussian multivariate process  $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{M})$  and the square root of a non-negative random scalar variable  $\tau$  called the *texture*. Thus, the SIRV is fully characterized by the texture (representing an unknown intensity) and the unknown covariance matrix of the Gaussian vector. Another statistical framework is based on the use of Elliptical Random Process which generalizes the SIRV processes. In that framework, the multidimensional vector  $\mathbf{x}$  (supposed to be Gaussian in SIRV theory) is here uniformly distributed on the hyper-sphere. That means that a SIRV is also elliptically distributed

One of the major challenging difficulties in SIRV or Spherically distributed random process modeling, is to estimate these two unknown quantities. For example, the classical Sample Covariance Matrix used in adaptive detection in Gaussian noise is not at all the best estimate and does not

<sup>1</sup> : SONDRRA Research Alliance, Supélec, France, <sup>2</sup> : DSO National Laboratories, Singapore, <sup>3</sup> : French Aerospace Lab, ONERA DEMR/TSI, France, <sup>4</sup> : GIPSA-LAB, Grenoble Institute of Technology, France

correspond to the Maximum Likelihood estimator. In SIRV context, these problems have been investigated in [6] for the texture estimation while [7] and [8] have proposed different estimates for the covariance matrix. A complete statistical analysis of these covariance matrix estimates has been realized in [9]. For Elliptical process, the estimation of the mean and covariance is known as  $M$ -estimation theory introduced by [10, 11].

## 2. ELLIPTICAL DISTRIBUTION

A  $m$ -dimensional random complex vector  $\mathbf{y}$  is said to have a complex elliptical distribution if its probability density function (PDF) has the form

$$f_{\mathbf{y}}(\mathbf{y}) = |\Sigma|^{-1} h_{\mathbf{y}}((\mathbf{y} - \boldsymbol{\mu})^H \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu})), \quad (1)$$

where  $H$  denotes the conjugate transpose operator and  $h_{\mathbf{y}} : [0, \infty) \rightarrow [0, \infty)$  is any function such as (1) defines a PDF,  $\boldsymbol{\mu}$  is the mean vector and  $\Sigma$  is the scatter matrix. The function  $h_{\mathbf{y}}$ , usually called density generator, is assumed to be only approximately known. Note that it produces density contours corresponding to elliptical surfaces. If the second moments exist, then  $\Sigma$  reflects the structure of the covariance matrix of the elliptically distributed random vector  $\mathbf{y}$ , i.e. the covariance matrix equates the scatter matrix up to a scaling constant. We shall denote this complex elliptical distribution by  $EC(\boldsymbol{\mu}, \Sigma, h)$ . It is worth pointing that the EC class includes an infinity of distributions, notably the Gaussian one, multivariate  $t$  distribution or multivariate Cauchy.

## 3. M- ESTIMATORS

Let  $(\mathbf{c}_1, \dots, \mathbf{c}_K)$  be a  $K$ -sample of  $m$ -dimensional complex independent vectors with  $\mathbf{c}_i \sim EC(\boldsymbol{\mu}, \Sigma, h)$ ,  $i = 1, \dots, K$ . The complex  $M$ -estimators of location and scatter are defined as the joint solutions to the estimating equations:

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \frac{\sum_{n=1}^K u_1(t_n) \mathbf{c}_n}{\sum_{n=1}^K u_1(t_n)} \\ \hat{\mathbf{M}} &= \frac{1}{K} \sum_{n=1}^K u_2(t_n^2) (\mathbf{c}_n - \hat{\boldsymbol{\mu}})(\mathbf{c}_n - \hat{\boldsymbol{\mu}})^H \end{aligned} \quad (2)$$

where  $t_n = ((\mathbf{c}_n - \hat{\boldsymbol{\mu}})^H \hat{\mathbf{M}}^{-1} (\mathbf{c}_n - \hat{\boldsymbol{\mu}}))^{1/2}$  and  $u_1, u_2$  are two weighting functions on the quadratic form  $t_n$ . Note  $t_n^2$  is in fact, the widely used Mahalanobis distance.  $M$ -estimators have first been studied in the real case, defined as solution of (2) with real samples. Existence and uniqueness have been proven in the real case, provided functions  $u_1, u_2$  satisfy a set of general assumptions stated by Maronna [11]. Olilla has shown in [12] that these conditions hold also in the complex

case.  $M$ -estimators are particularly suited for estimating the mean vector and the scatter matrix of an elliptical population. When dealing with heavy tailed clutter models, as in HSI, the use of robust estimates decreases the impact of highly impulsive samples and possible outliers.

Remark that if  $u_1$  and  $u_2$  are chosen to be constant and equal to one, the arising estimators correspond to the Sample Mean Vector and Sample Covariance Matrix respectively. They are indeed the Maximum Likelihood estimators when Gaussian distributions are considered.

We state below two particular estimates belonging to the family of  $M$ -estimators. Besides the indicated statistical robustness, they involve some CFAR properties useful for detection issues.

### 3.1. The Fixed Point estimates

According to the Fixed point approach, the joint estimation of  $\mathbf{M}$  and  $\boldsymbol{\mu}$  leads to [13]:

$$\hat{\mathbf{M}}_{FP} = \frac{m}{K} \sum_{k=1}^K \frac{(\mathbf{c}_k - \hat{\boldsymbol{\mu}})(\mathbf{c}_k - \hat{\boldsymbol{\mu}})^H}{(\mathbf{c}_k - \hat{\boldsymbol{\mu}})^H \hat{\mathbf{M}}_{FP}^{-1} (\mathbf{c}_k - \hat{\boldsymbol{\mu}})} \quad (3)$$

and

$$\hat{\boldsymbol{\mu}} = \frac{\sum_{k=1}^K \frac{\mathbf{c}_k}{(\mathbf{c}_k - \hat{\boldsymbol{\mu}})^H \hat{\mathbf{M}}_{FP}^{-1} (\mathbf{c}_k - \hat{\boldsymbol{\mu}})}}{\sum_{k=1}^K \frac{1}{(\mathbf{c}_k - \hat{\boldsymbol{\mu}})^H \hat{\mathbf{M}}_{FP}^{-1} (\mathbf{c}_k - \hat{\boldsymbol{\mu}})}} \quad (4)$$

Obtained when choosing  $u_1(s) = s^{-1}$  and  $u_2(s) = ms^{-1}$ . For the matrix estimate, existence and uniqueness have been established in [14]. Although the proof for simultaneous scatter and location estimates is still an open question, they have been found to be useful and reliable for elliptical contours estimation parameters because of its easy implementation. They are specified by implicit equations and can be easily computed using a recursive algorithm. We refer to [9] for a detailed performance analysis of the Fixed Point covariance matrix estimate. The main results of the statistical properties of the  $\hat{\mathbf{M}}_{FP}$  are summarized:  $\hat{\mathbf{M}}_{FP}$  is a consistent and unbiased estimate of  $\mathbf{M}$ ; its asymptotic distribution is Gaussian and its covariance matrix is fully characterized in [15]; its asymptotic distribution is the same as the asymptotic distribution of a Wishart matrix with  $mK/(m+1)$  degrees of freedom.

### 3.2. The Huber's $M$ -estimates

Using the well-known Huber's  $\psi$  function [10] defined as,

$$\psi_k(s) = \min(s, k) \quad (5)$$

with  $k > 0$ . One can obtain Huber's  $M$ -estimator by taking  $u_1(s) = \psi_k(s)/s$  and  $u_2(s) = \psi_{k^2}(s)/s$ . We remark that the Huber function can be seen as a mix between the Fixed Point

estimate and the conventional SCM estimate. Extreme values of  $t_n^2$  outside  $[0, k^2]$  are strongly attenuated by the  $1/s$  decreasing function (as for the Fixed Point) while normal values below  $k^2$  are uniformly kept (SCM behavior).

#### 4. THE ANMF BUILT WITH THE M-ESTIMATORS

Different types of adaptive non-Gaussian detectors were derived for target enhancement purposes in radar applications. We focus here on the study of the GLRT-Linear Quadratic [16], also known as Adaptive Cosine Estimate,

$$\Lambda(\hat{\mathbf{M}}, \hat{\boldsymbol{\mu}}) = \frac{|\mathbf{p}^H \hat{\mathbf{M}}^{-1}(\mathbf{y} - \hat{\boldsymbol{\mu}})|^2}{(\mathbf{p}^H \hat{\mathbf{M}}^{-1} \mathbf{p})(\mathbf{y} - \hat{\boldsymbol{\mu}})^H \hat{\mathbf{M}}^{-1}(\mathbf{y} - \hat{\boldsymbol{\mu}})} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda \quad (6)$$

where  $\mathbf{p}$  is the spectral steering vector,  $\mathbf{y}$  the cell under test and  $\lambda$  the decision threshold. Note that the mean  $\hat{\boldsymbol{\mu}}$  is generally omitted in radar detection (and therefore not estimated) as the noise is always zero mean. So, in hyperspectral imaging, as the data represent intensity values and are positive, we need to estimate it, jointly with the covariance matrix  $\mathbf{M}$ . Used with the Fixed Point estimates, this detector has essential CFAR properties like texture-CFAR (independent of density generator function), matrix-CFAR (independent of  $\mathbf{M}$ ) and mean-CFAR. Hence, the detector GLRT  $\Lambda(\hat{\mathbf{M}}_{FP}, \hat{\boldsymbol{\mu}})$  behaves according to the same distribution regardless of the elliptical distribution used and for different covariance matrices. This is of a major interest in areas of background transition, like for example, in coastal areas (ground and sea) or at the edge of forests (fields and trees) because the detector resulting distribution should be insensitive to the different clutter areas.

##### 4.1. Detector performance

The performance analysis has been realized over the data set provided by DSO National Laboratories, the normalized hypercube is shown in figure 1. The resulting ROC curves (Receiver Operating Characteristic) compare the output of the detector built with the Fixed Point estimates, the Huber  $M$ -estimators and the classical SCM. The test conducted consists in placing an artificial target with a fixed SNR through each pixel of the image. For all the possible threshold values, both probability of false alarm and probability of detection are computed. The outcome is illustrated in figure 2.

These preliminary results show the improvement in performance introduced by the use of  $M$ -estimators regarding the conventional SCM. The desired robustness properties previously mentioned lead to a higher  $P_d$  for small values of the  $P_{fa}$ .

##### 4.2. False Alarm Regulation

The ANMF test statistics distribution is well-known for zero-mean Gaussian model. When  $\mathbf{M}$  is estimated according to the SCM, it follows a complex Wishart distribution  $\mathcal{CW}(K, m; \mathbf{M})$ . Taking into account that the Fixed Point

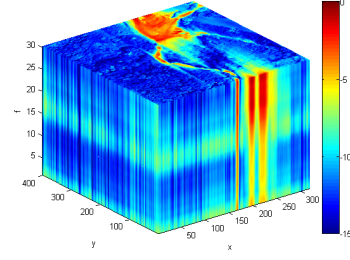


Fig. 1. Normalized data set.

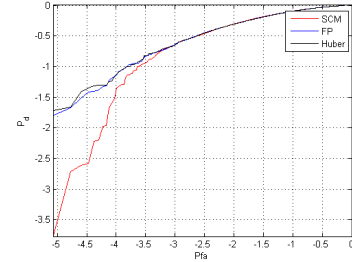


Fig. 2. ROC curves depicting the performance of the detector built with the SCM (in red), the Fixed Point (in blue) and the Huber type (in black) estimates. Probabilities are given in  $\log_{10}$  scale.

Matrix asymptotic distribution is the same as the asymptotic distribution of a Wishart matrix with  $mK/(m+1)$  degrees of freedom. A theoretical relationship between the detection threshold  $\lambda$  and the Probability of False Alarm  $P_{fa} = \mathbb{P}(\Lambda > \lambda | H_0)$  has been stated in [17]:

$$P_{fa} = (1 - \lambda)^{a-1} {}_2F_1(a, a-1; b-1; \lambda) \quad (7)$$

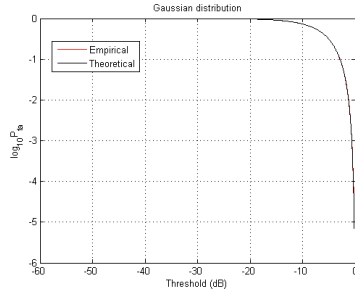
where  $a = \frac{m}{m+1}K - m + 2$ ,  $b = \frac{m}{m+1}K + 2$  and  ${}_2F_1$  is the Hypergeometric function. This expression holds when  $\boldsymbol{\mu}$  is completely known and can be removed from the data. As the joint estimation of  $\mathbf{M}$  and  $\boldsymbol{\mu}$  is needed, (7) is no longer valid and a gap is evidenced between theoretical and empirical curves.

When  $\boldsymbol{\mu}$  is included in the estimation of the SCM,  $\mathbf{M}$  is distributed as a complex Wishart with  $K-1$  degrees of freedom. The theoretical  $P_{fa}$ -threshold relationship for the Gaussian case and SCM estimation has been derived resulting in an expression as in (7) where  $a = (K-1) - m + 2$  and  $b = (K-1) + 2$ . Figure 3 (a) exhibits the regulation of the false alarm for the detector when Gaussian data model is considered.

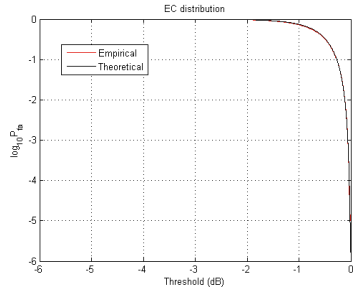
Simulations held in elliptical distributions context show the early empirical distribution for the detector built with the Fixed Point estimates, hinting its curve analytical expression and its CFAR-texture properties. The figure 4 (b) exemplifies

the empirical results for a  $K$ -distribution with shape parameter  $\nu = 0.1$ . Although a more detailed analysis need to be done.

Note that the previous Pfa-threshold has been derived assuming radar data being complex and is not valid for real data. As the hyperspectral data are real and positive, they have been passed through an Hilbert filter to render them complex.



**Fig. 3.**  $P_{fa}$ - threshold relationship for Gaussian model and SCM estimation.



**Fig. 4.**  $P_{fa}$ - threshold relationship for  $K$ -distribution with shape parameter  $\nu = 0.1$  and FP estimation.

## 5. CONCLUSIONS

We consider the family of elliptical contoured distributions for impulsive clutter characterization in hyperspectral imaging. In this context, we study different robust estimators for statistical modeling of the background. Particularly, we describe two  $M$  estimators, so the Fixed Point and the Huber type, pointing its robust behavior. The main contribution of this article is the analysis performed on the ACE detector when built with these newfangled estimates, showing a better performance in probability of detection terms. Finally we introduce a theoretical relationship for false alarm regulation, when covariance matrix and mean are jointly estimated, foremost for the gaussian case. Empirical simulation results are highlighted for elliptical distribution model.

## 6. REFERENCES

[1] S. Matteoli, M. Diani, and G. Corsini, "A tutorial overview of anomaly detection in hyperspectral images," *Aerospace and*

*Electronic Systems Magazine, IEEE*, vol. 25, no. 7, pp. 5–28, 2010.

[2] E. J. Kelly, "An adaptive detection algorithm," *IEEE Trans.-AES*, vol. 23, no. 1, pp. 115–127, November 1986.

[3] S. Kraut, L. L. Scharf, and L. T. Mc Whorter, "Adaptive subspace detectors," *IEEE Trans.-SP*, vol. 49, no. 1, pp. 1–16, January 2001.

[4] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *Signal Processing Magazine, IEEE*, vol. 19, no. 1, pp. 29–43, 2002.

[5] E. Jay, J.-P. Ovarlez, D. Declercq, and P. Duvaut, "Bord: Bayesian optimum radar detector," *Signal Processing*, vol. 83, no. 6, pp. 1151–1162, June 2003.

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

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

[8] F. Pascal, P. Forster, J.-P. Ovarlez, and P. Larzabal, "Performance analysis of covariance matrix estimates in impulsive noise," *IEEE Trans.-SP*, vol. 56, no. 6, pp. 2206–2217, June 2008.

[9] P. J. Huber, "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.

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

[11] E. Ollila and V. Koivunen, "Influence functions for array covariance matrix estimators," *Proc. IEEE Workshop on Statistical Signal Processing (SSP)*, ST Louis, MO, pp. 445–448, October 2003.

[12] D.E. Tyler, "A distribution-free  $m$ -estimator of multivariate scatter," *The Annals of Statistics*, vol. 15, no. 1, pp. 234–251, 1987.

[13] F. Pascal, Y. Chitour, J.-P. Ovarlez, P. Forster, and P. Larzabal, "Covariance structure maximum likelihood estimates in compound gaussian noise : Existence and algorithm analysis," *IEEE Trans.-SP*, vol. 56, no. 1, pp. 34–48, January 2008.

[14] F. Pascal, P. Forster, J.-P. Ovarlez, and P. Larzabal, "Theoretical analysis of an improved covariance matrix estimator in non-gaussian noise," in *Proc. IEEE-ICASSP*, Philadelphia, March 2005, vol. IV, pp. 69–72.

[15] F. Gini, "Sub-optimum coherent radar detection in a mixture of  $k$ -distributed and gaussian clutter," *IEE Proc. Radar, Sonar and Navigation*, vol. 144, no. 1, pp. 39–48, February 1997.

[16] JP Ovarlez, SK Pang, F. Pascal, V. Achard, and TK Ng, "Robust detection using the sirv background modelling for hyperspectral imaging," in *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*. IEEE, 2011, pp. 4316–4319.

B

Adaptive non-Zero Mean Gaussian  
Detection and Application to  
Hyperspectral Imaging





# Adaptive non Zero-Mean Gaussian Detection and Application to Hyperspectral Imaging

Joana Frontera-Pons, *Student Member, IEEE*, Frédéric Pascal, *Senior Member, IEEE*, and Jean-Philippe Ovarlez, *Member, IEEE*

**Abstract**—Classical target detection schemes are usually obtained by deriving the likelihood ratio under Gaussian hypothesis and replacing the unknown background parameters by their estimates. In most applications, interference signals are assumed to be Gaussian with zero mean (or with a known mean vector) and with unknown covariance matrix. When the mean vector is unknown, it has to be jointly estimated with the covariance matrix, as it is the case for instance in hyperspectral imaging. In this paper, adaptive versions of the classical Matched Filter and the Normalized Matched Filter, as well as two versions of the Kelly detector are first derived and then analyzed for the case where the mean vector of the background is unknown. More precisely, theoretical closed-form expressions for false-alarm regulation are derived and the Constant False Alarm Rate property is pursued to allow the detector to be independent of nuisance parameters. Finally, the theoretical contributions are validated through simulations and the proposed detectors are tested on real hyperspectral scenes.

**Index Terms**—Hyperspectral Imaging, adaptive target detection, non zero-mean Gaussian distribution, false alarm regulation.

## I. INTRODUCTION

**H**YPERSPECTRAL imaging (HSI) extends from the fact that for any given material, the amount of radiation emitted varies with wavelength. HSI sensors measure the radiance of materials within each pixel area at a very large number of contiguous spectral bands and provide image data containing both spatial and spectral information (see [1] for more details and reference therein). Hyperspectral processing involves various applications such as unmixing [2], classification and dimensionality reduction [3], detection, ... Among them, hyperspectral detection is an active research topic that has led to many publications e.g. [4], [5], [6], [7]. More precisely, hyperspectral target detection methods are commonly used to detect targets embedded in the background and that generally cannot be solved by spatial resolution [8]. Furthermore, Detection Theory [9] arises in many different military and civilian applications and has been widely investigated in several signal processing domains such as radar, sonar and communications, see [10] for various references. There are two different methodologies for target detection purposes in the HSI literature [11]: Anomaly

Detection [5], [6] and Target Detection [4].

In many practical situations, there is not enough information about the target to detect, thus *Anomaly Detection* methods are widely used. The most widespread detector, the RX detector [12] is based on the Mahalanobis distance [13]. This detector and most of its variants search for pixels in the image with spectral characteristics that differ from the background. On the other hand, when the spectral signature of the desired target is known, it can be used as a steering vector in *Target Detection* techniques [11].

Interestingly, target detection methods have been extensively developed and analyzed in the signal processing and radar processing [10], [14], [15], [16]. In all these works as well as in several signal processing applications, signals are assumed to be Gaussian with zero mean or with a known mean vector (MV) that can be removed. In such context, Statistical Detection Theory [9] has led to several well-known algorithms, for instance the Matched Filter (MF) and its adaptive versions, the Kelly detector [14] and the Adaptive Normalized Matched Filter [17]. Other interesting approaches based on subspace projection methods have been derived and analyzed in [15]. However, when the mean vector of the noise background is unknown, these techniques are no longer adapted and improved methods have to be derived by taking into account the mean vector estimation. For this purpose, some preliminary results have been given in [18]. One of the contributions of this work is to extend and generalize these original results.

More precisely, this work deals with the classical Adaptive Matched Filter (AMF), the Kelly detection test and the Adaptive Normalized Matched Filter (ANMF). These detectors have been derived under Gaussian assumptions and benefit from great popularity in HSI target detection literature, see e.g. [19], [20]. To evaluate the detector performance, the classical process, according to the Neyman-Pearson criterion is first to regulate the false-alarm, by setting a detection threshold for a given probability of false-alarm (PFA). Since the PFA is related to the cumulative distribution function (CDF) of the detection test, this process is equivalent to the derivation of the detection test statistic. Then, the probability of detection can be evaluated for different Signal-to-Noise Ratios (SNR). Therefore, keeping the false-alarm rate constant (CFAR) is essential to set a proper detection threshold [21], [22]. The aim is to build a CFAR detector which provides

J. Frontera-Pons is with SONDRRA, Supelec, Plateau du Moulon, 3 rue Joliot-Curie, F-91190 Gif-sur-Yvette, France (e-mail: joana.fronterapons@supelec.fr)

F. Pascal is with SONDRRA, Supelec, Plateau du Moulon, 3 rue Joliot-Curie, F-91190 Gif-sur-Yvette, France (e-mail: frederic.pascal@supelec.fr)

J.-P. Ovarlez is with ONERA, DEMR/TSI, Chemin de la Hunière, F-91120 Palaiseau, France (e-mail: jean-philippe.ovarlez@onera.fr)

detection thresholds that are relatively immune to noise and background variation, and allow target detection with a constant false-alarm rate. The theoretical analysis of CFAR methods for adaptive detectors is a challenging problem since in adaptive schemes, the statistical distribution of the detectors is not always available in a closed-form expression.

The theoretical contributions of this paper are twofold. First, we derive the expression of each adaptive detector under the Gaussian assumption where both the mean vector and the covariance matrix (CM) are assumed to be unknown. Then, the exact derivation of the distribution of each proposed detection scheme under null hypothesis, i.e. when no target is supposed to be present, is provided. Thus, through Gaussian assumption, closed-form expressions for the false-alarm regulation are obtained, which allow to theoretically set the detection threshold for a given PFA.

On the other hand, one difficulty for the background detection statistic is to assume a tractable model or at least to account for robustness to deviation from the assumed theoretical model in the detection scheme. Since Gaussian assumption is not always fulfilled for real hyperspectral data, alternative robust estimation techniques are proposed in [23]. However, it is essential to notice that the derivations for many results in robust detection contexts strongly rely on the results obtained in the Gaussian context. For instance, this is the case in [24] where the derivation of a robust detector distribution is based on its Gaussian counterpart.

This paper is organized as follows. Section II introduces the required background on classical detection techniques as well as the obtention of the adaptive detectors for both unknown MV and CM. Then, Section III provides the main theoretical contributions of the paper by deriving the exact "PFA-threshold" relationship for the AMF, the "plug-in" Kelly detector and the ANMF under Gaussian assumption while a generalized version of the Kelly detector is derived. Finally, in Section IV, the theoretical analyses are validated through Monte-Carlo simulations and real HS data are processed to, first, extract homogeneous, let's say Gaussian, data and then, highlight the agreement with the proposed theoretical results. Conclusions and perspectives are drawn in Section V.

In the following, vectors (resp. matrices) are denoted by bold-faced lowercase letters (resp. uppercase letters).  $^T$  and  $^H$  respectively represent the transpose and the Hermitian operators.  $|\mathbf{A}|$  represents the determinant of the matrix  $\mathbf{A}$  and  $\text{Tr}(\mathbf{A})$  its trace.  $j$  is used to denote the unit imaginary number.  $\sim$  means "distributed as".  $\Gamma(\cdot)$  denotes the gamma function. Eventually,  $\Re\{\mathbf{x}\}$  represents the real part of the complex vector  $\mathbf{x}$ .

## II. BACKGROUND

After providing the general background in non-zero mean Gaussian detection, this section is devoted to review the expressions of the adaptive detectors.

The problem of detecting a signal corrupted by an additive noise  $\mathbf{b}$  in a  $m$ -dimensional complex vector  $\mathbf{x}$  can be stated as a the following binary hypothesis test:

$$\begin{cases} \mathcal{H}_0 : \mathbf{x} = \mathbf{b} \\ \mathcal{H}_1 : \mathbf{x} = \mathbf{s} + \mathbf{b} \end{cases},$$

and the signal  $\mathbf{s}$  can be written in the form  $\alpha \mathbf{p}$ , where  $\alpha$  is an unknown complex scalar amplitude, and  $\mathbf{p}$  is the steering vector describing the signal which is sought. Since the background statistics, i.e. the MV and the CM, are assumed to be unknown, they have to be estimated from  $\mathbf{x}_1, \dots, \mathbf{x}_N \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  a sequence of  $N$  independent and identically distributed (IID) signal-free secondary data. Then, the adaptive detector is obtained by replacing the unknown parameters by their estimates. In practice, an estimate may be obtained from the pixels surrounding the pixel under test, which play the role of the  $N$  IID signal-free secondary data vectors. The sample size  $N$  has to be chosen large enough to ensure the invertibility of the covariance matrix and small enough to justify both spectral homogeneity (stationarity) and spatial homogeneity. Let us now recall the detectors under interest in this work

### A. Adaptive Matched Filter

The MF detector is the optimal linear filter for maximizing the SNR in the presence of additive Gaussian noise with known parameters [9]. It corresponds to the Generalized Likelihood Ratio Test (GLRT) when the amplitude  $\alpha$  of the target to be detected is an unknown parameter. The MF detection scheme can be written as:

$$\Lambda_{MF} = \frac{|\mathbf{p}^H \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})|^2}{(\mathbf{p}^H \boldsymbol{\Sigma}^{-1} \mathbf{p})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda, \quad (1)$$

where  $\mathcal{H}_0$  and  $\mathcal{H}_1$  denote respectively the hypothesis of the absence and the presence of a target to detect. Note that it differs from the classical MF (zero-mean Gaussian Noise) by the term  $\boldsymbol{\mu}$ , the background mean, but without any consequence since  $\mathbf{x} - \boldsymbol{\mu} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$ . Moreover, the "PFA-threshold" relationship is given by [9]:

$$PFA_{MF} = \exp(-\lambda).$$

The two-step GLRT, called the AMF and denoted  $\Lambda_{AMF\hat{\boldsymbol{\Sigma}}}^{(N)}$  to underline the dependency with  $N$ , is usually built replacing the covariance matrix  $\boldsymbol{\Sigma}$  by any estimate  $\hat{\boldsymbol{\Sigma}}$  obtained from the  $N$  secondary data  $\{\mathbf{x}_i\}_{i \in [1, N]} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . If we consider a known mean vector  $\boldsymbol{\mu}$ , the adaptive version becomes:

$$\Lambda_{AMF\hat{\boldsymbol{\Sigma}}}^{(N)} = \frac{|\mathbf{p}^H \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{x} - \boldsymbol{\mu})|^2}{(\mathbf{p}^H \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{p})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda. \quad (2)$$

By choosing  $\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Sigma}}_{CSCM}$  where  $\hat{\boldsymbol{\Sigma}}_{CSCM}$  is the Centered Sample Covariance Matrix (CSCM) defined in Appendix A, the theoretical "PFA-threshold" relationship related to the test given in (1) is given by [16]

$$PFA_{AMF\hat{\boldsymbol{\Sigma}}} = {}_2F_1 \left( N - m + 1, N - m + 2; N + 1; -\frac{\lambda}{N} \right), \quad (3)$$

where  ${}_2F_1(\cdot)$  is the hypergeometric function [25] defined as,

$${}_2F_1(a, b; c; z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 \frac{t^{b-1}(1-t)^{c-b-1}}{(1-tz)^a} dt.$$

This detector holds the CFAR properties in the sense that its false alarm expression only depends on the dimension of the vector  $m$  and the number  $N$  of secondary data used for the estimation. Note that it is also independent of the noise covariance matrix  $\Sigma$  and the mean vector  $\mu$ , therefore the detector is said to be CFAR.

### B. Adaptive Kelly detector

The adaptive Kelly detector was derived in [14] and it is based on the Generalized Likelihood Ratio Test (GLRT) assuming Gaussian distribution. In this case, only the covariance matrix  $\Sigma$  is unknown and the mean vector  $\mu$  is assumed to be known. The Kelly detection test is obtained according to:

$$\Lambda_{\text{Kelly} \hat{\Sigma}_{SCM}}^{(N)} \quad (4)$$

$$= \frac{|\mathbf{p}^H \hat{\Sigma}_{SCM}^{-1} (\mathbf{x} - \mu)|^2}{\left( \mathbf{p}^H \hat{\Sigma}_{SCM}^{-1} \mathbf{p} \right) \left( N + (\mathbf{x} - \mu)^H \hat{\Sigma}_{SCM}^{-1} (\mathbf{x} - \mu) \right)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda. \quad (5)$$

As shown in [14], the PFA for the Kelly test is given by:

$$PFA_{\text{Kelly}} = (1 - \lambda)^{N-m+1}. \quad (6)$$

The Kelly detector is a CFAR test, in which the PFA is independent of the true covariance matrix and the mean vector. The AMF (two-step GLRT-based) and the Kelly detector (GLRT-based) have been derived on the same assumptions about the nature of the observations. It is therefore interesting to compare their detection performance for a given PFA. Note that for large values of  $N$  the performances are substantially the same.

### C. Adaptive Normalized Matched Filter

The Normalized Matched Filter (NMF) [26] was obtained in Gaussian noise hypothesis but when considering that the covariance matrix is of the form  $\sigma^2 \Sigma$  with an unknown variance  $\sigma^2$  but known structure  $\Sigma$ . The GLRT leads to

$$\Lambda_{NMF} = \frac{|\mathbf{p}^H \Sigma^{-1} (\mathbf{x} - \mu)|^2}{(\mathbf{p}^H \Sigma^{-1} \mathbf{p}) ((\mathbf{x} - \mu)^H \Sigma^{-1} (\mathbf{x} - \mu))} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda, \quad (7)$$

The PFA-threshold relationship is given [26]:

$$PFA_{NMF} = (1 - \lambda)^{(m-1)}.$$

The two step-GLRT, called ANMF, is generally obtained when the unknown noise covariance matrix  $\Sigma$  is replaced by an estimate [15]:

$$\Lambda_{ANMF \hat{\Sigma}}^{(N)} = \frac{|\mathbf{p}^H \hat{\Sigma}^{-1} (\mathbf{x} - \mu)|^2}{\left( \mathbf{p}^H \hat{\Sigma}^{-1} \mathbf{p} \right) \left( (\mathbf{x} - \mu)^H \hat{\Sigma}^{-1} (\mathbf{x} - \mu) \right)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda.$$

For the choice for  $\hat{\Sigma} = \hat{\Sigma}_{SCM}$ , the PFA follows [15]:

$$PFA_{ANMF \hat{\Sigma}_{SCM}} = (1 - \lambda)^{a-1} {}_2F_1(a, a-1; b-1; \lambda), \quad (8)$$

where  $a = N - m + 2$  and  $b = N + 2$ .

## III. MAIN RESULTS

In this section, let us now assume that the mean vector  $\mu$  is an unknown parameter as it is the case for instance in HSI and let us derive the new corresponding detection schemes. Then, using standard calculus on Wishart distributions, recapped in Appendix B, the distributions of each detection test is provided.

### A. Adaptive Matched Filter Detector

When both covariance matrix  $\Sigma$  and mean vector  $\mu$  are unknown, the two-step GLRT procedure leads to replace them by their estimates  $\hat{\Sigma}$  and  $\hat{\mu}$  built from the  $N$  secondary data  $\{\mathbf{x}_i\}_{i \in [1, N]}$  in (1) leading to the AMF detector of the following form:

$$\Lambda_{AMF \hat{\Sigma}, \hat{\mu}}^{(N)} = \frac{|\mathbf{p}^H \hat{\Sigma}^{-1} (\mathbf{x} - \hat{\mu})|^2}{(\mathbf{p}^H \hat{\Sigma}^{-1} \mathbf{p})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda,$$

where the notation  $\Lambda_{AMF \hat{\Sigma}, \hat{\mu}}^{(N)}$  is used to stress now the dependency on the estimated mean vector  $\hat{\mu}$ . Under Gaussian assumption, and for the particular MLE choice  $\hat{\Sigma} = \hat{\Sigma}_{SCM}$  and  $\hat{\mu} = \hat{\mu}_{SMV}$  defined in Appendix A, the distribution of this detection test is given in the next Proposition, through its PFA.

**Proposition III.1** *Under Gaussian assumptions, the theoretical relationship between the PFA and the threshold  $\lambda$  is given by*

$$PFA_{AMF \hat{\Sigma}, \hat{\mu}} = {}_2F_1 \left( N - m, N - m + 1; N; -\frac{\lambda}{N + 1} \right), \quad (9)$$

where  $\hat{\Sigma} = \hat{\Sigma}_{SCM}$  and  $\hat{\mu} = \hat{\mu}_{SMV}$ .

Before turning into the proof, let us comment on this result.

- Interestingly, this detector also holds the CFAR property in the sense that its false-alarm expression depends only on the dimension  $m$  and on the number of secondary data  $N$ , but not on the noise parameters  $\mu$  and  $\Sigma$ . Note that the only effect of estimating the mean is the loss of one degree of freedom and the modification of the threshold compared to (3). Obviously, the impact of these modifications decrease as the number of secondary data  $N$  used to estimate the unknown parameters increases.

- Moreover, the result has been obtained when using the MLEs of the unknown parameters but the proof can be easily extended to other covariance matrix estimators

such as  $\hat{\Sigma} = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}_i - \hat{\mu})(\mathbf{x}_i - \hat{\mu})^H$  which is the unbiased covariance matrix estimate

*Proof:* For simplicity matters, the following notations are used:  $\hat{\Sigma} = \hat{\Sigma}_{SCM}$  and  $\hat{\mu} = \hat{\mu}_{SMV}$ .

Since the derivation of the PFA is done under hypothesis  $\mathcal{H}_0$ , let us set  $\{\mathbf{x}_i\}_{i \in [1, N]} \sim \mathcal{CN}(\mu, \Sigma)$  and  $\mathbf{x} \sim \mathcal{CN}(\mu, \Sigma)$ , where all these vectors are independent. Now, let us denote

$$\hat{\mathbf{W}}_{N-1} = \sum_{i=1}^N (\mathbf{x}_i - \hat{\mu})(\mathbf{x}_i - \hat{\mu})^H = N \hat{\Sigma}_{SCM}.$$

Note that as an application of the Cochran theorem (see e.g. [27]), one has

$$\hat{\mathbf{W}}_{N-1} \stackrel{\text{dist.}}{=} \sum_{i=1}^{N-1} (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^H = (N-1) \hat{\boldsymbol{\Sigma}}_{SCM},$$

where  $\stackrel{\text{dist.}}{=}$  means *is distributed as*.

Since  $\hat{\boldsymbol{\mu}} \sim \mathcal{CN}\left(\boldsymbol{\mu}, \frac{1}{N} \boldsymbol{\Sigma}\right)$ , one has  $\mathbf{x} - \hat{\boldsymbol{\mu}} \sim \mathcal{CN}\left(\mathbf{0}, \frac{N+1}{N} \boldsymbol{\Sigma}\right)$ . This can be equivalently rewritten as

$$\sqrt{N/(N+1)}(\mathbf{x} - \hat{\boldsymbol{\mu}}) \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}).$$

Now, let us set  $\mathbf{y} = \sqrt{\frac{N}{N+1}}(\mathbf{x} - \hat{\boldsymbol{\mu}})$  with  $\mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$ .

As we jointly estimate the mean and the covariance matrix, a degree of freedom is lost, compared to the only covariance matrix estimation problem.

Let us now consider the classical AMF test (i.e.  $\boldsymbol{\mu}$  known) built from  $N-1$  secondary data, rewritten in terms of  $\hat{\mathbf{W}}_{N-1}$ :

$$\Lambda_{AMF}^{(N-1)} \hat{\boldsymbol{\Sigma}} = (N-1) \frac{|\mathbf{p}^H \hat{\mathbf{W}}_{N-1}^{-1} \mathbf{y}|^2}{(\mathbf{p}^H \hat{\mathbf{W}}_{N-1}^{-1} \mathbf{p})},$$

where  $\mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$  and whose "PFA-threshold" relationship is given by (3) where  $N$  is replaced by  $N-1$ .

Now, for the joint estimation problem, the AMF can be rewritten as:

$$\begin{aligned} \Lambda_{AMF, \hat{\boldsymbol{\Sigma}}, \hat{\boldsymbol{\mu}}}^{(N)} &= N \frac{|\mathbf{p}^H \hat{\mathbf{W}}_{N-1}^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}})|^2}{(\mathbf{p}^H \hat{\mathbf{W}}_{N-1}^{-1} \mathbf{p})}, \\ &= N \frac{N+1}{N} \frac{|\mathbf{p}^H \hat{\mathbf{W}}_{N-1}^{-1} \mathbf{y}|^2}{(\mathbf{p}^H \hat{\mathbf{W}}_{N-1}^{-1} \mathbf{p})}, \\ &\stackrel{\text{dist.}}{=} \frac{(N+1)}{(N-1)} \Lambda_{AMF}^{(N-1)} \hat{\boldsymbol{\Sigma}}. \end{aligned}$$

where  $(\mathbf{x} - \hat{\boldsymbol{\mu}})$  has been replaced by  $\sqrt{N+1/N} \mathbf{y}$  with  $\mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$ , as previously.

Hence, one can determine the false-alarm relationship:

$$\begin{aligned} PFA_{AMF, \hat{\boldsymbol{\Sigma}}, \hat{\boldsymbol{\mu}}} &= \mathbb{P}\left(\Lambda_{AMF, \hat{\boldsymbol{\Sigma}}, \hat{\boldsymbol{\mu}}}^{(N)} > \lambda | \mathcal{H}_0\right), \\ &= \mathbb{P}\left(\frac{(N+1)}{(N-1)} \Lambda_{AMF}^{(N-1)} \hat{\boldsymbol{\Sigma}} > \lambda | \mathcal{H}_0\right), \\ &= \mathbb{P}\left(\Lambda_{AMF}^{(N-1)} \hat{\boldsymbol{\Sigma}} > \lambda' | \mathcal{H}_0\right), \end{aligned}$$

where  $\lambda' = \frac{(N-1)}{(N+1)} \lambda$ , which leads to the conclusion. ■

### B. Kelly Detector

The exact GLRT Kelly detector for both unknown mean vector  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$  has now to be derived since it does not correspond to the Kelly detector given in (4) in which an estimate of the mean is simply plugged (two-step GLRT). Following the same lines as in [14], we now

assume that both the mean vector and the covariance matrix are unknown. The likelihood functions under  $\mathcal{H}_0$  and  $\mathcal{H}_1$  are given by:

$$f_i(\mathbf{x}) = \left( \frac{1}{\pi^m |\boldsymbol{\Sigma}|} \exp[-\text{Tr}(\boldsymbol{\Sigma}^{-1} \mathbf{T}_i)] \right)^{N+1}, \quad (10)$$

for  $i = 0, 1$  and where

$$(N+1) \mathbf{T}_0 = (\mathbf{x} - \boldsymbol{\mu}_0)(\mathbf{x} - \boldsymbol{\mu}_0)^H + \sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu}_0)(\mathbf{x}_i - \boldsymbol{\mu}_0)^H,$$

$$(N+1) \mathbf{T}_1 = (\mathbf{x} - \alpha \mathbf{p} - \boldsymbol{\mu}_1)(\mathbf{x} - \alpha \mathbf{p} - \boldsymbol{\mu}_1)^H + \sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu}_1)(\mathbf{x}_i - \boldsymbol{\mu}_1)^H,$$

and

$$\boldsymbol{\mu}_0 = \frac{1}{N+1} \left( \mathbf{x} + \sum_{i=1}^N \mathbf{x}_i \right), \quad (11)$$

$$\boldsymbol{\mu}_1 = \frac{1}{N+1} \left( \mathbf{x} - \alpha \mathbf{p} + \sum_{i=1}^N \mathbf{x}_i \right). \quad (12)$$

Under  $\mathcal{H}_0$  and  $\mathcal{H}_1$ , the maxima are achieved at

$$\max_{\boldsymbol{\Sigma}, \boldsymbol{\mu}} f_i = \left( \frac{1}{(\pi e)^m |\mathbf{T}_i|} \right)^{N+1}, \quad \text{for } i = 0, 1,$$

And neglecting the exponent  $N+1$ , one obtains the following LR:

$$L(\alpha) = \frac{|\mathbf{T}_0|}{|\mathbf{T}_1|} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \eta.$$

Then, as this LR still depends on the unknown amplitude  $\alpha$  of the signal, thus, it has to be maximized w.r.t  $\alpha$ , which is equivalent to minimize  $\mathbf{T}_1$  w.r.t  $\alpha$ . A way to do this is to introduce the following sample covariance matrix:

$$\mathbf{S}_0 = \sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu}_0)(\mathbf{x}_i - \boldsymbol{\mu}_0)^H. \quad (13)$$

Then,  $(N+1) |\mathbf{T}_0|$  can be written as

$$(N+1) |\mathbf{T}_0| = |\mathbf{S}_0| (1 + (\mathbf{x} - \boldsymbol{\mu}_0)^H \mathbf{S}_0^{-1} (\mathbf{x} - \boldsymbol{\mu}_0)).$$

In the same way, and after some manipulations,  $(N+1) |\mathbf{T}_1|$  becomes

$$\begin{aligned} (N+1) |\mathbf{T}_1| &= |\mathbf{S}_0| \left( \sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu}_1)^H \mathbf{S}_0^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_1) \right. \\ &\quad \left. + (\mathbf{x} - \alpha \mathbf{p} - \boldsymbol{\mu}_1)^H \mathbf{S}_0^{-1} (\mathbf{x} - \alpha \mathbf{p} - \boldsymbol{\mu}_1) \right), \\ &= |\mathbf{S}_0| (A + B). \end{aligned}$$

Now, let us rewrite the two terms  $A$  and  $B$  to separate the terms involving  $\alpha$ . By recalling that  $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_0 - \frac{1}{N+1} \alpha \mathbf{p}$ ,

one obtains:

$$A = 1 + \frac{N|\alpha|^2}{(N+1)^2} \mathbf{p}^H \mathbf{S}_0^{-1} \mathbf{p} + \frac{2}{N+1} \Re \left\{ \bar{\alpha} \mathbf{p}^H \mathbf{S}_0^{-1} \sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu}_0) \right\},$$

$$B = (\mathbf{x} - \boldsymbol{\mu}_0)^H \mathbf{S}_0^{-1} (\mathbf{x} - \boldsymbol{\mu}_0) + \frac{N^2 |\alpha|^2}{(N+1)^2} \mathbf{p}^H \mathbf{S}_0^{-1} \mathbf{p} - \frac{2N}{N+1} \Re \left\{ \bar{\alpha} \mathbf{p}^H \mathbf{S}_0^{-1} (\mathbf{x} - \boldsymbol{\mu}_0) \right\}.$$

Notice that  $\mathbf{x} - \boldsymbol{\mu}_0 = -\sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu}_0)$ , then rearranging the expression of  $(N+1)|\mathbf{T}_1|$  leads to

$$\frac{(N+1)|\mathbf{T}_1|}{|\mathbf{S}_0|} = \frac{(N+1)|\mathbf{T}_0|}{|\mathbf{S}_0|} + \frac{N|\alpha|^2}{(N+1)} \mathbf{p}^H \mathbf{S}_0^{-1} \mathbf{p} - 2 \Re \left\{ \bar{\alpha} \mathbf{p}^H \mathbf{S}_0^{-1} (\mathbf{x} - \boldsymbol{\mu}_0) \right\}.$$

Now, the term depending on  $\alpha$  can be rewritten as follows

$$\frac{N}{(N+1)} \mathbf{p}^H \mathbf{S}_0^{-1} \mathbf{p} \left| \alpha - \frac{N+1}{N} \frac{\mathbf{p}^H \mathbf{S}_0^{-1} (\mathbf{x} - \boldsymbol{\mu}_0)}{\mathbf{p}^H \mathbf{S}_0^{-1} \mathbf{p}} \right|^2 - \frac{N+1}{N} \frac{|\mathbf{p}^H \mathbf{S}_0^{-1} (\mathbf{x} - \boldsymbol{\mu}_0)|^2}{\mathbf{p}^H \mathbf{S}_0^{-1} \mathbf{p}}.$$

Minimizing  $|\mathbf{T}_1|$  w.r.t  $\alpha$  is equivalent to cancelling the square term in the previous equation. Thus, the GLRT can now be written according to the following definition.

**Definition III.1 (The generalized Kelly detector)** *Under Gaussian assumptions, the extension of Kelly's test when both the mean vector and the covariance matrix of the background are unknown takes the following form:*

$$\Lambda = \frac{\beta(N) |\mathbf{p}^H \mathbf{S}_0^{-1} (\mathbf{x} - \boldsymbol{\mu}_0)|^2}{(\mathbf{p}^H \mathbf{S}_0^{-1} \mathbf{p}) (1 + (\mathbf{x} - \boldsymbol{\mu}_0)^H \mathbf{S}_0^{-1} (\mathbf{x} - \boldsymbol{\mu}_0))} \stackrel{\mathcal{H}_1}{\geq} \lambda, \quad (14)$$

where  $\beta(N) = \frac{N+1}{N}$  and

- $\mathbf{S}_0 = \sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu}_0)(\mathbf{x}_i - \boldsymbol{\mu}_0)^H,$
- $\boldsymbol{\mu}_0 = \frac{1}{N+1} \left( \mathbf{x} + \sum_{i=1}^N \mathbf{x}_i \right).$

Let us now comment on this new detector. One can notice that both the covariance matrix  $\mathbf{S}_0$  as well as the mean  $\boldsymbol{\mu}_0$  estimates depend on the data  $\mathbf{x}$  under test, which is not the case in other classical detectors where the unknown parameters are estimated from signal-free secondary data. Consequently,  $\mathbf{S}_0$  and  $\mathbf{x} - \boldsymbol{\mu}_0$  are not independent. Moreover, the covariance matrix estimate  $\mathbf{S}_0$  is not Wishart-distributed due to the non-standard mean estimate  $\boldsymbol{\mu}_0$ . Thus, the derivation of this ratio distribution is very difficult.

As for previous detector, it would be intuitive to think that the proposed test behaves as the classical Kelly's test but for

$N-1$  degrees of freedom. To prove that let us first rewrite (14) as follows:

$$\Lambda = \frac{|\mathbf{p}^H \mathbf{S}_0^{-1} \mathbf{y}|^2}{(\mathbf{p}^H \mathbf{S}_0^{-1} \mathbf{p}) \left( 1 + \frac{N}{N+1} \mathbf{y}^H \mathbf{S}_0^{-1} \mathbf{y} \right)} \stackrel{\mathcal{H}_1}{\geq} \lambda, \quad \stackrel{\mathcal{H}_0}{\geq}$$

where we use:

- $(\mathbf{x} - \boldsymbol{\mu}_0) = \frac{N}{N+1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}),$
- $\hat{\boldsymbol{\mu}}_{SMV} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i,$
- $\mathbf{y} = \sqrt{\frac{N}{N+1}} (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}) \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}).$

Now, let us set  $\mathbf{S}_0^{(i)} = \sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu}_0^{(i)}) (\mathbf{x}_i - \boldsymbol{\mu}_0^{(i)})^H,$

where  $\boldsymbol{\mu}_0^{(i)} = \frac{1}{N} \left( \sum_{j \neq i}^N \mathbf{x}_j + \mathbf{x} \right)$ . Then, the test becomes

$$\frac{\frac{N+1}{N} |\mathbf{p}^H \mathbf{S}_0^{(i)-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})|^2}{\left( \mathbf{p}^H \mathbf{S}_0^{(i)-1} \mathbf{p} \right) \left( 1 + (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})^H \mathbf{S}_0^{(i)-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}) \right)}.$$

One can notice that each  $\mathbf{x}_i$  (including  $\mathbf{x}$ ) plays the same role, thus the distribution of this test is the same for every permutation of the  $(N+1)$ -sample  $(\mathbf{x}, \mathbf{x}_1, \dots, \mathbf{x}_N)$ . However, the dependency between the covariance matrix estimate and the data under test  $\mathbf{x}$  still remains.

To fill this gap, another way of taking advantage of the Kelly's detector when the mean vector is unknown can be to use the classical scheme recalled in (4) and to plug the classical estimator of the mean, based only on the  $N$  secondary data, i.e.  $\hat{\boldsymbol{\mu}}_{SMV} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i$ . This leads to the two-step GLRT Kelly's detector:

$$\Lambda_{\text{Kelly}}^{(N)} \hat{\boldsymbol{\Sigma}}_{SCM}, \hat{\boldsymbol{\mu}}_{SMV} = \frac{|\mathbf{p}^H \hat{\boldsymbol{\Sigma}}_{SCM}^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})|^2}{\left( \mathbf{p}^H \hat{\boldsymbol{\Sigma}}_{SCM}^{-1} \mathbf{p} \right) \left( N + (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})^H \hat{\boldsymbol{\Sigma}}_{SCM}^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}) \right)} \stackrel{\mathcal{H}_1}{\geq} \lambda.$$

In this case, the distribution can be derived. This is the purpose of the following proposition.

**Proposition III.2** *The theoretical relationship between the PFA and the threshold is given by*

$$PFA_{\text{Kelly}} \hat{\boldsymbol{\Sigma}}_{SCM}, \hat{\boldsymbol{\mu}}_{SMV} = \frac{\Gamma(N)}{\Gamma(N-m+1) \Gamma(m-1)} \times \int_0^1 \left[ 1 + \frac{\lambda}{1-\lambda} \left( 1 - \frac{u}{N+1} \right) \right]^{m-N} u^{N-m} (1-u)^{m-2} du. \quad (15)$$

*Proof:* The detection test rewritten with  $\hat{\Sigma}_{SCM}^{-1} = N \mathbf{W}_{N-1}^{-1}$  becomes:

$$\Lambda_{Kelly \hat{\Sigma}, \hat{\mu}}^{(N)} = \frac{N^2 \left| \mathbf{p}^H \hat{\mathbf{W}}_{N-1}^{-1} (\mathbf{x} - \hat{\mu}) \right|^2}{N \left( \mathbf{p}^H \hat{\mathbf{W}}_{N-1}^{-1} \mathbf{p} \right) \left( N + N \mathbf{y}^H \hat{\mathbf{W}}_{N-1}^{-1} (\mathbf{x} - \hat{\mu}) \right)},$$

and replacing  $(\mathbf{x} - \hat{\mu})$  by  $\sqrt{\frac{N+1}{N}} \mathbf{y}$ , one obtains:

$$\begin{aligned} \Lambda_{Kelly \hat{\Sigma}, \hat{\mu}}^{(N)} &= \frac{\frac{N+1}{N} N^2 \left| \mathbf{p}^H \hat{\mathbf{W}}_{N-1}^{-1} \mathbf{y} \right|^2}{N \left( \mathbf{p}^H \hat{\mathbf{W}}_{N-1}^{-1} \mathbf{p} \right) \left( N + \frac{N+1}{N} N \mathbf{y}^H \hat{\mathbf{W}}_{N-1}^{-1} \mathbf{y} \right)} \\ &= \frac{\left| \mathbf{p}^H \hat{\mathbf{W}}_{N-1}^{-1} \mathbf{y} \right|^2}{\left( \mathbf{p}^H \hat{\mathbf{W}}_{N-1}^{-1} \mathbf{p} \right) \left( \frac{N}{N+1} + \mathbf{y}^H \hat{\mathbf{W}}_{N-1}^{-1} \mathbf{y} \right)} \end{aligned}$$

with  $\mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \Sigma)$ .

The classical Kelly detector obtained when the mean vector is known is recalled here, built with  $N-1$  zero-mean Gaussian data, and written with  $\hat{\mathbf{W}}_{N-1}$ :

$$\Lambda_{Kelly \hat{\Sigma}}^{(N-1)} = \frac{\left| \mathbf{p}^H \hat{\mathbf{W}}_{N-1}^{-1} \mathbf{y} \right|^2}{\left( \mathbf{p}^H \hat{\mathbf{W}}_{N-1}^{-1} \mathbf{p} \right) \left( 1 + \mathbf{y}^H \hat{\mathbf{W}}_{N-1}^{-1} \mathbf{y} \right)}. \quad (16)$$

It is worth pointing out that the term  $N/(N+1)$  resulting from the mean estimation in  $\Lambda_{Kelly \hat{\Sigma}, \hat{\mu}}^{(N)}$  does not appear in the classical Kelly detector (16). This fact prevents from relating the two expressions. Thus, a proof similar to the Proposition III.1 is not feasible.

According to [15], [28], an equivalent LR can be expressed as:

$$\hat{\kappa}^2 = \frac{\Lambda_{Kelly \hat{\Sigma}, \hat{\mu}}^{(N)}}{1 - \Lambda_{Kelly \hat{\Sigma}, \hat{\mu}}^{(N)}} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \frac{\lambda}{1 - \lambda}.$$

Following the same development proposed in [15], the statistic  $\hat{\kappa}^2$  can be identified as the ratio  $\theta/\beta$  between two independent scalar random variables  $\theta$  and  $\beta$ . For this particular development of Kelly distribution with non-centered data, the scalar random variable  $\beta$  is found to have the same distribution as the function  $1-u/(N+1)$  where  $u$  is a random variable following a complex central beta distribution with  $N-m+1, m-1$  degrees of freedom:

$$u \sim f_u(u) = \frac{\Gamma(N)}{\Gamma(N-m+1) \Gamma(m-1)} u^{N-m} (1-u)^{m-2},$$

whereas the p.d.f. of the variable  $\theta$  is distributed according to the complex  $F$ -distribution with  $1, N-m$  degrees of freedom scaled by  $1/(N-m)$ :

$$\theta \sim f_\theta(\theta) = (N-m) (1+\theta)^{m-N-1}$$

One can now derive the cumulative density function of the Kelly test as:

$$\begin{aligned} \mathbb{P} \left( \Lambda_{Kelly \hat{\Sigma}, \hat{\mu}}^{(N)} \leq \lambda \right) &= \mathbb{P} \left( \hat{\kappa}^2 \leq \frac{\lambda}{1-\lambda} \right) = \mathbb{P} \left( \theta \leq \beta \frac{\lambda}{1-\lambda} \right) \\ &= \int_0^1 \left[ \int_0^{\frac{\lambda}{1-\lambda} (1-u/(N+1))} f_\theta(v) dv \right] f_u(u) du. \end{aligned}$$

Solving the integral one obtains the "PFA-threshold" relationship:

$$\begin{aligned} PFA_{Kelly \hat{\Sigma}, \hat{\mu}} &= \frac{\Gamma(N)}{\Gamma(N-m+1) \Gamma(m-1)} \\ &\times \int_0^1 \left[ 1 + \frac{\lambda}{1-\lambda} \left( 1 - \frac{u}{N+1} \right) \right]^{m-N} u^{N-m} (1-u)^{m-2} du. \end{aligned}$$

However, the final expression can not be further simplified to a closed-form expression as those obtained for the other detectors can not be determined. ■

### C. Adaptive Normalized Matched Filter

Similarly, the ANMF for both mean vector and covariance matrix estimation becomes:

$$\Lambda_{ANMF \hat{\Sigma}, \hat{\mu}} = \frac{\left| \mathbf{p}^H \hat{\Sigma}^{-1} (\mathbf{x} - \hat{\mu}) \right|^2}{\left( \mathbf{p}^H \hat{\Sigma}^{-1} \mathbf{p} \right) \left( (\mathbf{x} - \hat{\mu})^H \hat{\Sigma}^{-1} (\mathbf{x} - \hat{\mu}) \right)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \lambda.$$

**Proposition III.3** The theoretical relationship between the PFA and the threshold is given by

$$PFA_{ANMF \hat{\Sigma}, \hat{\mu}} = (1-\lambda)^{a-1} {}_2F_1(a, a-1; b-1; \lambda), \quad (17)$$

where  $a = (N-1) - m + 2$ ,  $b = (N-1) + 2$  and where  $\hat{\Sigma} = \hat{\Sigma}_{SCM}$  and  $\hat{\mu} = \hat{\mu}_{SMV}$ .

*Proof:* The proof is similar to the proof of Proposition III.1. The main difference is due to the normalization term  $(\mathbf{x} - \hat{\mu})^H \hat{\Sigma}^{-1} (\mathbf{x} - \hat{\mu})$ . Indeed, the correction factor  $N/(N-1)$  appears both at the numerator and at the denominator, and consequently, it disappears. The same argument is also true for the factor  $N$  that arises from the covariance matrix estimates, i.e. since the detector is homogeneous of degree 0 in terms of covariance matrix estimates (i.e.  $\Lambda_{ANMF \hat{\Sigma}, \hat{\mu}} = \Lambda_{ANMF \gamma \hat{\Sigma}, \hat{\mu}}$  for any real  $\gamma$ ), this scalar also disappears. Thus, the distribution of the ANMF with an estimate of the mean is exactly the same as in (8) where  $N$  is replaced by  $N-1$ . ■

## IV. SIMULATIONS

In this section, we validate the theoretical analysis on simulated data. The experiments were conducted on  $m = 5$  dimensional Gaussian vectors, for different values of  $N$ , the number of secondary data and the computations have been made through  $10^6$  Monte-Carlo trials. The true covariance is chosen as a Toeplitz matrix whose entries are  $\Sigma_{i,j} = \rho^{|i-j|}$  and where  $\rho = 0.4$ . The mean vector is arbitrary set to have all entries equal to  $(3+4j)$ .

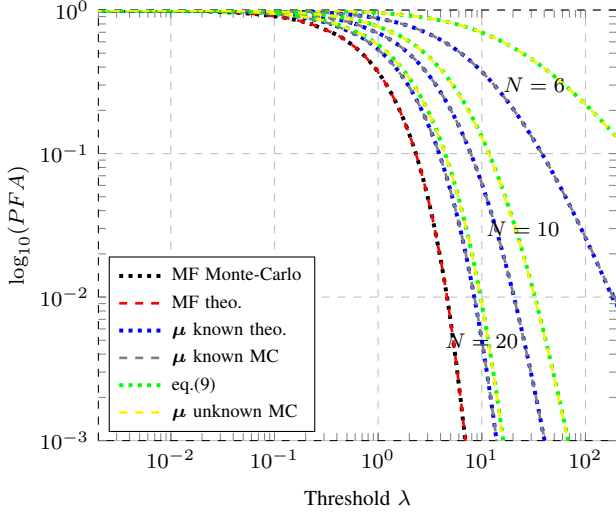


Fig. 1: PFA versus threshold for the AMF for different values of  $N$  ( $m = 5$ ,  $\mathbf{p} = [1, \dots, 1]^T$ ,  $\rho = 0.4$ ,  $\boldsymbol{\mu} = (4 + 3j)\mathbf{p}$ ) when a)  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  are known (MF) (red and black curves), b) only  $\boldsymbol{\mu}$  is known (gray and blue curves) and c) Proposition III.1: both  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  are unknown (yellow and green curves).

#### A. False Alarm Regulation

The FA regulation is presented for previous detection schemes having a closed-form expression, i.e. for all except the generalized Kelly detector. Fig. 1 shows the false-alarm regulation for the MF, the AMF when only the covariance matrix is unknown and the AMF for both covariance matrix and mean vector unknown. The steering vector used for the simulations is the unity vector  $\mathbf{p} = [1, \dots, 1]^T$  without loss of generality as all the PDF are found to be independent of the steering vector  $\mathbf{p}$ . The perfect agreement of the green and yellow curves illustrates the results of Proposition III.1. Moreover, remark that when  $N$  increases both AMF get closer to each other, and they approach the known parameters case MF.

Fig. 2 and Fig. 3 present the FA regulation for the Kelly detector and the ANMF respectively, under Gaussian assumption. For clarity purposes, the results are displayed in terms of the threshold  $\eta = (1 - \lambda)^{-(N+1)}$  for Adaptive Kelly detectors, and  $\eta = (1 - \lambda)^{-m}$  for ANMF and NMF detectors, respectively and a logarithmic scale is used. This validates results of Proposition III.2 and III.3 for the SCM-SMV.

Remark that the derived relationships given by eqs. (9) and (17) are quite similar to those for which the mean is known. However, as illustrated in Fig. 1 and Fig. 3, there is an important difference for small values of  $N$ . It is worth pointing out that the theoretical "PFA-threshold" relationships presented above depend only on the size of the vectors  $m$  and the number of secondary data used to estimate the parameters  $N$ . Thus, the detector outcome will not depend on the true value of the covariance matrix or the mean vector. These three detectors hold the CFAR property with respect to the background parameters. However, their distribution strongly relies on the underlying distribution of the background, i.e.

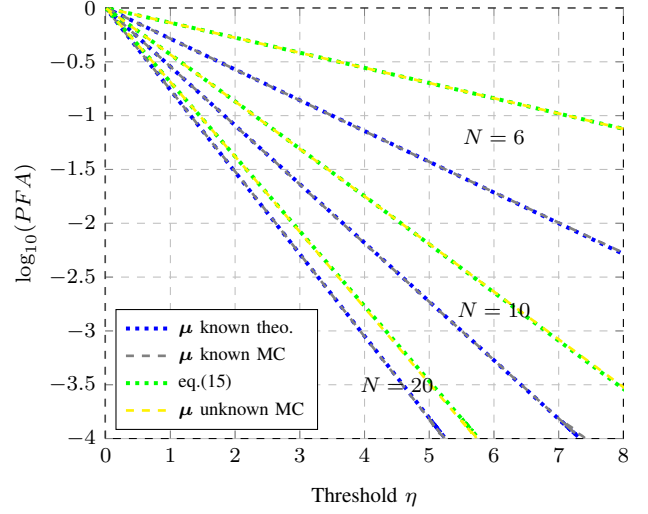


Fig. 2: PFA versus threshold for the "plug-in" Kelly detector for different values of  $N$  ( $m = 5$ ,  $\mathbf{p} = [1, \dots, 1]^T$ ,  $\rho = 0.4$ ,  $\boldsymbol{\mu} = (4 + 3j)\mathbf{p}$ ) when a) only  $\boldsymbol{\mu}$  is known (gray and blue curves) and b) Proposition III.2: both  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  are unknown (yellow and green curves).

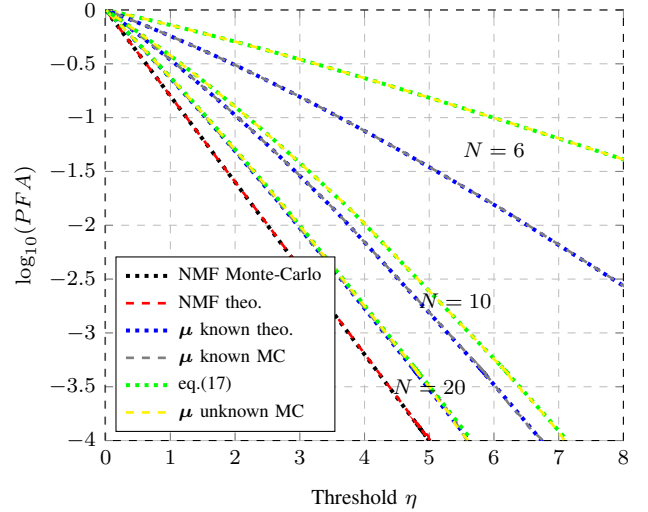


Fig. 3: PFA versus threshold for the ANMF for different values of  $N$  ( $m = 5$ ,  $\mathbf{p} = [1, \dots, 1]^T$ ,  $\rho = 0.4$ ,  $\boldsymbol{\mu} = (4 + 3j)\mathbf{p}$ ) when a)  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  are known (NMF) (red and black curves), b) only  $\boldsymbol{\mu}$  is known (gray and blue curves) and c) Proposition III.3: both  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  are unknown (yellow and green curves).

if Gaussian assumption is not fulfilled the "PFA-threshold" relationship will divert from the theoretical results derived in this paper.

#### B. Performance Evaluation

The four detection schemes are compared in terms of probability of detection. The experiments were conducted to detect a vector  $\alpha\mathbf{p}$  embedded in Gaussian noise with the same distribution parameters than for false alarm regulation. The Monte-Carlo simulation was set for dimensions  $m = 5$  and  $N = 10$  and for the probability of false alarm  $PFA = 10^{-3}$ .



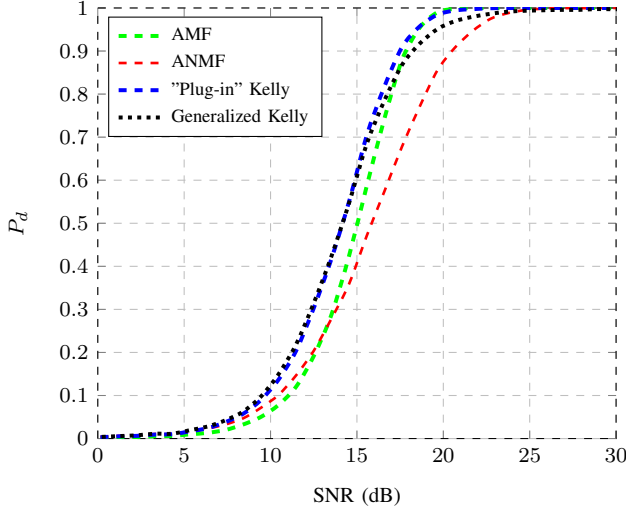


Fig. 4: Probability of detection for  $PFA = 10^{-3}$  corresponding to different values of  $SNR = \alpha^2 \mathbf{p}^H \mathbf{\Sigma}^{-1} \mathbf{p}$  in Gaussian case. ( $m = 5$ ,  $N = 20$ ,  $\mathbf{p} = [1, \dots, 1]^T$ ,  $\rho = 0.4$ ).

Then, the threshold  $\lambda$  has been adjusted according to the false alarm regulation relative to each detectors (AMF, ANMF, two-step GLRT Kelly, Generalized Kelly). Fig. 4 presents the detection probability versus the SNR defined as  $\alpha^2 \mathbf{p}^H \mathbf{\Sigma}^{-1} \mathbf{p}$  with the known steering vector  $\mathbf{p} = [1, \dots, 1]^T$ . The detectors delivering the best performance results are the Kelly detectors ("two-step GLRT" and generalized). Actually, these detectors lead to very similar performance with a small improvement of the generalized (resp. "two-step GLRT") one at low (resp. high) SNR. As expected, the AMF and the ANMF require a higher SNR to achieve same performance.

### C. Hyperspectral Real Data

The same experiments have been conducted on a experimental hyperspectral image. The scene analyzed is the NASA Hyperion sensor dataset displayed in Fig. 5. The image is constituted of  $798 \times 253$  pixels and 116 spectral bands after water absorption bands have been removed. The analysis has been done on a homogenous part of the image corresponding to the water region on the top left of the image (blue part). The part extracted consists in  $60 \times 20$  pixels. In order to ensure the validity of the proposed methods, we show in Fig. 6 the outcome of a classical Gaussianity test "Q-Q plot" for the selected region over the band 42. However, these techniques allow to "validate" the Gaussianity of each band but cannot ensure the Gaussianity of the corresponding vector.

Since hyperspectral data are real and positive, the data have been passed through a Hilbert filter to render them complex. A downsampling taking one over two consecutive bands is required to avoid redundant information that can reduce the covariance matrix rank. However, it is important to note that the real component after Hilbert transform is still the original signal (principle of analytic signals). The only goal of this procedure is simply to be in the same assumptions

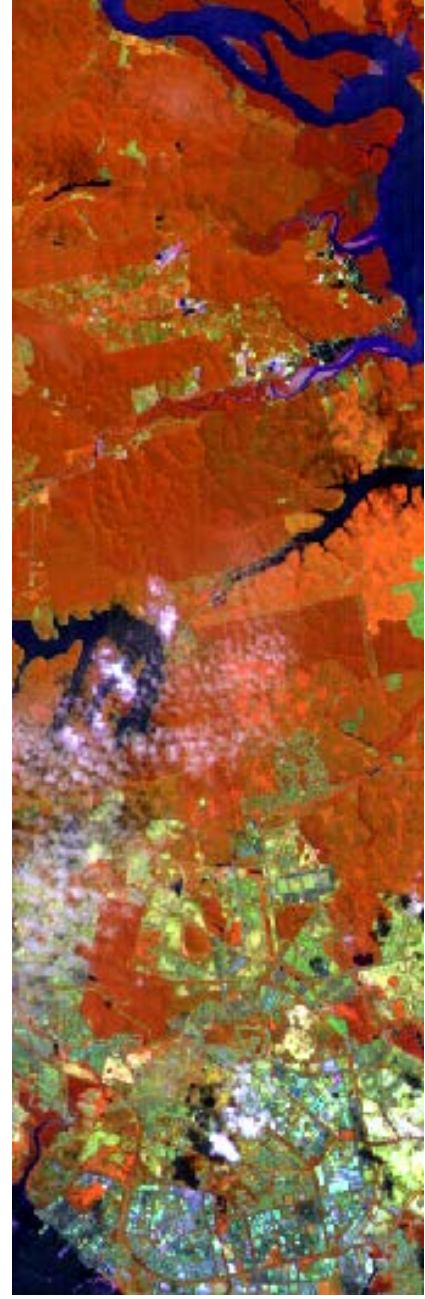


Fig. 5: True color composition of the Hyperion scene.

than those of the proposed PDF test statistics developed in complex case and which are very difficult (or impossible) to derive in real case. We have chosen sequentially six bands ( $m = 6$ ) in the complex representation. In this approach, both covariance matrix and mean vector are estimated using a sliding window of size  $5 \times 5$ , having  $N = 24$  secondary data.

The outcome of the detectors under  $\mathcal{H}_0$  hypothesis for this image are shown on the Fig. 7, Fig. 8 and Fig. 9 respectively. The results obtained on real HSI data on a Gaussian distributed region agree with the theoretical relationships presented above. Remark that the experimental false-alarm rate that can be achieved depends on the number of points on which the detector is calculated (in a similar manner to the Monte-Carlo

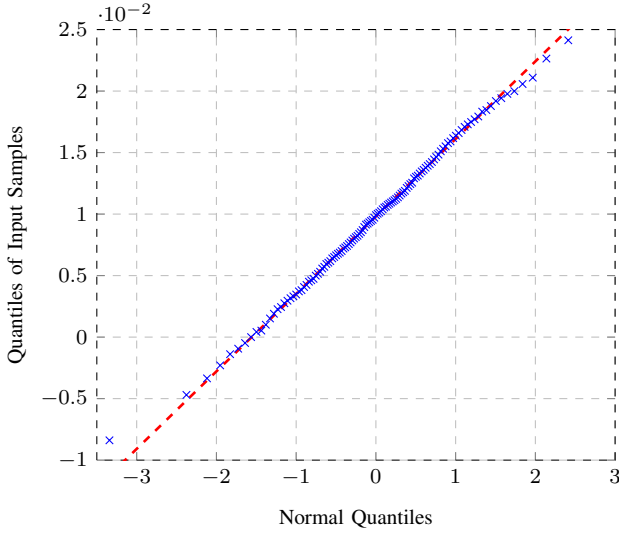


Fig. 6: Q-Q Plot of the data sample versus the Normal theoretical distribution.

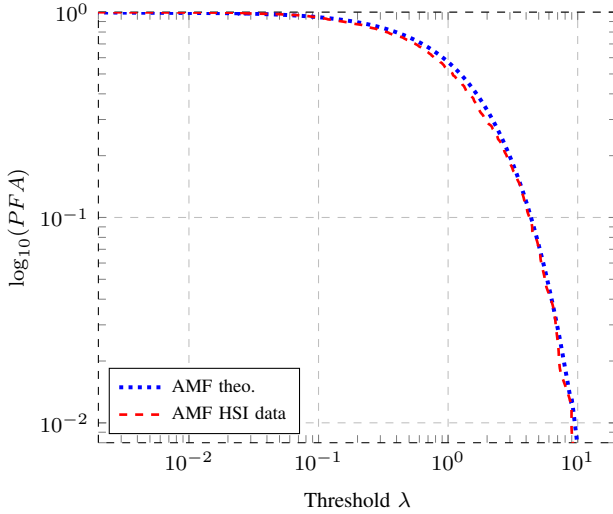


Fig. 7: AMF false-alarm regulation for a real HSI image.  $m = 6$ ,  $N = 24$ ,  $\mathbf{p} = [1, \dots, 1]^T$ .

trials). As the homogenous area is bounded and the data set is small, the distribution of the detectors may divert for small values of the PFA directly related to the size of the region.

Depending on the underlying material, the distribution of the detector might divert from the expected behavior when Gaussian distribution is assumed. This is the case on these real data since the extracted area is not perfectly Gaussian. This suggests the use of non-Gaussian distributions to model the background for hyperspectral imaging.

## V. CONCLUSION

Four adaptive detection schemes, the AMF, Kelly detectors with a "two-step GLRT" and a generalized versions as well as the ANMF, have been analyzed in the case where both the covariance matrix and the mean vector are unknown and need to be estimated. In this context, theoretical closed-form

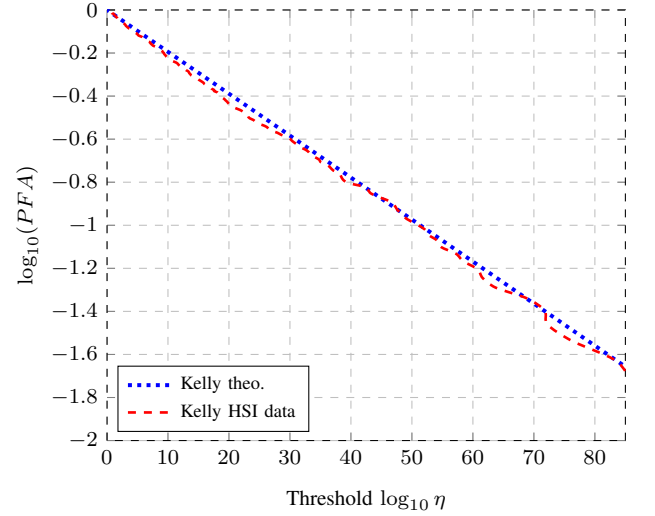


Fig. 8: Kelly false-alarm regulation for a real HSI image.  $m = 6$ ,  $N = 24$ ,  $\mathbf{p} = [1, \dots, 1]^T$ .

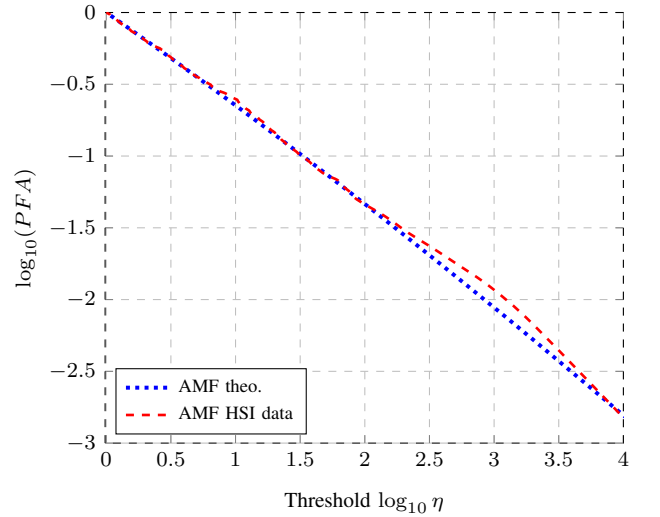


Fig. 9: ANMF false-alarm regulation for a real HSI image.  $m = 6$ ,  $N = 24$ ,  $\mathbf{p} = [1, \dots, 1]^T$ .

expressions for false-alarm regulation have been derived under Gaussian assumptions for the SCM-SMV estimates for three detection schemes. The resulting "PFA-threshold" expressions highlight the CFARness of these detectors since they only depend on the size and the number of data, but not on the unknown parameters. The theoretical analysis has been validated through Monte Carlo simulations and the performances of the detectors have been compared in terms of probability of detection. Finally, the analysis on experimental hyperspectral data validates the theoretical contribution through real application, in which a homogeneous subset of data has been extracted. But more generally, this work finds its purpose in signal processing methods for which both mean vector and covariance matrix are unknown.

## APPENDIX A COMPLEX NORMAL DISTRIBUTIONS

A  $m$ -dimensional vector  $\mathbf{x} = \mathbf{u} + j\mathbf{v}$  has a complex normal distribution with mean  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma} = E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^H]$ , denoted  $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , if  $\mathbf{z} = (\mathbf{u}^T, \mathbf{v}^T)^T \in \mathbb{R}^{2m}$  has a normal distribution [29]. If  $\text{rank}(\boldsymbol{\Sigma}) = m$ , the probability density function exists and is of the form

$$f_{\mathbf{x}}(\mathbf{x}) = \pi^{-m} |\boldsymbol{\Sigma}|^{-1} \exp\{-(\mathbf{x} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\}.$$

The resulting Maximum Likelihood Estimates (MLE) are the well-known SCM and SMV defined as:

$$\hat{\boldsymbol{\mu}}_{SMV} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i \quad \hat{\boldsymbol{\Sigma}}_{SCM} = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^H,$$

where the  $\mathbf{x}_i$  are IID  $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . Further, we shall denote the Centered SCM (CSCM) as:

$$\hat{\boldsymbol{\Sigma}}_{CSCM} = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^H. \quad (18)$$

## APPENDIX B WISHART DISTRIBUTION

Let  $\mathbf{x}_1, \dots, \mathbf{x}_N$  be an IID  $N$ -sample, where  $\mathbf{x}_i \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . Let us define  $\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}}_{SMV}$  and  $\hat{\mathbf{W}} = N \hat{\boldsymbol{\Sigma}}_{SCM}$  referred to as a Wishart matrix. Thus one has (see [30] for the real case):

- $\hat{\boldsymbol{\mu}}$  and  $\hat{\mathbf{W}}$  are independently distributed;
- $\hat{\boldsymbol{\mu}} \sim \mathcal{CN}(\boldsymbol{\mu}, \frac{1}{N} \boldsymbol{\Sigma})$ ;
- $\hat{\mathbf{W}} \sim \mathcal{CW}(N-1, \boldsymbol{\Sigma})$  is Whishart distributed with  $N-1$  degrees of freedom

## REFERENCES

- [1] C.-I. Chang, *Hyperspectral imaging: techniques for spectral detection and classification*. Springer, 2003, vol. 1.
- [2] N. Keshava and J. F. Mustard, "Spectral unmixing," *Signal Processing Magazine, IEEE*, vol. 19, no. 1, pp. 44–57, 2002.
- [3] J. C. Harsanyi and C.-I. Chang, "Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 32, no. 4, pp. 779–785, 1994.
- [4] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *Signal Processing Magazine, IEEE*, vol. 19, no. 1, pp. 29–43, 2002.
- [5] D. W. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker, "Anomaly detection from hyperspectral imagery," *Signal Processing Magazine, IEEE*, vol. 19, no. 1, pp. 58–69, 2002.
- [6] C.-I. Chang and S.-S. Chiang, "Anomaly detection and classification for hyperspectral imagery," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 40, no. 6, pp. 1314–1325, 2002.
- [7] H. Kwon and N. M. Nasrabadi, "Kernel matched subspace detectors for hyperspectral target detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 2, pp. 178–194, 2006.
- [8] S. Matteoli, M. Diani, and G. Corsini, "A tutorial overview of anomaly detection in hyperspectral images," *Aerospace and Electronic Systems Magazine, IEEE*, vol. 25, no. 7, pp. 5–28, 2010.
- [9] S. M. Kay, *Fundamentals of Statistical signal processing, Volume 2: Detection theory*. Prentice Hall PTR, 1998.
- [10] F. Gini, A. Farina, and M. Greco, "Selected list of references on radar signal processing," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 37, no. 1, pp. 329–359, 2001.
- [11] D. Manolakis, D. Marden, and G. Shaw, "Hyperspectral image processing for automatic target detection applications," *Lincoln Laboratory Journal*, vol. 14, no. 1, pp. 79–116, 2003.
- [12] I. Reed and X. Yu, "Adaptive multiple-band cfar detection of an optical pattern with unknown spectral distribution," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 38, no. 10, pp. 1760–1770, 1990.
- [13] P. C. Mahalanobis, "On the generalized distance in statistics," *Proceedings of the National Institute of Sciences (Calcutta)*, vol. 2, pp. 49–55, 1936.
- [14] E. J. Kelly, "An adaptive detection algorithm," *Aerospace and Electronic Systems, IEEE Transactions on*, no. 2, pp. 115–127, 1986.
- [15] S. Kraut, L. L. Scharf, and L. T. Mc Whorter, "Adaptive Subspace Detectors," *Signal Processing, IEEE Transactions on*, vol. 49, no. 1, pp. 1–16, January 2001.
- [16] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A cfar adaptive matched filter detector," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 28, no. 1, pp. 208–216, 1992.
- [17] S. Kraut and L. L. Scharf, "The CFAR adaptive subspace detector is a scale-invariant GLRT," *Signal Processing, IEEE Transactions on*, vol. 47, no. 9, pp. 2538–2541, 1999.
- [18] J. Frontera-Pons, F. Pascal, and J. Ovarlez, "False-alarm regulation for target detection in hyperspectral imaging," in *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2013 IEEE 5th International Workshop on*. IEEE, 2013, pp. 161–164.
- [19] D. Manolakis, R. Lockwood, T. Cooley, and J. Jacobson, "Is there a best hyperspectral detection algorithm?" in *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2009, pp. 733 402–733 402.
- [20] D. Manolakis, E. Truslow, M. Pieper, T. Cooley, M. Brueggeman, and S. Lipson, "The remarkable success of adaptive cosine estimator in hyperspectral target detection," in *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2013, pp. 874 302–874 302.
- [21] F. Gini and M. Greco, "Covariance matrix estimation for cfar detection in correlated heavy tailed clutter," *Signal Processing*, vol. 82, no. 12, pp. 1847–1859, 2002.
- [22] 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," *Signal Processing, IEEE Transactions on*, vol. 50, no. 8, pp. 1908–1915, August 2002.
- [23] J. Frontera-Pons, M. Mahot, J. Ovarlez, F. Pascal, and J. Chanussot, "Robust detection using  $M$ -estimators for hyperspectral imaging," in *Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, 2012.
- [24] F. Pascal, J. Ovarlez, P. Forster, and P. Larzabal, "On a SIRV-CFAR detector with radar experimentations in impulsive noise," in *Proc. European Signal Processing Conference, Florence, Italy*, vol. 134, September 2006.
- [25] M. E. Abramowitz et al., *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Courier Dover Publications, 1964, vol. 55.
- [26] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *Signal Processing, IEEE Transactions on*, vol. 42, no. 8, pp. 2146–2157, 1994.
- [27] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, New York, 1984.
- [28] C. D. Richmond, "Performance of a class of adaptive detection algorithms in nonhomogeneous environments," *Signal Processing, IEEE Transactions on*, vol. 48, no. 5, pp. 1248–1262, 2000.
- [29] A. van den Bos, "The multivariate complex normal distribution-a generalization," *Information Theory, IEEE Transactions on*, vol. 41, no. 2, pp. 537–539, 1995.
- [30] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*. Chapman & Hall/CRC, 2000.



# Hyperspectral Anomaly Detectors using Robust Estimators



# Hyperspectral Anomaly Detectors using Robust Estimators

Joana Frontera-Pons, *Student Member, IEEE*, Miguel A. Veganzones, *Member, IEEE*, Frédéric Pascal, *Senior Member, IEEE*, and Jean-Philippe Ovarlez, *Member, IEEE*

**Abstract**—Anomaly detection methods are used for target detection in which no a priori information about the spectra of the targets of interest is available. This paper reviews classical anomaly detection schemes such as the widely spread Reed-Xiaoli Detector and some of its variations. Moreover, the Mahalanobis distance based detector, rigorously derived from a Kelly's test-based approach, is analyzed and its exact distribution is derived when both mean vector and covariance matrix are unknown and have to be estimated. Although, most of these techniques are based on Gaussian distribution, we also propose here ways to extend them to non-Gaussian framework. For this purpose, elliptical distributions are considered for background statistical characterization. Through this assumption, this paper describes robust estimation procedures more suitable for non-Gaussian environment. We show that using them as plug-in estimators in anomaly detectors leads to some great improvement in the detection process. Finally, the theoretical contribution is validated through simulations and on real hyperspectral scenes.

**Index Terms**—Hyperspectral Imaging, anomaly detection, elliptical distributions,  $M$ -estimation.

## I. INTRODUCTION

**T**ARGET detection and anomaly detection of multidimensional signals have proved to be valuable techniques in a wide range of applications, including search-and-rescue, surveillance, rare mineral and land mines detection, etc (see for e.g. [1], [2], [3]). Target detection aims to discover the presence of a specific signal of interest among a set of signals. Statistical target detection is based on the Neyman-Pearson (NP) criterion, which maximizes the probability of detection for a given probability of false alarm.

Classical target detection methods require the knowledge of the spectra of the desired targets. One could be interested in a large number of possible targets each with different signatures. Thus, the variety of sought spectra corresponding to the different kind of targets and the difficulties due to the atmospheric compensation for the measured spectral signatures (used as steering vectors) have led to the derivation of new algorithms that intend to distinguish unusual materials in a scene without reference to target signatures. In this

work, we are focused on anomaly detection (see e.g. [4] and references therein). It can be interpreted as a particular case of target detection in which no a priori information about the spectra of the targets of interest is available. Hence, the aim of anomaly detection is to locate objects in the image that are anomalous with respect to the background. The type of interesting targets can differ significantly from one application to another, e.g. in forestry applications infected trees are the anomalies of interest, whereas in defense and intelligence applications, the anomalies to be detected are usually vehicles. Note that, since anomaly detectors do not use any a priori knowledge, they cannot distinguish between true targets and detections of bright pixels of the background or targets that are not of interest. This fact makes extremely difficult to define a false alarm rate for the detectors as highlighted in [5].

Anomalies are defined with reference to a model of the background. As for the previous target detection methods, the background model is developed adaptively using reference data. Most of these methods rely on the classical Gaussian distribution assumption and need for the statistical characterization of the background usually through first and second order parameters (i.e. the mean vector and the covariance matrix). In this case, the reference data are taken either from a local neighborhood around the observation vector either using all the pixels in the image. Both approaches have their benefits (see e.g. [6]). Local strategy provides more realistic scenario for the background characterization. However, it can be sensitive to the presence of false alarms due to isolated anomalies. While the global approach is not likely to generate this kind of false alarms, it will decrease the detection capability for isolated targets. Local procedures will be considered in the following for the different detection schemes. We consider on the first part of the paper the most popular Gaussian-based anomaly detectors (see e.g. [7] for a complete survey in anomaly detection methods).

However, the actual distribution of the background pixels differs from the theoretically predicted under Gaussian hypothesis. In fact, as stated in [8], the empirical distribution usually has heavier tails compared to the Gaussian distribution, and these tails strongly influence the observed false-alarm rate of the detector. Therefore, the class of Elliptical distributions is assumed for background statistics characterization and recalled in the second part of this work. The family of Elliptical distributions were originally introduced by Kelker

J. Frontera-Pons is with SONDR, Supelec, Plateau du Moulon, 3 rue Joliot-Curie, F-91190 Gif-sur-Yvette, France (e-mail: joana.fronterapons@supelec.fr)

M. A. Veganzones is with GIPSA-lab, CNRS, 11 rue des mathématiques, F-38402 Saint Martin d'Hères, France (e-mail: miguel-angel.veganzones@gipsa-lab.fr)

F. Pascal is with SONDR, Supelec, Plateau du Moulon, 3 rue Joliot-Curie, F-91190 Gif-sur-Yvette, France (e-mail: frederic.pascal@supelec.fr)

J.-P. Ovarlez is with ONERA, DEMR/TSI, Chemin de la Hunière, F-91120 Palaiseau, France (e-mail: jean-philippe.ovarlez@onera.fr)

in [9] and widely studied in [10]. They account for non-Gaussianity providing a long tailed alternative to multivariate normal model. They are proven to represent a more accurate characterization of HSI than models based on Gaussian assumption [8]. Therefore, the classical Gaussian-based estimators do not provide anymore optimal performance and the Fixed Point estimators (also known as Tyler's estimators [11]) are proposed for the parameters estimation. These can then be used as plug-in estimators in place of the unknown mean vector or/and of the covariance matrix in the detection scheme (see for e.g. [12], [13]). This is a simple but often efficient method to obtain robust properties for signal processors derived under the Gaussian assumption. One of the contributions of this work is to extend the results presented in [14].

More precisely, this paper provides a rigorous derivation of the Mahalanobis distance through a Kelly's test-based approach. Moreover, one of the theoretical contribution is the derivation of the exact distribution for the classical Mahalanobis-based anomaly detector when both mean vector and covariance matrix are unknown in Gaussian environment. Furthermore, robust estimation methods are considered in classical anomaly detection schemes and the improvement brought in most scenarios is pointed out.

This paper is organized as follows. Section II revisits classical anomaly detection schemes and provides the theoretical contribution of this paper by deriving the distribution of one of the detectors. Section III describes the family of elliptical distributions and the robust estimation methods studied in this paper for target detection purposes. Section IV illustrates the theoretical analysis through simulations and Section V reveals the theoretical improvement over real hyperspectral images. Finally Section VI concludes this work.

In the following, vectors (resp. matrices) are denoted by bold-faced lowercase letters (resp. uppercase letters).  $^T$  represents the transpose operator.  $|\mathbf{A}|$  represents the determinant of the matrix  $\mathbf{A}$  and  $\text{Tr}(\mathbf{A})$  its trace.  $j$  is used to denote the unit imaginary number.  $\sim$  means "distributed as".  $\Gamma(\cdot)$  denotes the gamma function. Eventually,  $\|\mathbf{x}\|$  represents the Euclidean norm of the vector  $\mathbf{x}$ .

## II. ADAPTIVE ANOMALY DETECTION METHODS

Before detailing the analysis of the corresponding detectors, let us recap the most common Gaussian-based estimators. Along with their well-known properties and their simplicity of analysis, the Sample Covariance Matrix (SCM) and the Sample Mean Vector (SMV) are the most extended estimates since they are the Maximum Likelihood Estimators (MLE) for Gaussian case, as shown in [15]:

$$\hat{\boldsymbol{\mu}}_{SMV} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i, \quad (1)$$

$$\hat{\boldsymbol{\Sigma}}_{SCM} = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{SMV})(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{SMV})^T. \quad (2)$$

Further, we shall denote the Centered SCM (CSCM) as:

$$\hat{\boldsymbol{\Sigma}}_{CSCM} = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T. \quad (3)$$

where  $N$  denotes the number of secondary data. However, such widespread techniques are suboptimal when the noise is a non-Gaussian stochastic process. Section III reviews some robust procedures particularly suited for estimating the covariance matrix and the mean vector of elliptical populations. Let us now detail the most popular Gaussian-based anomaly detectors.

### A. Reed-Xiaoli Detector

The original Reed-Xiaoli Detector (RXD) proposed in [16] is commonly considered as the benchmark anomaly detector for hyperspectral data. The considered signal model can be written as:

$$\begin{cases} \mathcal{H}_0 : \mathbf{x}_i = \mathbf{b}_i, & , i = 1, \dots, N \\ \mathcal{H}_1 : \mathbf{x}_i = \mathbf{p} \alpha_i + \mathbf{b}_i, & , i = 1, \dots, N, \end{cases}$$

where  $\mathbf{x}_i$  are the  $N$  available data vectors on the image of dimension  $m$ .  $\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$  represents the residual background,  $\mathbf{p}$  is the spectral signature of the possible anomalous material assumed to be unknown; and  $\alpha_i$  stands for the amplitude of the intended targets through the  $N$  available data, i.e. it is a known vector  $\boldsymbol{\alpha}$  of dimension  $N$  that indicates the strength and position of the sought targets over the image. Remark that each vector from the available data can potentially contain an anomaly while in classical detection problem secondary data are assumed to be signal free. Thus, one can arrange the vector data into a matrix as  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$ , and the detection scheme derived in [16] takes the form:

$$\Lambda(\mathbf{X}) = \frac{(\mathbf{X}\boldsymbol{\alpha}^T)^T (\mathbf{X}\mathbf{X}^T)^{-1} (\mathbf{X}\boldsymbol{\alpha}^T)}{\boldsymbol{\alpha}\boldsymbol{\alpha}^T} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda.$$

Since hyperspectral data are not zero mean, let us now consider that the background  $\mathbf{b}_i$  is distributed according to  $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  and the mean vector  $\boldsymbol{\mu}$  is supposed to be known. In the case just one anomaly in the data under test is intended to be detected, the corresponding amplitude vector can be written as  $\boldsymbol{\alpha}_i = [0 \dots 0 \ 1 \ 0 \dots 0]^T$  where 1 is at the  $i^{th}$  position and the previous detector takes the form:

$$\Lambda_{RXD} = (\mathbf{x}_i - \boldsymbol{\mu})^T \hat{\boldsymbol{\Sigma}}_{CSCM}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda.$$

Finally, since the mean vector is usually unknown, it can be replaced on the detector in by its estimate  $\hat{\boldsymbol{\mu}}_{SMV}$ . The resulting detector, commonly known as two-step Generalized Likelihood Ratio Test, yields:

$$\Lambda_{ARXD} = (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{SMV})^T \hat{\boldsymbol{\Sigma}}_{SCM}^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{SMV}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda. \quad (4)$$

The covariance matrix estimation  $\hat{\boldsymbol{\Sigma}}_{SCM}$  in Eq. (4), is performed over all the data set, i.e. including the vector  $\mathbf{x}_i$  under test. In the following, the test in Eq. (4) will be referred as the Adaptive RXD (ARXD), to underline the fact that the unknown mean vector is replaced by its estimate.

### B. Kelly Anomaly Detector

We detail here a classical anomaly detector often mistakenly referred as the RXD. Following the development proposed in [17], let us now assume the following signal model:

$$\begin{cases} \mathcal{H}_0 : \mathbf{x} = \mathbf{b}, & \mathbf{x}_i = \mathbf{b}_i, i = 1, \dots, N \\ \mathcal{H}_1 : \mathbf{x} = \alpha \mathbf{p} + \mathbf{b}, & \mathbf{x}_i = \mathbf{b}_i, i = 1, \dots, N, \end{cases}$$

and the  $\mathbf{b}_1, \dots, \mathbf{b}_N$  are assumed to an independent identically distributed (IID) sample from a Gaussian distribution  $\mathbf{b}_i \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . As in classical Kelly detector, the covariance matrix  $\boldsymbol{\Sigma}$  is unknown and the mean vector  $\boldsymbol{\mu}$  is supposed to be known. However for anomaly detector derivation, the amplitude of the signal  $\alpha$  is supposed to be known and the unknown parameter is now the steering vector  $\mathbf{p}$ . Therefore,  $N+1$   $m$ -dimensional vectors are observed under each hypothesis and the joint probability density function (p.d.f.) of the the  $N$  secondary data and the observation vector  $\mathbf{x}$  under the two hypotheses  $\mathcal{H}_i$  can be written as:

$$f_i(\mathbf{x}) = \left( \frac{1}{2\pi^{\frac{m}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2} \text{Tr}(\boldsymbol{\Sigma}^{-1} \mathbf{T}_i)\right] \right)^{N+1}, \quad (5)$$

where  $\mathbf{T}_i$  is the composite sample covariance matrix constructed from both the secondary data and observation vector:

$$\begin{aligned} \mathbf{T}_0 &= \frac{1}{N+1} \left( (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T + \hat{\mathbf{W}} \right), \\ \mathbf{T}_1 &= \frac{1}{N+1} \left( (\mathbf{x} - (\alpha \mathbf{p} + \boldsymbol{\mu}))(\mathbf{x} - (\alpha \mathbf{p} + \boldsymbol{\mu}))^T + \hat{\mathbf{W}} \right), \end{aligned}$$

and  $\hat{\mathbf{W}} = N \hat{\boldsymbol{\Sigma}}_{CSCM}$ . The first step is to maximize with respect to (w.r.t) the unknown covariance matrix  $\boldsymbol{\Sigma}$ . Thus, the matrix maximizing the PDF  $f_i$  is simply  $\mathbf{T}_i$ . When this estimator is replaced in the PDF, one obtains:

$$\max_{\boldsymbol{\Sigma}} f_i = \left( \frac{1}{(\pi e)^m |\mathbf{T}_i|} \right)^{\frac{N+1}{2}}. \quad (6)$$

and the GLRT, neglecting the exponent  $(N+1)/2$  is given by:

$$\Lambda(\mathbf{x}, \mathbf{p}) = \frac{|\mathbf{T}_0|}{|\mathbf{T}_1|} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta. \quad (7)$$

It remains to maximize this expression over the unknown spectral signature  $\mathbf{p}$  and the resulting MLE takes the form:

$$\hat{\mathbf{p}} = \frac{\mathbf{x} - \boldsymbol{\mu}}{\alpha}. \quad (8)$$

After replacing  $\mathbf{p}$  by Eq. (8) in Eq. (7), it is easy to show that the resulting GLRT test is equivalent to:

$$\Lambda_{KellyAD \hat{\boldsymbol{\Sigma}}}(\mathbf{x} - \boldsymbol{\mu})^T \hat{\boldsymbol{\Sigma}}_{CSCM}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda. \quad (9)$$

The quadratic form in Eq. (9) corresponds to the Mahalanobis distance detailed in [18]. It performs statistically as an outlier detector. When Gaussian assumption is valid, the quadratic form  $(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$  follows a  $\chi^2$ -distribution with  $m$  degrees of freedom for  $\boldsymbol{\Sigma}$  and  $\boldsymbol{\mu}$  perfectly known. In case the parameter  $\boldsymbol{\Sigma}$  is replaced by its MLE, the CSCM, the

distribution of the quadratic form can be written according to (see [19]):

$$\Lambda_{KellyAD \hat{\boldsymbol{\Sigma}}}^{(N)} = (\mathbf{x} - \boldsymbol{\mu})^T \hat{\boldsymbol{\Sigma}}_{CSCM}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \sim T^2, \quad (10)$$

becomes a Hotelling  $T^2$  distribution and thus,

$$\frac{N-m+1}{mN} \Lambda_{KellyAD \hat{\boldsymbol{\Sigma}}}^{(N)} \sim F_{m, N-m+1} \quad (11)$$

where  $F_{m, N-m+1}$  is the non-central  $F$ -distribution with  $m$  and  $N-m+1$  degrees of freedom [20] and the superscript  $(N)$  is used to stress the dependence on the number of secondary data  $N$ . For high values of  $N$ , ( $N > 10m$ ), the distribution can be approximated by the  $\chi^2$ -distribution.

As discussed above, when the mean vector is unknown, it can be replaced on the detector (two-step GLRT) by its MLE leading to:

$$\Lambda_{KellyAD \hat{\boldsymbol{\Sigma}}, \hat{\boldsymbol{\mu}}}^{(N)} = (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})^T \hat{\boldsymbol{\Sigma}}_{SCM}^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda. \quad (12)$$

**Remark II.1.** Interestingly, note that  $\Lambda_{RXD}$  (resp.  $\Lambda_{ARXD}$ ) and the  $\Lambda_{KellyAD \hat{\boldsymbol{\Sigma}}}$  (reps.  $\Lambda_{KellyAD \hat{\boldsymbol{\Sigma}}, \hat{\boldsymbol{\mu}}}^{(N)}$ ) differ only on the fact that the vector  $\mathbf{x}$  under test is also present in the covariance matrix estimation in Eq. (4). Therefore, in  $\Lambda_{RXD}$ , the  $N$  secondary data are not assumed to be signal free and the proposed detector aims to compare every sample to the covariance matrix over all the samples. While in the second approach,  $\Lambda_{KellyAD \hat{\boldsymbol{\Sigma}}}$ , one intends to differentiate the observation vector from the background statistically characterized using  $N$  samples. Hence,  $N+1$  vectors are available in the latter and  $\Lambda_{KellyAD \hat{\boldsymbol{\Sigma}}}$  does not represent anymore a benchmark structure. Often, the local Kelly detector is mistakenly referred as the local RXD when the users, either remove the vector  $\mathbf{x}_i$  from the secondary data or they prevent it to be part of this set by using a guard window.

The distribution of this detection test is given in the next Proposition.

**Proposition II.1.** The distribution of the detector under Gaussian assumption is given by

$$\frac{N-m}{m(N+1)} \Lambda_{KellyAD \hat{\boldsymbol{\Sigma}}, \hat{\boldsymbol{\mu}}}^{(N)} \sim F_{m, N-m}, \quad (13)$$

with  $F_{m, N-m}$  is the non-central  $F$ -distribution with  $m$  and  $N-m$  degrees of freedom.

*Proof:* For simplicity matters, the following notations are used:  $\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Sigma}}_{SCM}$  and  $\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}}_{SMV}$ .

Let us set  $\forall i = 1, \dots, N$ ,  $\mathbf{x}_i \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  and  $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , where all these vectors are independent. Now, let us denote

$$\hat{\mathbf{W}}_{N-1} = \sum_{i=1}^N (\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T = N \hat{\boldsymbol{\Sigma}}_{SCM}.$$

Note that as an application of the Cochran theorem (see e.g. [21]), one has

$$\hat{\mathbf{W}}_{N-1} \stackrel{dist.}{=} \sum_{i=1}^{N-1} (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T = (N-1) \hat{\boldsymbol{\Sigma}}_{CSCM},$$



where  $\stackrel{dist.}{=}$  means *is distributed as*.

Since  $\hat{\boldsymbol{\mu}} \sim \mathcal{N}(\boldsymbol{\mu}, \frac{1}{N}\boldsymbol{\Sigma})$ , one has  $\mathbf{x} - \hat{\boldsymbol{\mu}} \sim \mathcal{N}(\mathbf{0}, \frac{N+1}{N}\boldsymbol{\Sigma})$ . This can be equivalently rewritten as

$$\mathbf{y} = \sqrt{N/(N+1)}(\mathbf{x} - \hat{\boldsymbol{\mu}}) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}).$$

As we jointly estimate the mean and the covariance matrix, a degree of freedom is lost, compared with the only covariance matrix estimation problem.

Let us now consider  $\Lambda_{KellyAD, \hat{\boldsymbol{\Sigma}}}^{(N-1)}$  (i.e.  $\boldsymbol{\mu}$  known) built from  $N-1$  secondary data, rewritten in terms of  $\hat{\mathbf{W}}_{N-1}$ :

$$\Lambda_{KellyAD, \hat{\boldsymbol{\Sigma}}}^{(N-1)} = (N-1) \left( (\mathbf{x} - \boldsymbol{\mu})^T \hat{\mathbf{W}}_{N-1}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right)$$

where  $(\mathbf{x} - \boldsymbol{\mu}) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$  and whose distribution is given by Eq. (11) where  $N$  is replaced by  $N-1$ .

Now, for the joint estimation problem, the  $\Lambda_{KellyAD, \hat{\boldsymbol{\Sigma}}, \hat{\boldsymbol{\mu}}}$  can be rewritten as:

$$\begin{aligned} \Lambda_{KellyAD, \hat{\boldsymbol{\Sigma}}, \hat{\boldsymbol{\mu}}}^{(N)} &= N \left( (\mathbf{x} - \hat{\boldsymbol{\mu}})^T \hat{\mathbf{W}}_{N-1}^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}) \right) \\ &= N \frac{N+1}{N} \left( \mathbf{y}^T \hat{\mathbf{W}}_{N-1}^{-1} \mathbf{y} \right) \\ &\stackrel{dist.}{=} \frac{N+1}{N-1} \Lambda_{KellyAD, \hat{\boldsymbol{\Sigma}}}^{(N-1)} \end{aligned}$$

This concludes the proof.  $\blacksquare$

The "PFA-threshold" relationship is easily obtained as the complementary cumulative density function (c.d.f.) of the detector distribution.

It is worth pointing out from Eq. (12) that  $\Lambda_{KellyAD, \hat{\boldsymbol{\Sigma}}, \hat{\boldsymbol{\mu}}}$  performs similarly to a matched filter structure applied to  $\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}$ :

$$\Lambda(\mathbf{x}) = c \mathbf{H}^T (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}), \quad (14)$$

where  $\mathbf{H}^T$  is the matched signal and  $c$  a constant that can be also a function on  $\mathbf{x}$ . The expression in Eq. (14) is completely characterized by the matched signal  $\mathbf{H}^T$  and the scale constant  $c$ . Hence, one can identify from Eq. (12) the matched signal  $\mathbf{H}^T = (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})^T \hat{\boldsymbol{\Sigma}}_{SCM}^{-1}$  and  $c = 1$ .

### C. Normalized-RXD and Uniform Target Detector

Following the same approach than in Eq. (14), one can derive many different anomaly detection schemes. We recall here two popular variants of the Mahalanobis distance described in [4]: the Normalized-RXD (N-RXD) and the Uniform Target Detector (UTD).

The N-RXD takes the form:

$$\Lambda_{N-RXD} = \frac{(\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})^T}{\|\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}\|} \hat{\boldsymbol{\Sigma}}_{SCM}^{-1} \frac{(\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})^T}{\|\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}\|} \stackrel{\mathcal{H}_1}{\underset{\mathcal{H}_0}{\gtrless}} \lambda, \quad (15)$$

where  $\|\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}\|^2 = (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})^T (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})$  stands for the Euclidean norm of the vector. The detection test in Eq. (15) can be immediately identified as the normalized version of  $\Lambda_{KellyAD}$ . In addition,  $\Lambda_{N-RXD}$  takes also the form of a matched filter specified in Eq. (14) with matched signal  $\mathbf{H}^T = (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})^T \hat{\boldsymbol{\Sigma}}_{SCM}^{-1}$  the same as in Eq. (9) and a different scale constant  $c = \|\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}\|^{-2}$ .

The UTD is another widespread anomaly detection test. It was firstly introduced in [22] and can be defined as:

$$\Lambda_{UTD} = (\mathbf{1} - \hat{\boldsymbol{\mu}}_{SMV})^T \hat{\boldsymbol{\Sigma}}_{SCM}^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}) \stackrel{\mathcal{H}_1}{\underset{\mathcal{H}_0}{\gtrless}} \lambda. \quad (16)$$

with  $\mathbf{1} = [1, \dots, 1]^T$  is the  $m$ -dimensional unity vector. Once again the detector in Eq. (16) can be interpreted as a matched filter where  $\mathbf{H}^T = (\mathbf{1} - \hat{\boldsymbol{\mu}}_{SMV})^T \hat{\boldsymbol{\Sigma}}_{SCM}^{-1}$  is the matched signal. If there is no a priori information about the target spectra, the non-prior approach is the one that does not introduce any information into the detector and consists on assuming uniform distribution for the spectra over all the bands.

### D. Generalized Kelly Anomaly Detector

In the previous detection schemes, it has not been taken into account in the derivation of the test that both mean vector  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$  were unknown. One simply replaced the mean vector by a plug-in estimate in the detector (two-step GLRT). In case both covariance matrix and mean vector are unknown, we need to derive a new detector. This strategy is similar to the one proposed in [23] for the generalized Kelly detection test. The likelihood functions under  $\mathcal{H}_0$  and  $\mathcal{H}_1$  are given in (5). Under  $\mathcal{H}_0$  and  $\mathcal{H}_1$ , the maxima are achieved at

$$\max_{\boldsymbol{\Sigma}, \boldsymbol{\mu}} f_i = \left( \frac{1}{(\pi e)^m |\mathbf{T}_i|} \right)^{\frac{N+1}{2}}, \quad \text{for } i = 0, 1,$$

where

$$\begin{aligned} (N+1)\mathbf{T}_0 &= (\mathbf{x} - \boldsymbol{\mu}_0)(\mathbf{x} - \boldsymbol{\mu}_0)^T + \sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu}_0)(\mathbf{x}_i - \boldsymbol{\mu}_0)^T, \\ (N+1)\mathbf{T}_1 &= (\mathbf{x} - \alpha \mathbf{p} - \boldsymbol{\mu}_1)(\mathbf{x} - \alpha \mathbf{p} - \boldsymbol{\mu}_1)^T \\ &\quad + \sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu}_1)(\mathbf{x}_i - \boldsymbol{\mu}_1)^T, \end{aligned}$$

and

$$\boldsymbol{\mu}_0 = \frac{1}{N+1} \left( \mathbf{x} + \sum_{i=1}^N \mathbf{x}_i \right), \quad (17)$$

$$\boldsymbol{\mu}_1 = \frac{1}{N+1} \left( \mathbf{x} - \alpha \mathbf{p} + \sum_{i=1}^N \mathbf{x}_i \right). \quad (18)$$

Following the same lines than in [23], one has to maximize the LR in Eq. (7) w.r.t.  $\mathbf{p}$ . This is obtained by taking:

$$\hat{\mathbf{p}} = \frac{N+1}{N} \frac{(\mathbf{x} - \boldsymbol{\mu}_0)}{\alpha}. \quad (19)$$

Hence, the resulting detector can be written according to:

$$\Lambda_{G-KellyAD} = (\mathbf{x} - \boldsymbol{\mu}_0)^H \mathbf{S}_0^{-1} (\mathbf{x} - \boldsymbol{\mu}_0) \stackrel{\mathcal{H}_1}{\underset{\mathcal{H}_0}{\gtrless}} \lambda, \quad (20)$$

where  $\mathbf{S}_0 = \sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu}_0)(\mathbf{x}_i - \boldsymbol{\mu}_0)^H$ , and  $\boldsymbol{\mu}_0 = \frac{1}{N+1} \left( \mathbf{x} + \sum_{i=1}^N \mathbf{x}_i \right)$ . Once again the mean vector estimate  $\boldsymbol{\mu}_0$  and the covariance matrix  $\mathbf{S}_0$  depend on the data under test  $\mathbf{x}$ . Hence,  $\mathbf{x} - \boldsymbol{\mu}_0$  and  $\mathbf{S}_0$  are not independent. Remark that

one can write  $(\mathbf{x} - \boldsymbol{\mu}_0) = \frac{N}{N+1}(\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})$ . Neglecting the multiplicative constants, the test in Eq. (20) appears to be equivalent to the classical  $\Lambda_{RXD}$  obtained throughout a different approach but built with  $N+1$  available data.

### III. ROBUST ANOMALY DETECTION

In this section, the class of elliptical distributions and robust estimation procedures are reviewed.

#### A. Elliptical Distributions

Hyperspectral data have been proven not to be multivariate normal but long tailed distributed [8]. In order to take into account these features, the class of elliptical distributions is considered to describe clutter statistical behavior (see for e.g. [10], [24] for a complete survey on elliptical distributions). An  $m$ -dimensional random complex vector  $\mathbf{x}$  has a multivariate elliptical distribution if its characteristic function is of the form:

$$\Phi_{\mathbf{x}}(\mathbf{c}) = \exp(j\mathbf{c}^T \boldsymbol{\mu}) \phi\left(\frac{1}{2}\mathbf{c}^T \boldsymbol{\Sigma} \mathbf{c}\right), \quad (21)$$

for some function  $\phi: \mathbb{R}^+ \rightarrow \mathbb{R}$ , called characteristic generator, a positive semidefinite matrix  $\boldsymbol{\Sigma}$ , called *scatter* matrix and  $\boldsymbol{\mu} \in \mathbb{C}^m$  the *location* vector. We shall write  $\mathbf{x} \sim \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ . From Eq. (21), it does not follow that  $\mathbf{x}$  has a p.d.f.  $f_{\mathbf{x}}(\cdot)$ , but if exists, it has the form:

$$f_{\mathbf{x}}(\mathbf{x}) = c_{m,h} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} h_m\left(\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right), \quad (22)$$

where  $c_{m,h}$  is a normalization constant and  $h_m(\cdot)$  is any function such as Eq. (22) defines a p.d.f. in  $\mathbb{R}^m$ . The function  $h_m$  is usually called density generator and it is assumed to be only approximately known. In this case, we shall write  $\mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h_m)$  instead of  $\mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ . Remark that the p.d.f. in Eq. (22) depends on  $\mathbf{x}$  only through the quadratic form  $(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$ . Thus, the level sets of the density  $f_{\mathbf{x}}(\mathbf{x})$  are ellipsoids in the Euclidean  $m$ -space.

If the second-order moment exists, then  $\boldsymbol{\Sigma}$  reflects the structure of the covariance matrix of the elliptically distributed random vector  $\mathbf{x}$ , i.e. the covariance matrix is equal to the scatter matrix up to a scalar constant. It serves to characterize the correlation structure existing within the spectral bands. It is worth pointing out that the family of elliptical distributions includes a large number of distributions, notably the Gaussian distribution, multivariate  $t$ -distribution,  $K$ -distribution or multivariate Cauchy. Thus, it allows for heterogeneity of the background power with the texture.

#### B. Robust parameters estimation

1) *Fixed Point Estimators*: The Fixed Point (FP) estimators, according to the definition proposed by Tyler in [11], satisfy

the following equations:

$$\hat{\boldsymbol{\mu}}_{FP} = \frac{\sum_{i=1}^N \frac{\mathbf{x}_i}{\left((\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP})^T \hat{\boldsymbol{\Sigma}}_{FP}^{-1}(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP})\right)^{1/2}}}{\sum_{i=1}^N \frac{1}{\left((\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP})^T \hat{\boldsymbol{\Sigma}}_{FP}^{-1}(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP})\right)^{1/2}}} \quad (23)$$

$$\hat{\boldsymbol{\Sigma}}_{FP} = \frac{m}{N} \sum_{i=1}^N \frac{(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP})(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP})^T}{((\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP})^T \hat{\boldsymbol{\Sigma}}_{FP}^{-1}(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP}))} \quad (24)$$

The FP estimates have been widely investigated in statistics and signal processing literature. We refer to [25] for a detailed performance analysis. It is worth pointing out that  $\hat{\boldsymbol{\Sigma}}_{SCM}$  and  $\hat{\boldsymbol{\Sigma}}_{FP}$  have the same asymptotic Gaussian distribution which differs on their second order moment by a factor  $\frac{m+1}{m}N$ , i.e. for  $N$  sufficiently large,  $\hat{\boldsymbol{\Sigma}}_{FP}$  behaves as a Wishart matrix with  $\frac{m}{m+1}$  degrees of freedom. Indeed, these estimators belong to the wider class of robust  $M$ -estimators.

2) *Shrinkage estimators*: We present now shrinkage methods that are suitable for high dimensional problems with small number of samples (large  $m$  small  $N$ ). In these "large  $m$  small  $N$ " problems, classical estimators suffer from a distorted eigen-structure and improved estimators are required.

A common regularization approach has been widely studied, the Diagonal Loading (DL) approach introduced in [26], [27]. Thus, in Gaussian context the regularized SCM takes the form:

$$\hat{\mathbf{M}}_{SCM-DL}(\beta) = \frac{1-\beta}{N} \sum_{i=1}^N (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{SMV})(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{SMV})^T + \beta \mathbf{I}_m. \quad (25)$$

In presence of non-Gaussian, impulsive background the estimate in Eq. (25) suffers from the same drawbacks than the SCM and the class of robust estimates are more appropriate. Yet, the FP estimators described above exhibit important shortcomings in high dimensional context and they can not be computed for the undersampling case when  $m > N$ .

Moreover, we extend here FP covariance matrix estimator to the high dimensional setting using shrinkage regularization. Let us consider now the shrinkage FP introduced in [28] and defined as the solution of the following fixed point equation:

$$\hat{\mathbf{M}}_{FP}(\beta) = (1-\beta) \frac{m}{N} \sum_{i=1}^N \frac{(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP})(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP})^T}{((\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP})^T \hat{\mathbf{M}}_{FP}^{-1}(\beta)(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP}))} + \beta \mathbf{I}, \quad (26)$$

for  $\beta \in (0, 1]$  and  $\hat{\boldsymbol{\mu}}_{FP}$  the same as given in Eq. (23). It was shown in [28] that when  $\beta$  tends to 0, the proposed shrinkage estimator in Eq. (26) tends to the FP estimator in Eq. (24) whose inverse has its trace equal to  $m$ . A different approach that introduces a normalization constraint in the algorithm for the shrinkage FP estimates is found in [29]. Moreover, in [30], [31], [32], this estimator has been used within the Expected Likelihood framework. The optimization of the shrinkage parameter  $\beta$  has been discussed in [33].

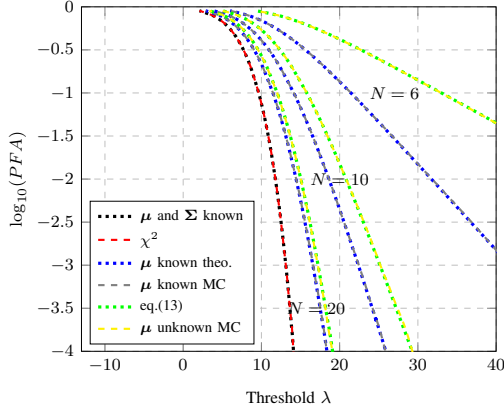


Fig. 1: PFA versus threshold for the  $\Lambda_{Kelly\ AD}$  when (1)  $\mu$  and  $\Sigma$  are known (Mahalanobis) (red and black curves) (2) only  $\mu$  is known (gray and blue curves) (3) Proposition II.1: both  $\mu$  and  $\Sigma$  are unknown (yellow and green curves).

The basis of the proposed method are the FP estimators. However, the approach presented here could be extended to other  $M$ -estimators.

#### IV. SIMULATIONS

In this section, we validate the theoretical analysis on simulated data. Firstly, we validate through Monte-Carlo simulations the distribution of  $\Lambda_{Kelly\ AD}$  detailed above. The experiments have been conducted on Gaussian vectors of dimension  $m = 5$  and for different values of  $N$ . The computations have been made through  $10^6$  Monte-Carlo trials. The true covariance is chosen as a Toeplitz matrix whose entries are  $\Sigma_{i,j} = \rho^{|i-j|}$  and where  $\rho = 0.4$ . The mean vector is arbitrary set to have all entries equal to 3. Exceedance plot shows the fraction of points in the data set whose Mahalanobis distance is larger than the indicated value. This is essentially a cumulative histogram of Mahalanobis distance values which correspond to the "PFA-threshold" relationship. Remark that the definition of false alarms is not unique and it depends on the application. Thus, we will rather refer to the distribution of the detector in target absent hypothesis.

Fig. 1 illustrates the distribution of the detector under null hypothesis. The case where both covariance matrix and mean vector are perfectly known corresponds to the  $\chi^2$ -distribution and the adaptive versions of the quadratic form become a  $T^2$  Hotelling. The perfect agreement of the green and yellow curves bears out the results of Proposition II.1.

Furthermore, we compare the five proposed anomaly detectors in terms of probability of detection. The experiments were on Gaussian vectors of dimension  $m = 5$ , for  $N = 10$  and the artificial targets signature used for the simulations is the unity vector  $\mathbf{p} = [1, \dots, 1]^T$ . On a first step, the threshold is determined to ensure a  $PFA = 10^{-3}$  obtained empirically from the data. The best results are obtained for the Mahalanobis-based detectors, i.e. the classical RXD, Kelly AD and generalized Kelly AD. The two detectors derived according to Kelly's

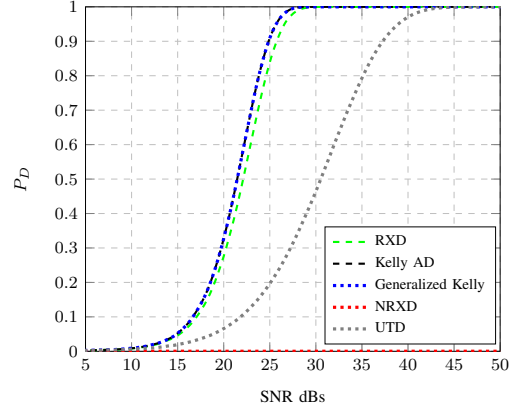


Fig. 2: Probability of detection for different SNR values and  $PFA = 10^{-3}$  in Gaussian case.

approach perform fundamentally the same and slightly better than the RXD. This improvement may be due to the fact that  $N + 1$  data are available for the Kelly's strategies, while only  $N$  samples are used in the classical RXD. The matched filter based detectors deliver poor performance in the case of the UTD, as the matching signal is the unity vector, which is shown to be not optimal even in the case the artificial targets signature used for the simulations is the unity vector. The NRXD presents an almost flat curve as the normalization factor grows as the SNR gets higher. The outcome of this detector can be assimilated to the residual background level and its use in Gaussian environment should be avoided.

#### V. REAL HYPERSPECTRAL DATA

##### A. Gaussian Background

The same experiments that in simulations have been conducted on a real hyperspectral image. The scene analyzed is the NASA Hyperion sensor dataset displayed in Fig. 3. The image is constituted of  $798 \times 253$  pixels and 116 spectral bands after water absorption bands have been removed. The analysis has been done on a homogenous part of the image corresponding to the water region on the top left of the image. The part extracted consists on  $60 \times 20$  pixels. In order to ensure the validity of the proposed methods, we show in Fig. 4 the outcome of a classical Gaussianity test "Q-Q plot" for the selected region over the band 42. Even if this allows to "validate" the Gaussianity of each band, it cannot ensure the Gaussianity of the corresponding multivariate vector.

To avoid the well-known problem due to high dimensionality, we have chosen sequentially six bands in the complex representation. In this approach, both covariance matrix and mean vector are estimated using a sliding window of size  $5 \times 5$ , having  $N = 24$  secondary data. Fig. 5 shows the distribution of the  $\Lambda_{Kelly\ AD, \hat{\Sigma}, \hat{\mu}}$  in real hyperspectral data (red curve). We also plot the theoretical relationship defined by Eq. (13). The results obtained on real HSI data on a Gaussian distributed region agree with the theoretical relationships presented above.

Finally, we illustrate the detection capability of the proposed methods when artificial anomalies with known spectral



Fig. 3: True color composition of the Hyperion scene.

signature are inserted on the real hyperspectral image. For this purpose, we extract the spectral signature from ground materials in Fig. 3 and the anomaly spectra is depicted in Fig. 6. Fig. 7 (a) details the position and the shape of the targets. For a fixed value of FA  $PFA = 0.1$ , we present in Fig. 7 the outcome of the different detectors. The detectors based on the Mahalanobis distance deliver best results for detection purposes as expected and the matched filter based detectors do not detect properly the artificial targets. These detection maps are in agreement with the SNR figure detailed above. Remark that the two-pixel targets are not detected by any of the detection schemes. This problem is due to the presence of a strong target in the secondary data that pollutes the covariance matrix estimation. Its occurrence has a significant

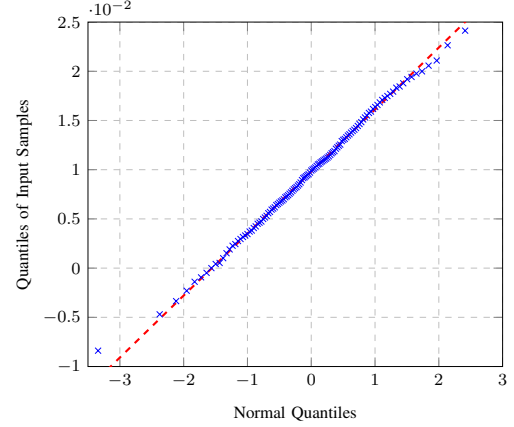


Fig. 4: Q-Q Plot of the data sample versus the Normal theoretical distribution.

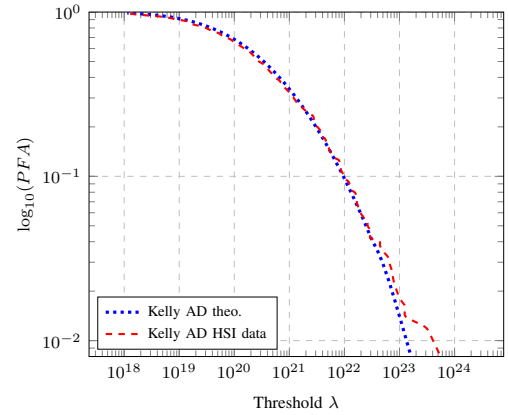


Fig. 5: Kelly AD complementary CDF of the Mahalanobis distance for a real HSI image

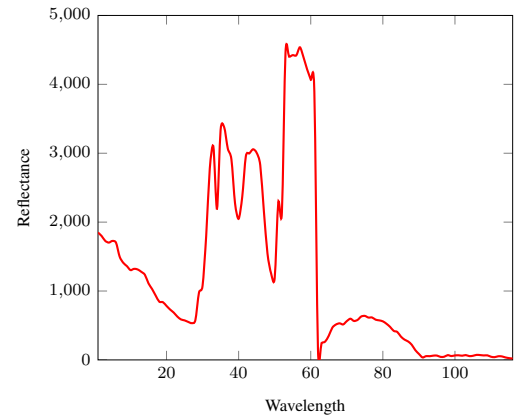


Fig. 6: Endmember used in the experiment which corresponds to land materials.

impact on the detection process and it degrades the detectors' performances.

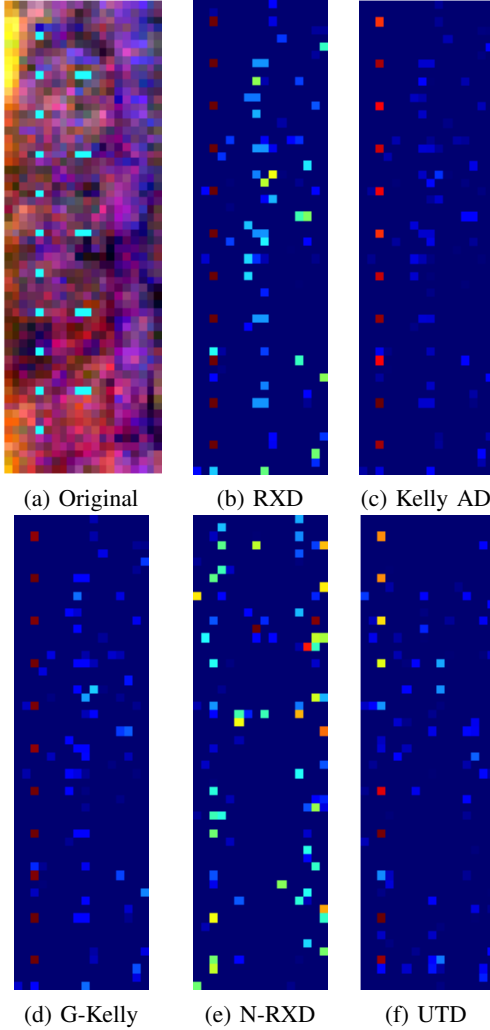


Fig. 7: Anomaly detection for artificial targets in real HSI.

Fig. 8 shows the results of the  $\Lambda_{Kelly AD}$  built now with FP estimators and with Shrinkage estimators, both SCM and FP. The same value of FA  $PFA = 10^{-1}$  is considered. Remark that all the anomalies of interest are now detected even those bigger than one pixel. Thereafter, the proposed estimation methods allow for better detection results in Gaussian case.

### B. Non-Gaussian Background

Let us now present some results on a real hyperspectral image in which the background can not be characterized with Gaussian distribution and artificial targets were introduced as anomalies. The original data set consists on  $50 \times 50$  pixels with 126 bands, from which we have chosen sequentially nine bands, see Fig. 9 (a). For this example, both covariance matrix and mean vector are estimated using a sliding window of size  $9 \times 9$  having  $N = 80$  secondary data. The results for the  $\Lambda_{Kelly AD}$  built with classical SMV-SCM, FP estimates and shrinkage estimators are shown in Fig 9, the FA is fixed

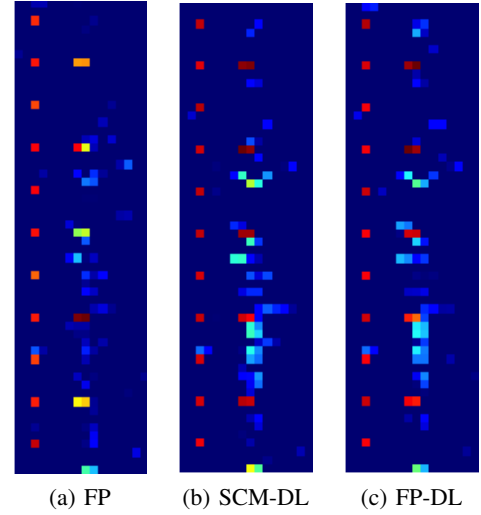


Fig. 8: Kelly AD built with robust estimates for artificial targets in real HSI.

at a value of  $PFA = 0.03$ . In this case, FP estimators and notably shrinkage FP estimates are capable of locating all the artificial targets and exhibit a lower number of false alarms. This improvement is due to the fact that FP estimates treat the outliers and impulsive samples in order for them to have a smaller contribution to the background characterization process, while the SMV-SCM estimates (and its respective diagonal loaded version) suffer from the presence of strong reflectance pixels in the secondary data. Remark that the shrinkage FP estimates allow for a better detection compared to FP estimates.

The algorithm has also been applied for galaxy detection on the MUSE data cube. The Multi Unit Spectroscopic Explorer (MUSE) project (see [34]) aims to provide astronomers with a new generation of optical instrument, capable of simultaneously imaging the sky (in 2D) and measuring the optical spectra of the light received at a given position on the sky. MUSE was installed on the VLT telescope and operational in 2013, and its performances are expected to allow observation of far galaxies up to 100 times fainter than those presently detectable. MUSE will deliver a 3D data-cube made of a stack of images recorded at 3578 different wavelengths over the range 465- 930 nm. Each monochromatic image represents a field of view of  $60 \times 60$  arcsec, recorded with a spatial sampling of 0.2 arcsec. Each record results in a data cube of size 1570 MB encoding 3578 images of  $300 \times 300$  pixels, possibly containing thousands of objects (galaxies) existing over different subsets of wavelengths.

An example of MUSE data cube image is displayed in Fig. 10 (a), from the 3578 available bands, we have chosen one band of each 100. The results for anomaly detection are presented in Fig.10 for a fixed value FA  $PFA = 10^{-3}$ . Note that detection with FP estimators (c) provides results with lower false alarm rate than classical ones (b).

These examples illustrate the robust behavior of FP estimators in non-Gaussian environments or for close targets

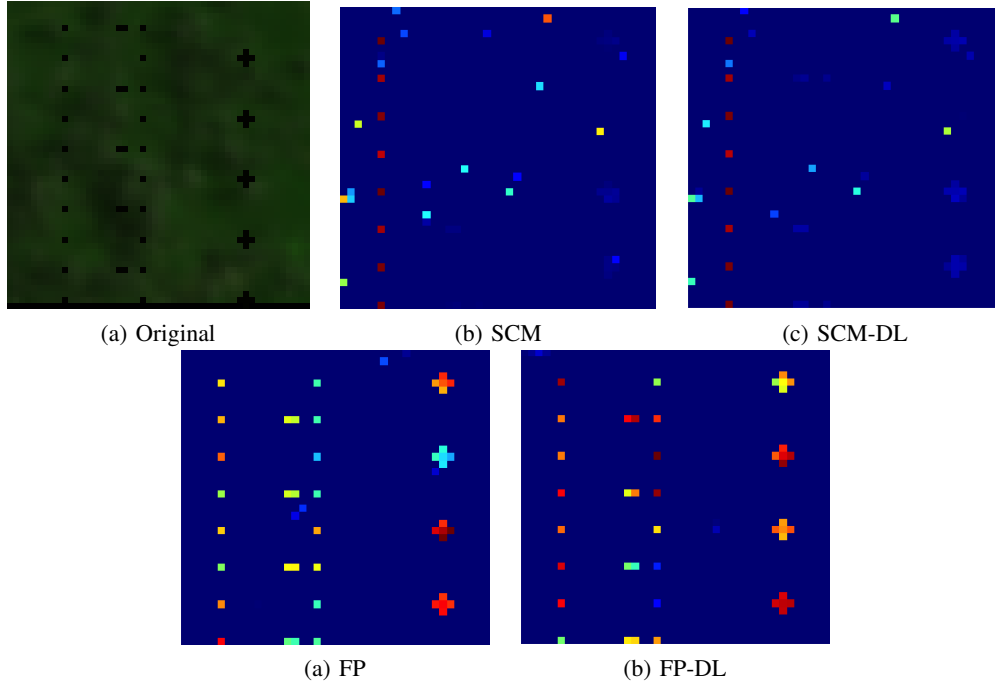


Fig. 9: Kelly AD built with robust estimates for artificial targets in real HSI.

detection problems.

## VI. CONCLUSION

The classical RXD test is explored and compared to other four detectors. The different advantages and drawbacks for the different detection schemes are commented. Furthermore, the comparison is performed through Monte Carlo simulations in Gaussian context and extended to real hyperspectral data with simulated anomalies. The family of elliptical distributions is considered for impulsive background characterization in hyperspectral imaging. In this context, robust estimation methods for mean vector and covariance matrix are used to overcome the non-Gaussianity of the background and the presence of outliers or strong scatters in the secondary data. Moreover the robust methods presented in this work outperform significantly the classical Gaussian-based SMV-SCM. Therefore, robust estimators offer a versatile alternative to Gaussian estimates. They allow to obtain better performances in impulsive environments while keeping good results in Gaussian background. The theoretical improvement provided by the robustness of the estimators is borne out through two real hyperspectral images.

## REFERENCES

- [1] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *Signal Processing Magazine, IEEE*, vol. 19, no. 1, pp. 29–43, 2002.
- [2] D. Manolakis, E. Truslow, M. Pieper, T. Cooley, and M. Brueggeman, "Detection algorithms in hyperspectral imaging systems: An overview of practical algorithms," *Signal Processing Magazine, IEEE*, vol. 31, no. 1, pp. 24–33, 2014.
- [3] M. T. Eismann, A. D. Stocker, and N. M. Nasrabadi, "Automated hyperspectral cueing for civilian search and rescue," *Proceedings of the IEEE*, vol. 97, no. 6, pp. 1031–1055, 2009.
- [4] C.-I. Chang and S.-S. Chiang, "Anomaly detection and classification for hyperspectral imagery," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 40, no. 6, pp. 1314–1325, 2002.
- [5] S. Matteoli, M. Diani, and G. Corsini, "A tutorial overview of anomaly detection in hyperspectral images," *IEEE Aero. and Electr. Systems Mag.*, vol. 25, no. 7, pp. 5–28, july 2010.
- [6] D. W. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker, "Anomaly detection from hyperspectral imagery," *Signal Processing Magazine, IEEE*, vol. 19, no. 1, pp. 58–69, 2002.
- [7] N. Nasrabadi, "Hyperspectral target detection: An overview of current and future challenges," *Signal Processing Magazine, IEEE*, vol. 31, no. 1, pp. 34–44, 2014.
- [8] D. Manolakis and D. Marden, "Non gaussian models for hyperspectral algorithm design and assessment," in *Geoscience and Remote Sensing Symposium, 2002. IGARSS'02. 2002 IEEE International*, vol. 3. IEEE, 2002, pp. 1664–1666.
- [9] D. Kelker, "Distribution theory of spherical distributions and a location-scale parameter generalization," *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 419–430, 1970.
- [10] K.-T. Fang, S. Kotz, and K. W. Ng, *Symmetric multivariate and related distributions*. Chapman and Hall, 1990.
- [11] D. Tyler, "A distribution-free  $m$ -estimator of multivariate scatter," *The Annals of Statistics*, vol. 15, no. 1, pp. 234–251, 1987.
- [12] F. Gini and M. V. Greco, "Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter," *Signal Processing, special section on SP with Heavy Tailed Distributions*, vol. 82, no. 12, pp. 1847–1859, December 2002.
- [13] J. Frontera-Pons, M. Mahot, J. Ovarlez, F. Pascal, S. Pang, and J. Chanussot, "A class of robust estimates for detection in hyperspectral images using elliptical distributions background," in *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*. IEEE, 2012, pp. 4166–4169.
- [14] J. Frontera-Pons, M. A. Veganzones, S. Velasco-Forero, F. Pascal, J.-P. Ovarlez, J. Chanussot *et al.*, "Robust anomaly detection in hyperspectral imaging," in *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*, 2014.
- [15] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*. Chapman & Hall/CRC, 2000.
- [16] I. Reed and X. Yu, "Adaptive multiple-band cfar detection of an optical pattern with unknown spectral distribution," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 38, no. 10, pp. 1760–1770, 1990.



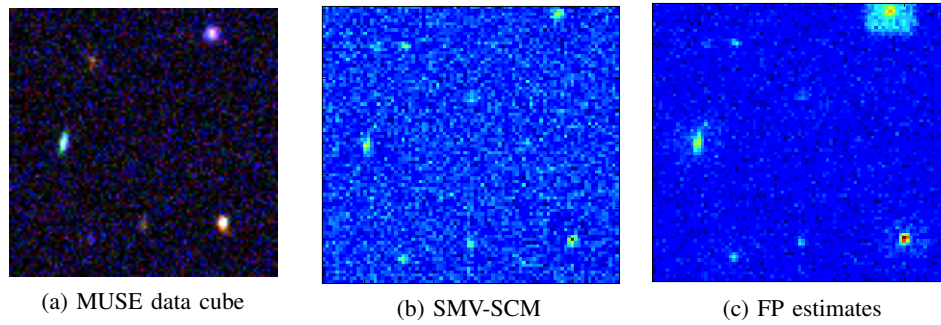


Fig. 10: Classical and Fixed-point anomaly detection in a hyperspectral image of  $300 \times 300$  in 3578 channels.

- [17] E. J. Kelly, "An adaptive detection algorithm," *Aerospace and Electronic Systems, IEEE Transactions on*, no. 2, pp. 115–127, 1986.
- [18] P. C. Mahalanobis, "On the generalized distance in statistics," *Proceedings of the National Institute of Sciences (Calcutta)*, vol. 2, pp. 49–55, 1936.
- [19] M. Bilodeau and D. Brenner, *Theory of multivariate statistics*. New York: Springer, 1999.
- [20] E. W. Weisstein, *CRC concise encyclopedia of mathematics*. CRC press, 2010.
- [21] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, New York, 1984.
- [22] J. C. Harsanyi, "Detection and classification of subpixel spectral signatures in hyperspectral image sequences," Ph.D. dissertation, University of Maryland Baltimore County, 1993.
- [23] J. Frontera-Pons, F. Pascal, and J.-P. Ovarlez, "Adaptive non-zero mean gaussian detection and application to hyperspectral imaging," *arXiv preprint arXiv:1404.2977*, 2014.
- [24] E. Ollila, D. Tyler, V. Koivunen, and H. Poor, "Complex elliptically symmetric distributions: survey, new results and applications," 2012.
- [25] F. Pascal, P. Forster, J.-P. Ovarlez, and P. Larzabal, "Performance analysis of covariance matrix estimates in impulsive noise," *IEEE Trans.-SP*, vol. 56, no. 6, pp. 2206–2217, June 2008.
- [26] Y. Abramovich, "Controlled method for adaptive optimization of filters using the criterion of maximum snr," *Radio Eng. Electron. Phys.*, vol. 26, no. 3, pp. 87–95, 1981.
- [27] B. D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 24, no. 4, pp. 397–401, 1988.
- [28] F. Pascal, Y. Chitour, and Y. Quek, "Generalized robust shrinkage estimator and its application to stap detection problem," *arXiv preprint arXiv:1311.6567*, 2013.
- [29] Y. Chen, A. Wiesel, and A. O. Hero, "Robust shrinkage estimation of high-dimensional covariance matrices," *Signal Processing, IEEE Transactions on*, vol. 59, no. 9, pp. 4097–4107, 2011.
- [30] Y. I. Abramovich and O. Besson, "Covariance matrix estimation in complex elliptic distributions using the expected likelihood approach," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 6476–6480.
- [31] Y. Abramovich and O. Besson, "Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach-part 1: The over-sampled case," 2012.
- [32] O. Besson and Y. Abramovich, "Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach-part 2: The under-sampled case," 2013.
- [33] R. Couillet and M. R. McKay, "Large dimensional analysis and optimization of robust shrinkage covariance matrix estimators," *arXiv preprint arXiv:1401.4083*, 2014.
- [34] "Official website of the muse project "http://muse.univ-lyon1.fr/".", [Online]. Available: "http://muse.univ-lyon1.fr/"

# List of publications

## Peer-reviewed scientific journal

1. Frontera-Pons, Pascal, F., and Ovarlez, J.P. Adaptive non-Zero Mean Gaussian Detection and Application to Hyperspectral submitted to *Signal Processing, IEEE Transactions on* in April 2014 .
2. Frontera-Pons, J., Veganzones, M.A., Pascal, F., and Ovarlez, J.P. Hyperspectral Anomaly Detectors using robust estimators, submitted to *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of* in September 2014.

## Conferences with proceedings

1. Frontera-Pons, J., Mahot, M., Ovarlez, J., Pascal, F., and Chanussot, J. Robust detection using  $M$ -estimators for hyperspectral imaging. In *IEEE-Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, (WHISPERS)*, Shanghai, Chine, June 2012,.
2. Frontera-Pons, J., Mahot, M., Ovarlez, J., Pascal, F., Pang, S., and Chanussot, J. A class of robust estimates for detection in hyperspectral images using elliptical distributions background. In *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, Munich, Germany, July 2012,.
3. Veganzones, M.A., Frontera-Pons, J., Chanussot, J., and Ovarlez, J.P. On the use of the Hotellings  $T^2$  statistic of the hierarchical clustering of Hyperspectral data. In *IEEE-Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, (WHISPERS)*, Florida, USA, June 2013,
4. Frontera-Pons, J., Ovarlez, J.P., Pascal, F., and Chanussot, J. Performance analysis of Robust Detectors for Hyperspectral Imaging. In *Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International*, Melbourne, Australie, July 2013,
5. Formont, P., Veganzones, M.A., Frontera-Pons, J., Pascal, F., Ovarlez, J.P. and Chanussot, J. CFAR hierarchical clustering of Polarimetric SAR data. In *Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International*, Melbourne, Australie, July 2013,



6. Frontera-Pons, J., Pascal, F., and Ovarlez, J.P. False-alarm regulation for target detection in Hyperspectral Imaging. In *IEEE-Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Saint Martin, France, December 2013.
7. Frontera-Pons, J., Veganzones, M.A., Velasco-Forero, S., Pascal, F., Ovarlez, J.P. and Chanussot, J. Robust Anomaly Detection in Hyperspectral Imaging. In *Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International*, Qubec, Canada, July 2014,
8. Veganzones, M.A., Frontera-Pons, J., Pascal, F., Ovarlez, J.P. and Chanussot, J. Binary Partition Trees-based Adaptive Hyperspectral RX Anomaly Detection. In *Image Processing (ICIP), 2013 IEEE International Conference on*, Paris, France, October 2014.

# Bibliography

- Abramovich, Y. (1981). Controlled method for adaptive optimization of filters using the criterion of maximum SNR. *Radio Eng. Electron. Phys*, 26(3):87–95.
- Abramovich, Y. and Besson, O. (2013a). Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach-part 1: The over-sampled case. *Signal Processing, IEEE Transactions on*, 61(23):5807–5818.
- Abramovich, Y. I. and Besson, O. (2013b). Covariance matrix estimation in complex elliptic distributions using the expected likelihood approach. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 6476–6480. IEEE.
- Abramowitz, M. E. et al. (1964). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, volume 55. Courier Dover Publications.
- Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, New York.
- Besson, O. and Abramovich, Y. (2013). Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach-part 2: The under-sampled case. *Signal Processing, IEEE Transactions on*, 61(23):5819–5829.
- Bilodeau, M. and Brenner, D. (1999). *Theory of multivariate statistics*. New York: Springer.
- Bioucas-Dias, J., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N., and Chanussot, J. (2013). Hyperspectral remote sensing data analysis and future challenges. *Geoscience and Remote Sensing Magazine, IEEE*, 1(2):6–36.
- Bioucas-Dias, J. M., Plaza, A., Dobigeon, N., Parente, M., Du, Q., Gader, P., and Chanussot, J. (2012). Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 5(2):354–379.
- Breloy, A., Ginolhac, G., Pascal, F., and Forster, P. (2014). Robust estimation of the clutter subspace for a low rank heterogeneous noise under high clutter to noise

- ratio assumption. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 66–70. IEEE.
- Carlson, B. D. (1988). Covariance matrix estimation errors and diagonal loading in adaptive arrays. *Aerospace and Electronic Systems, IEEE Transactions on*, 24(4):397–401.
- Chang, C.-I. (2003). *Hyperspectral imaging: techniques for spectral detection and classification*, volume 1. Springer.
- Chang, C.-I. and Chiang, S.-S. (2002). Anomaly detection and classification for hyperspectral imagery. *Geoscience and Remote Sensing, IEEE Transactions on*, 40(6):1314–1325.
- Chang, C.-I., Du, Q., Sun, T.-L., and Althouse, M. L. (1999). A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification. *Geoscience and Remote Sensing, IEEE Transactions on*, 37(6):2631–2641.
- Chen, J., Richard, C., and Honeine, P. (2013). Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model. *Signal Processing, IEEE Transactions on*, 61(2):480–492.
- Chen, M., Kanade, T., Pomerleau, D., and Rowley, H. A. (1999). Anomaly detection through registration. *Pattern Recognition*, 32(1):113–128.
- Chen, Y., Nasrabadi, N. M., and Tran, T. D. (2011a). Effects of linear projections on the performance of target detection and classification in hyperspectral imagery. *Journal of Applied Remote Sensing*, 5(1):053563–053563.
- Chen, Y., Wiesel, A., and Hero, A. O. (2011b). Robust shrinkage estimation of high-dimensional covariance matrices. *Signal Processing, IEEE Transactions on*, 59(9):4097–4107.
- Conte, E., De Maio, A., and Ricci, G. (2002). Recursive estimation of the covariance matrix of a compound-Gaussian process and its application to adaptive CFAR detection. *Signal Processing, IEEE Transactions on*, 50(8):1908–1915.
- Conte, E., Longo, M., and Lops, M. (1991). Modelling and simulation of non-Rayleigh radar clutter. In *Radar and Signal Processing, IEE Proceedings F*, volume 138, pages 121–130. IET.
- Couillet, R., Debbah, M., et al. (2011a). *Random matrix methods for wireless communications*. Cambridge University Press Cambridge, MA.
- Couillet, R., Debbah, M., and Silverstein, J. W. (2011b). A deterministic equivalent for the analysis of correlated mimo multiple access channels. *Information Theory, IEEE Transactions on*, 57(6):3493–3514.

- Couillet, R. and McKay, M. R. (2014). Large dimensional analysis and optimization of robust shrinkage covariance matrix estimators. *Journal of Multivariate Analysis*, 131:99–120.
- Couillet, R., Pascal, F., and Silverstein, J. W. (2012). Robust estimates of covariance matrices in the large dimensional regime. *arXiv preprint arXiv:1204.5320*.
- Couillet, R., Silverstein, J. W., Bai, Z., and Debbah, M. (2011c). Eigen-inference for energy estimation of multiple sources. *Information Theory, IEEE Transactions on*, 57(4):2420–2439.
- Eberlein, E. and Keller, U. (1995). Hyperbolic distributions in finance. *Bernoulli*, pages 281–299.
- Edelman, A. and Rao, N. R. (2005). Random matrix theory. *Acta Numerica*, 14:233–297.
- Eismann, M. T., Stocker, A. D., and Nasrabadi, N. M. (2009). Automated hyperspectral cueing for civilian search and rescue. *Proceedings of the IEEE*, 97(6):1031–1055.
- Frahm, G. (2004). *Generalized elliptical distributions: theory and applications*. PhD thesis, Universitätsbibliothek.
- Frahm, G., Junker, M., and Schmidt, R. (2005). Estimating the tail-dependence coefficient: properties and pitfalls. *Insurance: Mathematics and Economics*, 37(1):80–100.
- Frontera-Pons, J., Mahot, M., Ovarlez, J., Pascal, F., and Chanussot, J. (2012a). Robust detection using  $M$ -estimators for hyperspectral imaging. In *Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*.
- Frontera-Pons, J., Mahot, M., Ovarlez, J., Pascal, F., Pang, S., and Chanussot, J. (2012b). A class of robust estimates for detection in hyperspectral images using elliptical distributions background. In *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, pages 4166–4169. IEEE.
- Gini, F. (1997). Sub-optimum coherent radar detection in a mixture of K-distributed and Gaussian clutter. *IEE Proc. Radar, Sonar and Navigation*, 144(1):39–48.
- Gini, F., Farina, A., and Greco, M. (2001). Selected list of references on radar signal processing. *Aerospace and Electronic Systems, IEEE Transactions on*, 37(1):329–359.
- Gini, F. and Greco, M. V. (2002). Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter. *Signal Processing*, 82(12):1847–1859.
- Goetz, A. F., Vane, G., Solomon, J. E., and Rock, B. N. (1985). Imaging spectrometry for earth remote sensing. *Science*, 228(4704):1147–1153.

- Goodman, N. (1963). Statistical analysis based on a certain multivariate complex gaussian distribution (an introduction). *The Annals of Mathematical Statistics*, 34(1):152–177.
- Green, R. O., Eastwood, M. L., Sarture, C. M., Chrien, T. G., Aronsson, M., Chippendale, B. J., Faust, J. A., Pavri, B. E., Chovit, C. J., Solis, M., et al. (1998). Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS). *Remote Sensing of Environment*, 65(3):227–248.
- Gupta, A. K. and Nagar, D. K. (2000). *Matrix Variate Distributions*. Chapman & Hall/CRC.
- Harsanyi, J. C. (1993). *Detection and classification of subpixel spectral signatures in hyperspectral image sequences*. PhD thesis, University of Maryland Baltimore County.
- Harsanyi, J. C. and Chang, C.-I. (1994). Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach. *Geoscience and Remote Sensing, IEEE Transactions on*, 32(4):779–785.
- Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101.
- Jakeman, E. (1980). On the statistics of k-distributed noise. *Journal of Physics A: Mathematical and General*, 13(1):31.
- Kano, Y. (1994). Consistency property of elliptic probability density functions. *Journal of Multivariate Analysis*, 51(1):139–147.
- Kay, S. M. (1993). *Fundamentals of Statistical Signal Processing - Estimation Theory*, volume 1. Prentice-Hall PTR, Englewood Cliffs, NJ.
- Kay, S. M. (1998). *Fundamentals of Statistical signal processing - Detection theory*, volume 2. Prentice-Hall PTR, Englewood Cliffs, NJ.
- Kelker, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 419–430.
- Kelly, E. J. (1986). An adaptive detection algorithm. *Aerospace and Electronic Systems, IEEE Transactions on*, (2):115–127.
- Kent, J. T. and Tyler, D. E. (1991). Redescending m-estimates of multivariate location and scatter. *The Annals of Statistics*, pages 2102–2119.
- Keshava, N. and Mustard, J. F. (2002). Spectral unmixing. *Signal Processing Magazine, IEEE*, 19(1):44–57.

- Kirsteins, I. P. and Tufts, D. W. (1994). Adaptive detection using low rank approximation to a data matrix. *Aerospace and Electronic Systems, IEEE Transactions on*, 30(1):55–67.
- Korado, V. A. (1968). Optimum detection of signals with random parameters against the background of noise of unknown intensity under conditions of constant false alarm probability. *Radio Engineering and Electronic Physics*, 13:969–972.
- Kraut, S. and Scharf, L. L. (1999). The CFAR adaptive subspace detector is a scale-invariant GLRT. *Signal Processing, IEEE Transactions on*, 47(9):2538–2541.
- Kraut, S., Scharf, L. L., and Mc Whorter, L. T. (2001). Adaptive Subspace Detectors. *Signal Processing, IEEE Transactions on*, 49(1):1–16.
- Krishnaiah, P. and Lin, J. (1986). Complex elliptically symmetric distributions. *Communications in Statistics-Theory and Methods*, 15(12):3693–3718.
- Kwon, H. and Nasrabadi, N. M. (2006). Kernel matched subspace detectors for hyperspectral target detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(2):178–194.
- Lazarevic, A., Ertöz, L., Kumar, V., Ozgur, A., and Srivastava, J. (2003). A comparative study of anomaly detection schemes in network intrusion detection. In *SDM*, pages 25–36. SIAM.
- Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Sciences (Calcutta)*, 2:49–55.
- Mahot, M. (2012). *Estimation robuste de la matrice de covariance en traitement du signal*. PhD thesis, École normale supérieure de Cachan-ENS Cachan.
- Mahot, M., Pascal, F., Forster, P., and Ovarlez, J. (2013). Asymptotic properties of robust complex covariance matrix estimates. *Signal Processing, IEEE Transactions on*, 61(13):3348–3356.
- Mahot, M., Pascal, F., Forster, P., and Ovarlez, J.-P. (2012). Robust ANMF test using huber’s m-estimator. In *Sensor Array and Multichannel Signal Processing Workshop (SAM), 2012 IEEE 7th*, pages 373–376. IEEE.
- Manolakis, D., Lockwood, R., Cooley, T., and Jacobson, J. (2007). Robust matched filters for target detection in hyperspectral imaging data. In *Acoustics, Speech and Signal Processing (ICASSP), 2007 IEEE International Conference on*, volume 1, pages I–529. IEEE.
- Manolakis, D., Lockwood, R., Cooley, T., and Jacobson, J. (2009). Is there a best hyperspectral detection algorithm? In *SPIE Defense, Security, and Sensing*, pages 733402–733402. International Society for Optics and Photonics.

- Manolakis, D. and Marden, D. (2002). Non gaussian models for hyperspectral algorithm design and assessment. In *Geoscience and Remote Sensing Symposium, (IGARSS), 2002 IEEE International*, volume 3, pages 1664–1666. IEEE.
- Manolakis, D., Marden, D., and Shaw, G. (2003). Hyperspectral image processing for automatic target detection applications. *Lincoln Laboratory Journal*, 14(1):79–116.
- Manolakis, D., Rossacci, M., Cipar, J., Lockwood, R., Cooley, T., and Jacobson, J. (2005). Statistical characterization of natural hyperspectral backgrounds using t-elliptically contoured distributions. In *Defense and Security*, pages 56–65. International Society for Optics and Photonics.
- Manolakis, D. and Shaw, G. (2002). Detection algorithms for hyperspectral imaging applications. *Signal Processing Magazine, IEEE*, 19(1):29–43.
- Manolakis, D., Truslow, E., Pieper, M., Cooley, T., and Brueggeman, M. (2014). Detection algorithms in hyperspectral imaging systems: An overview of practical algorithms. *Signal Processing Magazine*, 31(1):24–33.
- Manolakis, D., Truslow, E., Pieper, M., Cooley, T., Brueggeman, M., and Lipson, S. (2013). The remarkable success of adaptive cosine estimator in hyperspectral target detection. In *SPIE Defense, Security, and Sensing*, pages 874302–874302. International Society for Optics and Photonics.
- Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Sbornik: Mathematics*, 1(4):457–483.
- Maronna, R. A. (1976). Robust  $M$ -estimators of multivariate location and scatter. *Annals of Statistics*, 4(1):51–67.
- Matteoli, S., Diani, M., and Corsini, G. (2010a). A tutorial overview of anomaly detection in hyperspectral images. *Aerospace and Electronic Systems Magazine*, 25(7):5–28.
- Matteoli, S., Diani, M., and Corsini, G. (2010b). A tutorial overview of anomaly detection in hyperspectral images. *Aerospace and Electronic Systems Magazine*, 25(7):5–28.
- Mausel, P., Kramber, W., and Lee, J. (1990). Optimum band selection for supervised classification of multispectral data. *Photogrammetric Engineering and Remote Sensing*, 56:55–60.
- Nasrabadi, N. (2014). Hyperspectral target detection: An overview of current and future challenges. *Signal Processing Magazine, IEEE*, 31(1):34–44.
- Niu, S., Ingle, V. K., Manolakis, D. G., and Cooley, T. W. (2010). Tests for the elliptical symmetry of hyperspectral imaging data. In *SPIE Optical Engineering+ Applications*, pages 78120D–78120D. International Society for Optics and Photonics.

- Ollila, E. and Koivunen, V. (2003a). Influence functions for array covariance matrix estimators. *Proc. IEEE Workshop on Statistical Signal Processing (SSP)*, ST Louis, MO, pages 445–448.
- Ollila, E. and Koivunen, V. (2003b). Robust antenna array processing using m-estimators of pseudo-covariance. In *Personal, Indoor and Mobile Radio Communications, 2003. PIMRC 2003. 14th IEEE Proceedings on*, volume 3, pages 2659–2663. IEEE.
- Ollila, E. and Tyler, D. E. (2012). Distribution-free detection under complex elliptically symmetric clutter distribution. In *Sensor Array and Multichannel Signal Processing Workshop (SAM), 2012 IEEE 7th*, pages 413–416. IEEE.
- Ollila, E., Tyler, D. E., Koivunen, V., and Poor, H. V. (2012). Complex elliptically symmetric distributions: survey, new results and applications. *Signal Processing, IEEE Transactions on*, 60(11):5597–5625.
- Ovarlez, J., Pang, S., Pascal, F., Achard, V., and Ng, T. (2011). Robust detection using the sirv background modelling for hyperspectral imaging. In *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*, pages 4316–4319. IEEE.
- Pascal, F., Bombrun, L., Tournet, J.-Y., and Berthoumieu, Y. (2013a). Parameter estimation for multivariate generalized gaussian distributions. *Signal Processing, IEEE Transactions on*, 61(23):5960–5971.
- Pascal, F., Chitour, Y., Ovarlez, J.-P., Forster, P., and Larzabal, P. (2008a). Covariance structure maximum likelihood estimates in compound gaussian noise: Existence and algorithm analysis. *Signal Processing, IEEE Transactions on*, 56(1):34–48.
- Pascal, F., Chitour, Y., and Quek, Y. (2013b). Generalized robust shrinkage estimator and its application to stap detection problem. *arXiv preprint arXiv:1311.6567*.
- Pascal, F., Forster, P., Ovarlez, J.-P., and Larzabal, P. (2005). Theoretical analysis of an improved covariance matrix estimator in non-Gaussian noise. In *Acoustics, Speech and Signal Processing (ICASSP), 2005 IEEE International Conference on*, pages 69–72, Philadelphia.
- Pascal, F., Forster, P., Ovarlez, J.-P., and Larzabal, P. (2008b). Performance analysis of covariance matrix estimates in impulsive noise. *Signal Processing, IEEE Transactions on*, 56(6):2206–2217.
- Pascal, F., Ovarlez, J.-P., Forster, P., and Larzabal, P. (2006). On a SIRV-CFAR detector with radar experimentations in impulsive noise. In *Proc. of the European Signal Processing Conf.*, Florence.



- Plaza, A., Benediktsson, J. A., Boardman, J. W., Brazile, J., Bruzzone, L., Camps-Valls, G., Chanussot, J., Fauvel, M., Gamba, P., Gualtieri, A., et al. (2009). Recent advances in techniques for hyperspectral image processing. *Remote sensing of environment*, 113:S110–S122.
- Rangaswamy, M., Lin, F. C., and Gerlach, K. R. (2004). Robust adaptive signal processing methods for heterogeneous radar clutter scenarios. *Signal Processing*, 84(9):1653–1665.
- Rangaswamy, M., Weiner, D. D., and Ozturk, A. (1993). Non-gaussian random vector identification using spherically invariant random processes. *Aerospace and Electronic Systems, IEEE Transactions on*, 29(1):111–124.
- Reed, I. and Yu, X. (1990). Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 38(10):1760–1770.
- Richmond, C. D. (2000). Performance of a class of adaptive detection algorithms in nonhomogeneous environments. *Signal Processing, IEEE Transactions on*, 48(5):1248–1262.
- Robey, F. C., Fuhrmann, D. R., Kelly, E. J., and Nitzberg, R. (1992). A CFAR adaptive matched filter detector. *Aerospace and Electronic Systems, IEEE Transactions on*, 28(1):208–216.
- Salembier, P. and Garrido, L. (2000). Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval. *Image Processing, IEEE Transactions on*, 9(4):561–576.
- Scharf, L. L. and Friedlander, B. (1994). Matched subspace detectors. *Signal Processing, IEEE Transactions on*, 42(8):2146–2157.
- Schreier, P. J. and Scharf, L. L. (2010). *Statistical signal processing of complex-valued data: the theory of improper and noncircular signals*. Cambridge University Press.
- Snyder, D., Kerekes, J., Fairweather, I., Crabtree, R., Shive, J., and Hager, S. (2008). Development of a web-based application to evaluate target finding algorithms. In *Geoscience and Remote Sensing Symposium, (IGARSS) 2008. IEEE International*, volume 2, pages II–915. IEEE.
- Soloveychik, I. and Wiesel, A. (2014). Tyler’s covariance matrix estimator in elliptical models with convex structure. *Signal Processing, IEEE Transactions on*, 62(20):5251–5259.
- Song, X., Wu, M., Jermaine, C., and Ranka, S. (2007). Conditional anomaly detection. *Knowledge and Data Engineering, IEEE Transactions on*, 19(5):631–645.

- Stein, D. W., Beaven, S. G., Hoff, L. E., Winter, E. M., Schaum, A. P., and Stocker, A. D. (2002). Anomaly detection from hyperspectral imagery. *Signal Processing Magazine, IEEE*, 19(1):58–69.
- Tarassenko, L., Hayton, P., Cerneaz, N., and Brady, M. (1995). Novelty detection for the identification of masses in mammograms.
- Theiler, J., Cao, G., Bachega, L. R., and Bouman, C. A. (2011). Sparse matrix transform for hyperspectral image processing. *Selected Topics in Signal Processing, IEEE Journal of*, 5(3):424–437.
- Theiler, J., Scovel, C., Wohlberg, B., and Foy, B. R. (2010). Elliptically contoured distributions for anomalous change detection in hyperspectral imagery. *Geoscience and Remote Sensing Letters, IEEE*, 7(2):271–275.
- Tochon, G., Feret, J., Martin, R., Tupayachi, R., Chanussot, J., and Asner, G. (2012). Binary partition tree as a hyperspectral segmentation tool for tropical rainforests. In *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, pages 6368–6371.
- Tyler, D. (1987). A distribution-free  $m$ -estimator of multivariate scatter. *The Annals of Statistics*, 15(1):234–251.
- Tyler, D. E. (1982). Radial estimates and the test for sphericity. *Biometrika*, 69(2):429–436.
- Tyler, D. E. (1988). Some results on the existence, uniqueness, and computation of the  $m$ -estimates of multivariate location and scatter. *SIAM Journal on Scientific and Statistical Computing*, 9(2):354–362.
- Valero, S., Salembier, P., and Chanussot, J. (2011). Hyperspectral image segmentation using binary partition trees. In *Image Processing (ICIP), 2011 IEEE International Conference on*, pages 1273–1276. IEEE.
- van den Bos, A. (1995). The multivariate complex normal distribution—a generalization. *Information Theory, IEEE Transactions on*, 41(2):537–539.
- Van Trees, H. L. (2004). *Detection, estimation, and modulation theory*. John Wiley & Sons.
- Veganzones, M., Tochon, G., Dalla Mura, M., Plaza, A., and Chanussot, J. (2013). Hyperspectral image segmentation using a new spectral mixture-based binary partition tree representation. In *Image Processing (ICIP), 2013 IEEE International Conference on*.
- Weisstein, E. W. (2010). *CRC concise encyclopedia of mathematics*. CRC press.
- Yao, K. (1973). A representation theorem and its applications to spherically invariant random processes. *Information Theory, IEEE Transactions on*, 19(5):600–608.

- Yao, K., Simon, M. K., and Bigiieri, E. (2004). Unified theory on wireless communication fading statistics based on sirp. In *Signal Processing Advances in Wireless Communications, 2004 IEEE 5th Workshop on*, pages 135–139. IEEE.
- Younsi, A., Greco, M., Gini, F., and Zoubir, A. (2009). Performance of the adaptive generalised matched subspace constant false alarm rate detector in non-gaussian noise: an experimental analysis. *Radar, Sonar & Navigation, IET*, 3(3):195–202.

## Abstract

Hyperspectral imaging (HSI) extends from the fact that for any given material, the amount of emitted radiation varies with wavelength. HSI sensors measure the radiance of the materials within each pixel area at a very large number of contiguous spectral bands and provide image data containing both spatial and spectral information. Classical adaptive detection schemes assume that the background is zero-mean Gaussian or with known mean vector that can be exploited. However, when the mean vector is unknown, as it is the case for hyperspectral imaging, it has to be included in the detection process. We propose in this work an extension of classical detection methods when both covariance matrix and mean vector are unknown.

However, the actual multivariate distribution of the background pixels may differ from the generally used Gaussian hypothesis. The class of elliptical distributions has already been popularized for background characterization in HSI. Although these non-Gaussian models have been exploited for background modeling and detection schemes, the parameters estimation (covariance matrix, mean vector) is usually performed using classical Gaussian-based estimators. We analyze here some robust estimation procedures ( $M$ -estimators of location and scale) more suitable when non-Gaussian distributions are assumed. Jointly used with  $M$ -estimators, these new detectors allow to enhance the target detection performance in non-Gaussian environment while keeping the same performance than the classical detectors in Gaussian environment. Therefore, they provide a unified framework for target detection and anomaly detection in HSI.

**Keywords :** Target detection, Hyperspectral imaging, Elliptical distributions, Robust estimation

## Résumé

L'imagerie hyperspectrale (HSI) repose sur le fait que, pour un matériau donné, la quantité de rayonnement émis varie avec la longueur d'onde. Les capteurs HSI mesurent le rayonnement des matériaux au sein de chaque pixel pour un très grand nombre de bandes spectrales contiguës et fournissent des images contenant des informations à la fois spatiale et spectrale. Les méthodes classiques de détection adaptative prennent généralement pour hypothèse un fond gaussien à vecteur moyenne nul ou connu. Cependant, quand le vecteur moyenne est inconnu, comme c'est le cas pour l'image hyperspectrale, il doit être inclus dans le processus de détection. Nous proposons dans ce travail d'étendre les méthodes classiques de détection pour lesquelles la matrice de covariance et le vecteur moyenne sont tous deux inconnus.

Cependant, la distribution statistique multivariée des pixels de l'environnement peut s'éloigner de l'hypothèse gaussienne classiquement utilisée. La classe de distributions elliptiques a déjà été popularisée pour la caractérisation de fond pour l'HSI. Bien que ces modèles non gaussiens aient déjà été exploités dans la modélisation du fond et dans la conception de détecteurs, l'estimation des paramètres (matrice de covariance, vecteur moyenne) est encore généralement effectuée en utilisant des estimateurs conventionnels gaussiens. Dans ce contexte, nous analysons des méthodes d'estimation robuste plus appropriées à ces distributions non-gaussiennes : les  $M$ -estimateurs. Ces méthodes de détection couplées à ces nouveaux estimateurs permettent, d'une part d'améliorer les performances de détection dans un environnement non-gaussien, mais d'autre part de garder les mêmes performances que celles des détecteurs conventionnels dans un environnement gaussien. Elles fournissent ainsi un cadre unifié pour la détection de cibles et la détection d'anomalies pour la HSI.

**Mots clés :** Détection de cibles, Imagerie Hyperspectrale, Distributions elliptiques, Estimation robuste