université PARIS-SACLAY

NNT : 2018SACLC043

CentraleSupélec

Exploitation of Sparsity for Hyperspectral Target Detection

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

Science et Technologies de l'Information et de la Communication (STIC)
Spécialité de doctorat: Traitement du signal et des images

Thèse présentée et soutenue à Gif, le 06 Juin 2018, par

Ahmad Bitar

Composition du Jury :

Sylvie MARCOS Directrice de Recherche, CNRS / L2S / CentraleSupélec	Président
Jean-Yves TOURNERET Professeur, INP de Toulouse	Rapporteur
Jocelyn CHANUSSOT Professeur, GIPSA-Lab, Grenoble INP	Rapporteur
Sylvie MARCOS Directrice de Recherche, CNRS / L2S / CentraleSupélec	Examinateur
Philippe FORSTER Professeur des universités, ENS Cachan	Examinateur
Nabil EL-KORSO Maître de Conférences, Université de Nanterre	Examinateur
Mauro DALLA MURA Maître de Conférences, GIPSA-Lab, Grenoble INP	Examinateur
Jean-Philippe OVARLEZ Maître de Recherche, ONERA & CentraleSupélec	Directeur de thèse
Loong-Fah Cheong Professeur, National University of Singapore	Co-Directeur de thèse

Thèse de doctorat

Exploitation of Sparsity for Hyperspectral Target Detection

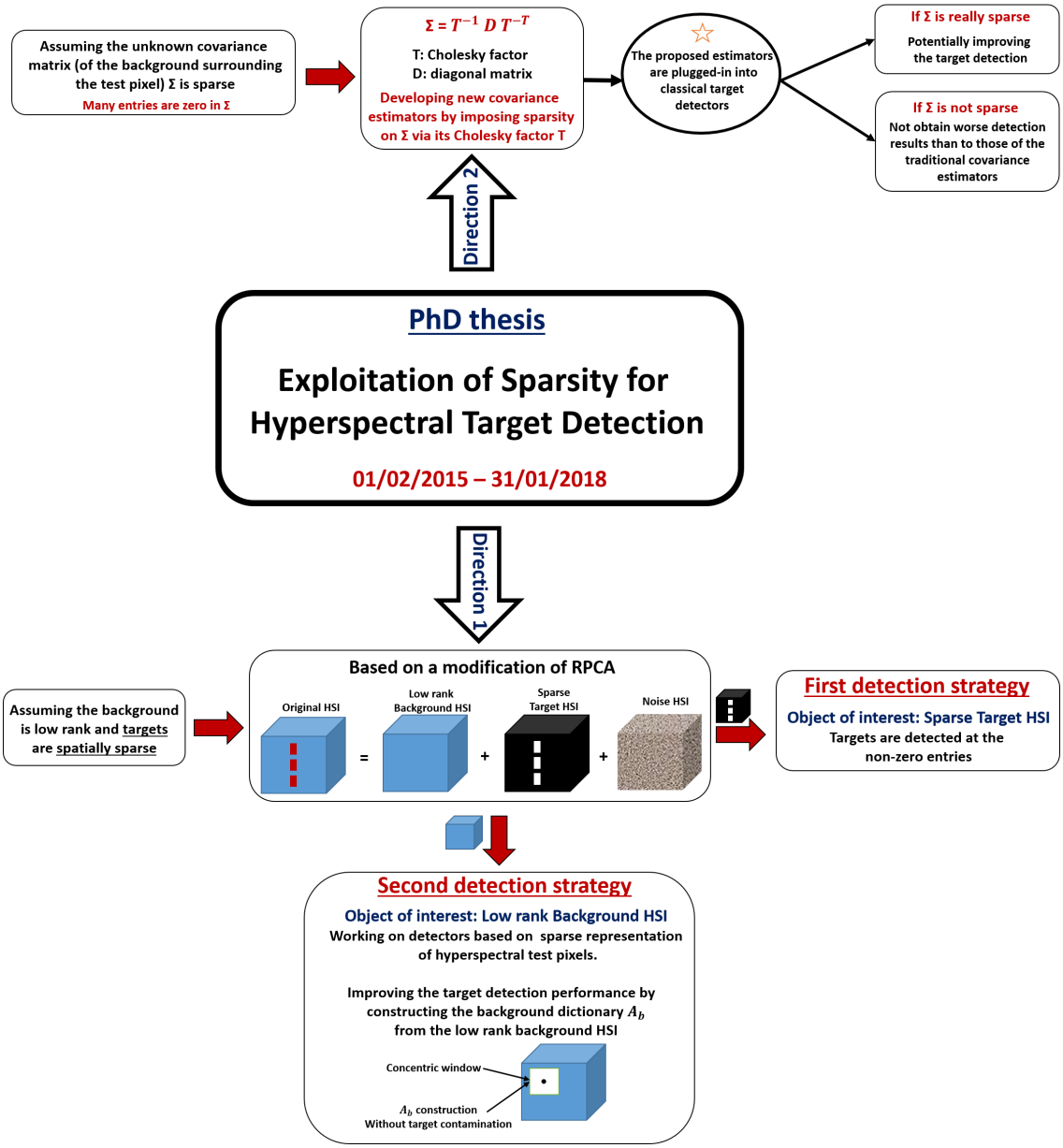
Ahmad W. Bitar

SONDRA laboratory
CentraleSupélec

Submitted in partial fulfilment of the requirements for the Degree of Doctor of Philosophy in Signal and Image Processing

Gif-sur-Yvette, April 2018

Graphical Abstract



Exploitation of Sparsity for Hyperspectral Target Detection

Ahmad W. Bitar

SONDRA laboratory
CentraleSupélec

Submitted in partial fulfilment of the requirements for the Degree of

Doctor of Philosophy in Signal and Image Processing

☐ I hereby certify that this is entirely
my own work unless otherwise stated.

Gif-sur-Yvette, April 2018

Dedicated to the memory of my dad
who has left but is still with me in my heart.
To my family for being with me despite the distance.

Acknowledgements

Because this may be my only opportunity to thank these individuals in writing, I may be a bit more verbose in my thanks than necessary.

As my thesis director, Jean-Philippe Ovarlez deserves thanks for many things. It has been an honor to be his Ph.D. student. I came to France in 02 February 2015 intending to join his group in SONDRA laboratory at CentraleSupélec for two primary reasons: the Ph.D. topic that he provided to me and by which I was very interested to work on, and what I have noticed about his personality. I have not been disappointed in either regard. He has provided guidance at key moments in my work and also a lot of support and motivations especially when I was sad after each rejected conference paper.

I would also like to thank my supervisor Loong-Fah Cheong who has taught me so much and has been endlessly patient with my countless questions. I appreciate a lot his support and helpful suggestions and advices he has given to me by mails, via all the Skype meetings we organized together, and especially when I worked with him at the National University of Singapore.

My utmost gratitude to Sylvain Azarian and Marc Lesturgie for welcoming me in the laboratory and for all the opportunities SONDRA has offered to me.

I would also like to extend my appreciation to all of the members of my Ph.D. evaluation committee. The reviewers Jean-Yves Tourneret and Jocelyn Chanussot for their careful review of my PhD thesis report. Appreciated! I also wish to say thank you to the examiners, Sylvie Marcos, Philippe Forster, Nabil EL-Korso and Mauro Dalla Mura, for having agreed to be in my thesis jury.

But none of this could have happened without my family. For as long as I can remember, my mom, Amal, has encouraged me to pursue all my interests. Thank you mom for your sincere love, kindness and making all my dreams come true. I would also like to express my sincere love to my sisters Lina, Nada and Racha; my brother, Ghassan; as well as Lina's husband, Hadi; Racha's husband, Hayssam; and Ghassan's wife, Suzan.

Lastly, I would like to give special thanks to my lovely nieces who are the main reason of my joy in this life. Thank you Ali, Wasfi, Nour, Yara, Carla, Perla and Julia.

Ahmad W. Bitar

Résumé

Une image hyperspectrale (HSI) est constituée d'une série d'images de la même scène spatiale, mais prises dans plusieurs dizaines de longueurs d'onde contiguës et très étroites, qui correspondent à autant de "couleurs". Lorsque la dimension spectrale est très grande, la détection de cibles devient délicate et caractérise une des applications les plus importantes de l'imagerie hyperspectrale. En imagerie hyperspectrale, il existe toujours des défis majeurs, e.g. la grande dimension, les effets d'atmosphères, et le bruit du capteur hyperspectrale. Cette thèse de doctorat se concentre principalement à exploiter la parcimonie (qui signifie généralement "petit en nombre ou quantité, souvent répartie sur une grande zone") afin d'atténuer le plus possible les défis mentionnés avant, ce qui implique une amélioration de la performance de détection de cibles. Les travaux dans cette thèse de doctorat sont divisés en deux directions différentes :

La première direction de la thèse :

La première direction de la thèse se concentre à exploiter la méthode d'analyse robuste en composantes principales (RPCA) afin de décomposer une image hyperspectrale d'intérêt en la somme d'une image de rang faible (associée au fond) et une image parcimonieuse de cibles contenant seulement les cibles avec le fond supprimé. L'utilisation du RPCA pour la reconnaissance de visages ou la détection de background/foreground dans une image a fait ses preuves. Cépendant, cette partie de thèse prouve que l'utilisation directe du RPCA est insuffisante pour distinguer les vraies cibles du fond en imagerie hyperspectrale. Plus particulièrement, le RPCA est evalué sur des images hyperspectrales réelles afin de démontrer qu'il ne cherche dans l'image que les petites régions hétérogènes et à contraste élevé, qui ne forment pas nécessairement la cible d'intérêt. Dans ce contexte, une version modifiée du RPCA est développée en tenant compte du fait que de l'information *a priori* est disponible sur la cible d'intérêt. Ceci consiste à modéliser la cible comme une mélange linéaire de signatures spectrales contenues dans un dictionnaire formé à partir des librairies spectrales en ligne. La méthode proposée pour estimer les différentes matrices intervenant dans la décomposition matricielle proposée est une méthode d'optimisation alternée très classique qui permet de découpler le problème en sous-problèmes plus simples via l'algorithme *Alternating Direction Method of Multipliers* (ADMM). À partir de la version du RPCA modifiée, deux stratégies de détection sont développées :

- La première stratégie de détection considère l'image parcimonieuse de cibles comme objet d'intérêt qui est utilisée directement pour la détection. Plus précisément, les cibles sont tout

simplement détectées aux entrées non nulles de l'image parcimonieuse de cibles. Cette image parcimoneuse de cibles constitue notre nouveau détecteur de cibles qui est caractérisé par les avantages suivants :

1. indépendant de la matrice de covariance;
2. invariant aux effets d'atmosphère;
3. a la capacité de détecter de cibles ayant une petite fraction de remplissage même lorsque la dimension spectrale est grande;
4. ne suppose aucune distribution au préalable.

- La deuxième stratégie de détection consiste à améliorer la construction adaptative du dictionnaire du fond dans les méthodes basées sur la représentation parcimonieuse des pixels de test hyperspectraux. Cette amélioration est effectuée à partir de l'image rang faible générée afin de construire un dictionnaire du fond pure de cibles d'intérêts.

La deuxième direction de la thèse:

La deuxième direction de la thèse pose le problème de la détection d'anomalies sous la forme d'un test d'hypothèse, qui, dans le cas d'un bruit Gaussian de moyenne nulle, se ramène à comparer une forme quadratique des vecteurs observés à un seuil adapté (e.g. le détecteur de Kelly). Puisque cette forme quadratique dépend de la matrice de covariance du bruit de mesure, il est alors intéressant de développer des estimateurs de cette matrice de covariance adaptés au problème de détection d'anomalies. Dans cette partie de la thèse, des nouveaux estimateurs de matrice de covariance sont développés en rendant la matrice de covariance parcimonieuse à partir de sa matrice unitaire triangulaire inférieure (aussi connue comme "facteur de Cholesky").

Evaluations des méthodes proposées :

Les méthodes proposées dans cette thèse sont évaluées sur des données synthétiques et epérimentales. Les résultats démontrent leur efficacité pour la détection de cibles en imagerie hyperspectrale.

Abstract

In hyperspectral imagery, pixels are represented by vectors whose entries correspond to spectral bands, and images are represented by 3-D hypercubes. Due to the high spectral dimensionality, one of the most important applications of hyperspectral imagery is target detection, which can be viewed as a binary classification problem where pixels are labeled as target or background based on their spectral characteristics. Hyperspectral imagery has many applications in areas such as military, astronomy, agricultural, mineralogy, and medical fields. Its rich spectral information allows for more accurate material identification.

In hyperspectral, the main challenges lie in large spectral dimensionality, and data variation modelling due to material spectral variability, atmospheric effects, and sensor noise. This thesis mainly concentrates on exploiting sparsity in order to alleviate the aforementioned challenges, and thus, improving the target detection for hyperspectral imagery. The proposed works are split into two different directions:

First thesis direction (Part II of this dissertation)
The first direction mainly concentrates on exploiting the recently developed Robust Principal Component Analysis (RPCA) for hyperspectral target detection. For a given hyperspectral image (HSI), and by considering similar assumptions to those used in RPCA, the background is assumed to have a low-rank property and the targets are spatially sparse. In particular, a given HSI is regarded as being made up of the sum of a low-rank background HSI (consisting only of background without the targets) and a sparse target HSI that only contains the targets with the background is suppressed. Some evaluations are made in this thesis to prove that the direct use of RPCA is inadequate to distinguish the true targets from their surrounding background. More precisely, the evaluations show that for a given HSI, only small heterogeneous and high contrast objects are deemed as targets under the general RPCA. In this regard, a modified version of RPCA is proposed by taking into consideration that a prior target information is provided to the user. This is done by introducing a subspace target dictionary, taken from online spectral libraries, into the sparse component, and thus, yielding a modified version of RPCA that is able to greatly identifying the true targets and separate them from the background (to be deposited in the sparse component). From the proposed target and background separation method (that is, the proposed modified RPCA), two detection strategies are available to realize the target detection:

- *The first detection strategy* (in Chapter 2) considers the sparse component as the object of interest, and which is used directly for the detection. That is, the targets are simply

detected at the non-zero entries of the sparse target HSI. In this detection strategy, a novel target detector is developed and which is simply a sparse HSI generated automatically from the original HSI, but containing only the targets with the background is suppressed.

- *The second detection strategy* (in Chapter 3) considers the low-rank component as the object of interest. It is known that most natural signals are inherently sparse in a certain basis or with respect to a given dictionary, and thus, they can be approximately represented by a few coefficients carrying the most relevant information. Hence, instead of modelling target and background signals using one single predefined model, sparsity has been exploited in the literature for hyperspectral imagery signal representations. The given dictionary which is used either for background or target, has a very important impact on the target detection performance. The main challenge is that there is always a lack of a sufficiently universal dictionary, especially for the background, and that should be constructed without contamination by the target pixels. That is why, this detection strategy mainly focuses on automatically constructing a locally adaptive background dictionary but on the low-rank background HSI (since it is pure from the targets), and thus, avoiding contamination by the target pixels. This leads to potentially improving the detection performances of the already developed target detectors which are based on the sparse representation approach.

Second thesis direction (Part III of this dissertation)

This thesis direction (in Chapter 6) consists on exploiting sparsity for covariance matrix estimation in order to improve the detection performance of the classical target detectors. These detectors depend on the true unknown covariance matrix (of the background surrounding the test pixel) $\mathbf{\Sigma}$ whose entries have to be carefully estimated especially in large dimensions. It is well known that when the spectral dimension is considered large compared to the sample size, such traditional covariance estimators lead to a very poor estimates performance unless some regularizations are applied such as constraining the covariance estimate to be sparse, namely, many entries are zero. In this thesis direction, the covariance matrix $\mathbf{\Sigma}$ is first regarded as being made up of $\mathbf{\Sigma} = \mathbf{T}^{-1}\,\mathbf{D}\,\mathbf{T}^{-T}$, where $\mathbf{T}$ is a unit lower triangular matrix (aka Cholesky factor) and $\mathbf{D}$ is a diagonal matrix with positive entries. Then, new covariance estimators are developed by assuming $\mathbf{\Sigma}$ is sparse and thus imposing sparsity via its Cholesky factor $\mathbf{T}$. Next, the proposed covariance estimators are plugged-in into classical target detectors (especially for anomaly detection). The proposed covariance estimators are always guaranteed to be positive definite. An important findings is that by taking advantage of the possible sparsity of $\mathbf{\Sigma}$, the target detection performance can be potentially improved; and even if $\mathbf{\Sigma}$ is not sparse (or not highly sparse), the proposed estimators do not achieve worse detection results than to those of the traditional covariance estimators.

Both thesis directions are evaluated on synthetic as well as real experiments, and the results of which demonstrate their effectiveness for hyperspectral target detection.

Contents

List of Figures

List of Abbreviations

RS . . . . . Remote Sensing
HSI . . . . . Hyperspectral Image
AVIRIS . . . . . Airborne Visible / Infrared Imaging Spectrometer
RPCA . . . . . Robust Principal Component Analysis
PCA . . . . . Principal Component Analysis
PCP . . . . . Principal Component Pursuit
SPCP . . . . . Stable Principal Component Pursuit
OP . . . . . Outlier Pursuit
SOP . . . . . Stable Outlier Pursuit
SVD . . . . . Singular Value Decomposition
ADMM . . . . . Alternating Direction Method of Multipliers
USGS . . . . . the United States Geological Survey
ASTER . . . . . Advanced Spaceborne Thermal Emission and Reflection
JHU . . . . . Johns Hopkins University
JPL . . . . . Jet Propulsion Laboratory
MP . . . . . Matching Pursuit
OMP . . . . . Orthogonal Matching Pursuit
SRBBH . . . . . Sparsity Representation-Based Binary Hypothesis
OWR . . . . . Outer Window Region
IWR . . . . . Inner Window Region
P_d . . . . . Probability of detection
P_{fa} . . . . . Probability of false alarm
ROC . . . . . Receiver Operating Characteristic
AUC . . . . . Area Under Curve
w.r.t . . . . . with respect to
OLS . . . . . Ordinary Least Squares
LASSO . . . . . Least Absolute Shrinkage and Selection Operator
SCAD . . . . . Smoothly Clipped Absolute Deviation
SCM . . . . . Sample Covariance Matrix
FP . . . . . Fixed Point
PDF . . . . . Probability Density Function
MLE . . . . . Maximum Likelihood Estimation
CG . . . . . Compound Gaussian
ANMF . . . . . Adaptive Normalized Matched Filter
i.i.d . . . . . Independent and identically distributed
LRT . . . . . Likelihood Ratio Test
SNR . . . . . Signal to Noise Ratio
AR . . . . . Autoregressive model
MUSE . . . . . Multi Unit Spectroscopic Explorer

List of Notations

Throughout this dissertation, we depict vectors in lowercase boldface letters and matrices in uppercase boldface letters.

For instance, $\mathbf{M}$ is a matrix, $[\mathbf{M}]_{:,j}$ is the j-th column, $[\mathbf{M}]_{j,:}$ is the j-th row.

$\mathbb{R}$ . . . . . Set of real numbers,

$\mathbb{E}$ . . . . . denotes the expectation,

$\mathbb{1}(.)$ . . . . . denotes the indicator function,

$(.)^T$ . . . . . denotes the transpose operator,

Tr . . . . . denotes the trace operator,

det(.) . . . . . denotes the determinant operator,

sgn . . . . . is for the sign: for any $z \in \mathbb{R}$, we define $\text{sgn}(z) = 1$ if $z > 0$, $\text{sgn}(z) = 0$ if $z = 0$ and $\text{sgn}(z) = -1$ if $z < 0$,

$(.)_+$ . . . . . For any $z \in \mathbb{R}$, we define $(z)_+ = 0$ if $z < 0$ and $(z)_+ = z$, if $z \geq 0$,

$\mathbf{0}_p$ . . . . . denotes a zero vector of dimension p,

H_0 . . . . . denotes the signal absent (or null) hypothesis,

H_1 . . . . . denotes the signal present (or alternative) hypothesis,

$\underset{H_0}{\overset{H_1}{\gtrless}}$ . . . . . decides H_0 (resp. H_1) if the result is inferior (resp. larger) than a certain value,

$\sim$ . . . . . means "distributed as",

✓ . . . . . is for "success",

✗ . . . . . is for "failure",

$\mathbf{I}$ . . . . . denotes the identity matrix,

rank(.) . . . . . is for the rank of a matrix,

$\hat{\mathbf{M}}$ . . . . . is the estimate of $\mathbf{M}$,

$||\mathbf{M}||_*$ . . . . . represents the nuclear norm of $\mathbf{M} \implies ||\mathbf{M}||_* = \text{Tr}(\sqrt{(\mathbf{M}^T\,\mathbf{M})})$,

$||\mathbf{M}||_0$ . . . . . represents the l_0 norm of $\mathbf{M}$. It counts the total number of non-zero entries in $\mathbf{M}$,

$||\mathbf{M}||_1$ . . . . . represents the l_1 norm of $\mathbf{M}$. It computes the sum of absolute values of the entries in $\mathbf{M}$,

$||\mathbf{M}||_F$ . . . . . represents the Frobenius norm of $\mathbf{M}$,

$||\mathbf{M}||_{2,0}$ . . . . . represents the $l_{2,0}$ norm of $\mathbf{M}$. It counts the total number of non-zero columns in $\mathbf{M} \implies ||\mathbf{M}||_{2,\,0} = \#\{j : ||[\mathbf{M}]_{:,\,j}||_2 \neq 0\}$,

$||\mathbf{M}||_{2,1}$ . . . . . represents the $l_{2,1}$ norm of $\mathbf{M} \implies ||\mathbf{M}||_{2,\,1} = \sum_j ||[\mathbf{M}]_{:,\,j}||_2$,

$||\mathbf{M}||_{0,2}$ . . . . . represents the $l_{0,2}$ norm of $\mathbf{M}$. It counts the total number of non-zero rows in $\mathbf{M} \implies ||\mathbf{M}||_{0,\,2} = \#\{j : ||[\mathbf{M}]_{j,\,:}||_2 \neq 0\}$,

$||\mathbf{M}||_{1,2}$ . . . . . represents the $l_{1,2}$ norm of $\mathbf{M} \implies ||\mathbf{M}||_{1,\,2} = \sum_j ||[\mathbf{M}]_{j,\,:}||_2$.

Part I

Introduction and Thesis Overview

Research is what I'm doing when I don't know what I'm doing.

— Wernher von Braun

1 Introduction

Contents

► *Synopsis* This chapter first provides a general introduction concerning hyperspectral imagery and target detection. Then, it gives a brief overview of this thesis, specifically by answering the question "*How and why is sparsity exploited for hyperspectral target detection?*". Next, it details the structure of this report and ends by providing some of the previously published materials.
In particular, this chapter is split into two parts:

- **The first part 1.1** "*Introduction to Hyperspectral Imagery and Target Detection*" introduces some basic information that should be known to well understand the concept of hyperspectral imagery and target detection. More precisely, it starts by giving a definition of remote sensing, and then briefly overviews the hyperspectral remote sensing where all necessary information concerning a hyperspectral image and hyperspectral sensors are given in detail. It ends by briefly discussing the main concept of target detection in hyperspectral imagery.

- **The second part 1.2** "*Thesis Overview: How and why is sparsity exploited for hyperspectral target detection?*" starts by giving a general definition of sparsity, and then overviews this

thesis by answering both questions "*Why is sparsity exploited?*" and "*How is sparsity exploited?*". Next, it details the structure of this report and ends by listing the previously published materials between 01 February 2015 and 31 January 2018.

1.1 Introduction to Hyperspectral Imagery and Target Detection

1.1.1 Remote Sensing

Remote sensing (RS) can be defined as the process of acquiring information of a scene without physically contacting it (that is, at a distance from the imaging surface or targeted scene). The RS images are usually collected by sensors which mainly rely on the energy emitted and reflected from the imaging surface [126]. Hyperspectral RS imaging is the main topic of this dissertation. Depending on the source of the energy involved in the image acquisition, two kinds of RS imaging systems are distinguished: passive and active. In passive systems, the sensors mainly rely on an external illumination source (i.e. the sun) to capture the targeted scene; whereas in active systems, the sensor has its own source of light or illumination.

1.1.2 Hyperspectral Remote Sensing

The hyperspectral remote sensing system [103] has four basic parts: the illumination source (i.e. the "sun" in passive remote sensing), the atmospheric path, the imaging surface (or targeted scene), and the airborne hyperspectral imaging sensor (see Figure 1.1). Due to the presence of the atmospheric path, the reflected energy (initially captured by the solar illumination and then modified by the atmosphere) from the material surface will be different from the one who reaches the sensor since it passes back through the atmosphere. In this regard, some effects that produce variability to the material spectra (e.g. atmospheric conditions, sensor noise, material composition, and scene geometry) have to be taken into consideration [77, 138].

Data cubes and spectral pixel representation

What is a hyperspectral image (HSI)?:
An airborne hyperspectral imaging sensor is capable of simultaneously acquiring the same spatial scene in contiguous and multiple narrow (0.01μm - 0.02μm) spectral wavelength (color) bands [103, 107, 123, 128]. When all the spectral bands are stacked together, the resulting hyperspectral image (HSI) is a three dimensional data cube (with spatial-spatial-spectral components). Thanks to the narrow acquisition, the HSI could have hundred to thousands of contiguous spectral bands. Having this very high level of spectral detail gives better capability to see the unseen.
For example, the Hyperion imaging sensor produces images of the same spatial scene in 220 contiguous spectral bands and in wavelengths ranging from 0.385μm to 2.5μm. The NASA's Airborne Visible / Infrared Imaging Spectrometer (AVIRIS) produces images in 224 contiguous

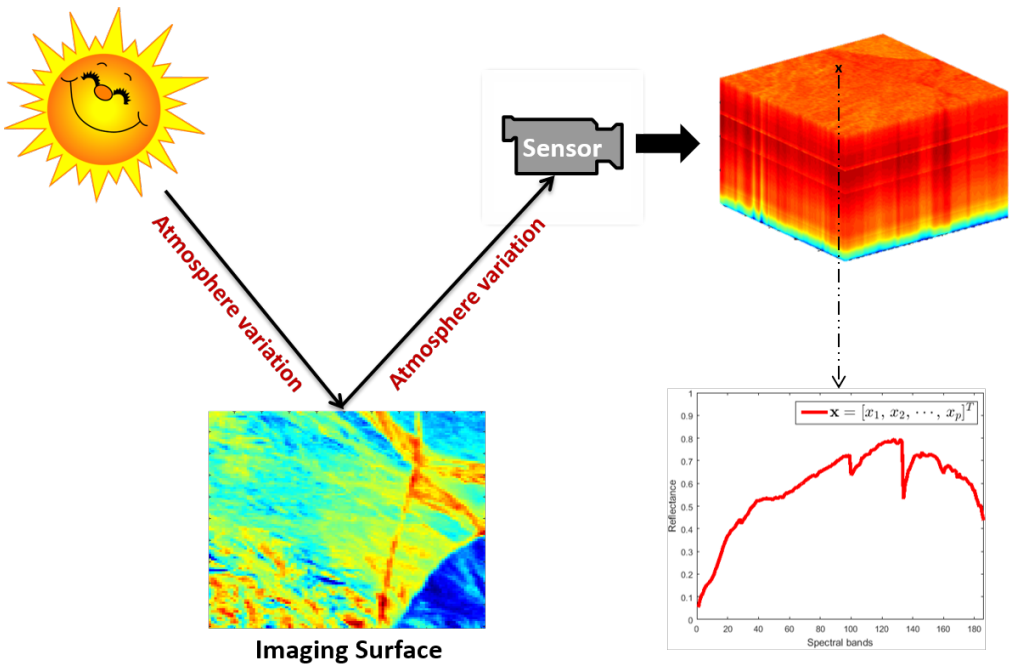


Figure 1.1: Passive hyperspectral remote sensing system

spectral bands with wavelengths ranging from 0.385μm to 2.5μm.

Spectral pixel representation:
In the spectral representation, each pixel in the HSI can be seen as a p-dimensional vector, $\mathbf{x} = [x_1, x_2, \cdots, x_p]^T \in \mathbb{R}^p$, where p designates the total number of spectral bands. The "spectral signature" of $\mathbf{x}$ (also known as "reflectance spectrum"), shows the fraction of incident energy, typically sunlight, that is reflected by a material from the surface of interest as a function of the wavelength of the energy [123].

Pure Vs Mixed pixels:
The HSI usually contains both pure and mixed pixels. A pure pixel contains only one single material, whereas a mixed pixel contains multiple materials, with its spectral signature representing the aggregate of all the materials in the corresponding spatial location [101, 103]. The latter situation often arises because hyperspectral images are collected hundred to thousands of meters away from an object, so that the object becomes smaller than the size of a pixel.

Characteristics of the airborne hyperspectral imaging sensors

Hyperspectral sensors are advanced colored digital cameras [103] which are characterized by 4 essential resolutions: spatial, spectral, radiometric and temporal.

- The *spatial resolution* determines how well the sensor can record spatial details in the targeted surface that are being imaged.

- The *spectral resolution* measures the width of the spectral bands used to measure the radiance at different wavelengths.
- The *radiometric resolution* corresponds to the number of bits used to describe the reflectance value measured by the sensor at each spectral band.
- The *temporal resolution* corresponds to how often the sensor revisits a scene to obtain a new set of data.

Most of hyperspectral sensors are ineffective at night, and measure radiation in the solar illumination portion (0.385μm - 2.5μm) of the electromagnetic spectrum. This in fact limits the use of these passive sensors to daylight hours [123].
When the range of the hyperspectral sensor is extended into thermal infrared portion (8μm to 14μm), materials emit more radiation than they reflect from the sun allowing spectrometers to operate all daylong [103, 123].

The electromagnetic spectrum portion 0.385μm - 2.5μm

- *The portion 0.385μm - 0.7μm*: it corresponds to the visible spectrum and which is the only portion perceptible to the human eye (with a maximum sensibility in around 0.55μm). In the 17th century, Isaac Newton (at the age of 23) held a prism of glass in the path of a beam of sunlight coming through a hole in the blind of his darkened room. He observed the white sunlight was split into red, orange, yellow, green, cyan and blue light. Newton has named the red, green and blue as "primary colors", since they regenerate white light when they are recombined all together. In addition, the mixture of any two of these primary colors will give the secondary colors. For example, mixing blue and green light gives cyan light; mixing green and red light gives yellow light. But when mixing red and blue light, Newton has observed a magenta light. Thus, his findings showed that the "primary colors" are red, green and blue, separated by the three "secondary colors" – yellow, cyan, and magenta.

  In fact, this magenta color has posed a mystery since it is a non-spectral color of light. It is impossible to find the magenta color on the visible spectrum due to the fact that there is no wavelength of light that makes it. In this regard, an English physician called Thomas young, was very curious to solve this magenta puzzle. He made an assumption that the human eye perceives only Newton's three primary colors (red, green, and blue), and that the eye perceived all of the variations in color by combining these internally. Hence, when red and blue are mixed together, the eye sees a magenta color, despite that the light is not actually magenta.

- *The portion 0.7μm - 2.5μm*: This portion includes the near infrared (from 0.7μm to 1.6μm) plus a part of the Medium infrared. The infrared light is not perceptible to the human eye. On the other hand, animals like goldfish for example can see in infrared light.

"The airborne hyperspectral imaging sensors which are effective in the portion 0.385μm - 2.5μm help us to see the world with the eyes of human and goldfish! "

1.1.3 Hyperspectral target detection: concept and Applications

Hyperspectral imagery has an emerging but narrow audience. With the rich information afforded by the high spectral dimensionality, hyperspectral imagery has found many applications in various fields such as agriculture [43, 120], mineralogy [79], military [51, 104, 131], and in particular, target detection [57, 59, 101, 103, 104, 106, 123, 128]. The latter is being not surprisingly one of the most important applications of hyperspectral imagery.
Usually, the detection is built using a binary hypothesis test that chooses between the following competing null and alternative hypothesis: target absent (H_0), that is, the test pixel $\mathbf{x}$ consists only of background; and target present (H_1) where $\mathbf{x}$ may be either fully or partially occupied by the target material.

Full vs subpixel targets

It is well known that the signal model for hyperspectral test pixels is fundamentally different from the additive model used in radar and communications applications [101, 107]. We can regard each test pixel $\mathbf{x}$ as being made up of $\mathbf{x} = \alpha\,\mathbf{t} + (1-\alpha)\,\mathbf{b}$, where $0 \leq \alpha \leq 1$ designates the target fill-fraction, $\mathbf{t}$ is the spectrum of the target, and $\mathbf{b}$ is the spectrum of the background (at the same spatial location). This model is known as replacement signal model, and hence, when a target is present in a given HSI, it replaces (that is, removes) an equal part of the background. For notational convenience, sensor noise has been incorporated into the target and background spectra (the vectors $\mathbf{t}$ and $\mathbf{b}$ include noise).

In particular, when $\alpha = 0$, the pixel $\mathbf{x}$ is fully occupied by the background material (the target is not present). When $\alpha = 1$, the pixel $\mathbf{x}$ is fully occupied by the target material and is usually referred to as the full or resolved target pixel. Whereas when $0 < \alpha < 1$, the pixel $\mathbf{x}$ is partially occupied by the target material, and is usually referred to as the subpixel or unresolved target.

The detection of a resolved target pixel depends on the spectral contrast between the target and background. On the other hand, the detection of an unresolved target depends on the target fill-fraction α, since it determines the amount of background interference on the observed target spectrum [101]. Additional factors affecting spectral detectability include environmental conditions and sensor noise.

Signal model used in target detection algorithms

The hyperspectral target detection algorithms are instead based on the additive model since the constraint $0 \leq \alpha \leq 1$ in the replacement model complicates their theoretical derivation and

practical implementation [101, 107]. The additive signal model can, however, be regarded as a good approximation of the replacement signal model for $0 < \alpha \ll 1$ (that is, $(1-\alpha) \approx 1$). However, the additive signal model assumption is not valid for real hyperspectral imaging data.

Applications to target detection

In hyperspectral imagery, the target detection algorithms can be grouped into two kinds of applications:

- *Application to target detection when the target* $\mathbf{t}$ *is known*:
  A prior target information can often be provided to the user. In real world hyperspectral imagery, this prior information may not be only related to its spatial properties (e.g. size, shape, texture) and which is usually not at our disposal, but to its spectral signature. The latter usually hinges on the nature of the given HSI where the spectra of the targets of interest have been already measured by some laboratories or with some hand-held spectrometers.
  Different target detectors (e.g. Matched Filter [105, 114], Normalized Matched Filter [93], and Kelly detector [91]) have been developed and which require that the target spectra to be known and provided to the user.

- *Application to target detection when the target* $\mathbf{t}$ *is not known (anomaly detection)*:
  This application is usually known as "hyperspectral anomaly detection". In many situations of practical interest, we do not have sufficient a priori information to specify the statistics of the target class. More precisely, the target spectra is not provided to the user (that is, not known). This unknown target is usually referred as « anomaly» [110] having a very different spectra from the surrounding background (e.g. a ship at sea).
  Different target detectors (e.g. Reed and Xiaoli detector [125], Kelly anomaly detector [92]) have been developed and which are independent on the target $\mathbf{t}$.

For any target detection to be successful, the target spectra must be distinguishable from the background spectra. In addition, detectors that depend on the target of interest $\mathbf{t}$ can have higher chance to achieve better detection performance than to those that do not depend on the target (the anomaly detectors). This is to be expected since anomaly detectors do not look for specific spectrally defined targets [101]. However, the application to anomaly detection for hyperspectral imaging is potentially interesting, but still not a promising area.

1.1.4 Serious challenges in hyperspectral target detection

The aforementioned classical target detectors, either when the target $\mathbf{t}$ is known or not, present several limitations in real world hyperspectral imagery. The task of understanding and solving these limitations presents significant challenges for hyperspectral target detection.

[Challenge one] The dependency on the unknown covariance matrix $\boldsymbol{\Sigma}$ (of the background surrounding the test pixel), and the estimation challenges of $\boldsymbol{\Sigma}$ in large dimensions and under different environments.

One of the major drawbacks of the aforementioned classical target detectors is that they depend on the unknown covariance matrix (of the background surrounding the test pixel) $\boldsymbol{\Sigma}$ whose entries have to be carefully estimated especially in large dimensions and to ensure success under different environments. In addition, there is always an explicit assumption on the statistical distribution of the observed data. Hence, estimating the covariance matrix in these classical target detectors has always been one of the stickiest points. This covariance matrix estimation can be interpreted as a whitening process in order to suppress image background.

- *Estimating the covariance matrix*: The sample covariance, for example, is the well known traditional covariance estimator derived by the maximum likelihood under Gaussianity. It is important to note that the estimation of covariance matrices through optimization of an objective function (e.g. a log-likelihood function) is usually a difficult numerical problem, since the resulting estimates should be positive definite matrices.

  In this regard, and in order to address the positivity definiteness constraint problem of the sample covariance, a method was developed to estimate the covariance matrix via linear regressions [124]. More precisely, for a positive-definite covariance matrix $\boldsymbol{\Sigma}$, its modified Cholesky decomposition can be written as $\mathbf{T}\,\boldsymbol{\Sigma}\,\mathbf{T}^T = \mathbf{D}$, where $\mathbf{T}$ is a unit lower triangular matrix (aka Cholesky factor) and $\mathbf{D}$ is a diagonal matrix with positive entries. The elements of $\mathbf{T}$ and $\mathbf{D}$ are uniquely defined and have interpretations as the successive regression coefficients and prediction error variances when measurements are regressed on their predecessors. A usual estimation of $\mathbf{T}$ and $\mathbf{D}$ can be simply done using the ordinary least squares method.

  However, some researchers have been interested on the covariance estimation by taking into consideration that in real world hyperspectral imagery, the background is usually non-homogeneous, and thus, the distribution of the background pixels can differ significantly from the theoretically predicted under Gaussian hypothesis. In this context, other traditional but robust covariance estimation approaches (i.e. the Tyler covariance estimator [41, 66, 119]) were proposed to take into account that the empirical distribution has instead long tails compared to the Gaussian assumptions [11, 12, 14, 15].

- *Covariance estimation in large dimensions*: Estimating large covariance matrices has been a longstanding important problem in many applications and has attracted increased attention over several decades. When the spectral dimension is considered large compared to the sample size (which is the usual case), the aforementioned traditional covariance estimators

are estimated with a lot of errors. It implies that the largest or smallest estimated coefficients in the matrix tend to take on extreme values not because this is "the truth", but because they contain an extreme amount of error [94, 95]. This is one of the main reasons why the classical target detectors usually behave poorly in detecting the targets of interest.

[Challenge two] Sensor noise, atmospheric conditions, and material composition

The classical target detectors that depend on the target to detect $\mathbf{t}$, use only a single reference spectrum for the target of interest. This might be inadequate since in real world hyperspectral imagery, various effects that produce variability to the material spectra (e.g. atmospheric conditions, sensor noise, and material composition) are inevitable [77, 138].

For instance, target signatures are typically measured in laboratories or in the field with hand-held spectrometers that are at most a few inches from the target surface. Hyperspectral images, however, are collected at huge distances away from the target and have significant atmospheric effects present.

1.2 Thesis Overview: How and why is sparsity exploited for hyperspectral target detection?

1.2.1 General definition of sparsity

I thought I will be late to my work! Fortunately, the traffic was sparse ☺

The word "**sparse**" (from Latin "sparsus") is an adjective from the early 18th century (used to describe writing in the sense "widely spaced"). It generally means "**small in number or amount, often spread over a large area**". This word is widely used everywhere and in everyday life (i.e. between people on the street, and between researchers in several domains).
As the general definition of sparsity is well defined, the researcher who is working on such application (i.e. Video surveillance, face recognition, hyperspectral target detection, hyperspectral image classification, and hyperspectral unmixing) might be interested to exploit this definition to develop contributions for solving or alleviating some specific problems and challenges. However, "*How and why is sparsity exploited?*" remains a very ambiguous question since its answer depends on the application of interest (in this thesis, it is the hyperspectral target detection) and the problems to solve in a specific application.

1.2.2 How and Why is sparsity exploited in this thesis?

There are several reasons that motivate our interest to exploit sparsity for hyperspectral target detection. We put forth three reasons, and for each of them, we briefly outline how is sparsity exploited to alleviate all or some of the aforementioned challenges of the classical target detectors, and hence, potentially improving the target detection performance.

[Reason one] In a given HSI, the targets usually occupy a small part of the entire image scene

In a given HSI, the targets usually occupy a very small part of the entire image scene and thus are assumed to be randomly distributed in the image scene and have low probability to be present. Thus, the targets present can be characterized by the sparsity property in the spatial domain (spatially sparse).

How is this sparsity exploited?

The main objective is to develop a new target detector that alleviates all the aforementioned challenges of the classical target detectors. Hence, our new target detector is independent on the true unknown covariance, behaves well in large dimensions, distributional free, and invariant to atmospheric effects. Our detector is based on similar assumptions to those used in Robust Principal Component Analysis (RPCA) [25]. That is, we exploit the aforementioned sparsity property on the targets (*the targets are spatially sparse*) aside an additional property that assumes the background is low-rank. More precisely, we take into consideration that in real world hyperspectral imagery:

1. The targets are spatially sparse (e.g. few pixels in a million pixel image),
2. The background is not too heavily cluttered with many different materials with multiple spectra, so that the background signals should span a low-dimensional subspace, a property that can be expressed as the low-rank condition of a suitably formulated matrix.

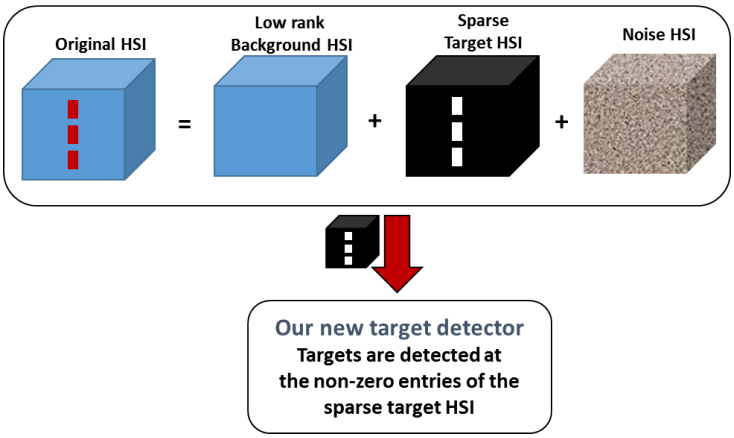


Figure 1.2: Flow Chart of our contribution for Reason one

From the aforementioned analysis, a method for separating targets of interest from the background in hyperspectral imagery is developed based on a modification of the well-known classical RPCA. More precisely, we first prove that a modification of RPCA is essential to substantially improving the discrimination of the targets of interest from their surrounding background. Then, we build the necessary assumptions for our modification of RPCA by taking into consideration that a prior target information is fully provided to the user and that the atmospheric influence can be

accounted via a pre-learned target dictionary $\mathbf{A}_t$ specified by the user.

From our proposed modification of RPCA, a given hyperspectral image (HSI) is thus regarded as being made up of the sum of a low-rank background HSI (consisting only of background), and a sparse target HSI that only contains the targets of interest based on the pre-learned target dictionary $\mathbf{A}_t$. From this target and background separation method, **only the sparse component (the sparse target HSI) will be the object of interest**, and hence, we aim to directly consider the sparse target HSI as our novel target detector (see Figure 1.2). More precisely, the targets are deemed to be present at the non-zero entries of the sparse target HSI.

To sum up, our main contribution for reason one is to develop a new target detector which is simply a sparse HSI generated automatically from the original HSI, but containing only the targets of interest with the background is suppressed (see the Flow Chart in Figure 1.2).

[Reason two] A hyperspectral test pixel lies in a low-dimensional subspace

It has been proven by Healey *et al.* in [77] that for any hyperspectral test pixel $\mathbf{x} \in \mathbb{R}^p$, it lies approximately in a low-dimensional subspace of the p-dimensional spectral-measurement space. Thus, instead of modelling target and background signals using one single predefined model, sparsity can be exploited for hyperspectral imagery signal representations.

How is this sparsity exploited?

Recently, some target detection algorithms based on sparse representation for HSI data have been developed [33, 163]. More precisely, dictionaries of target and background have been used (denoted as $\mathbf{A}_t$ and $\mathbf{A}_b$ in this thesis), and the test signal is then modeled as a sparse linear combination of the prototype signals taken from the dictionaries, and the recovered sparse representation can be used directly for the detection to constitute the target detector. The latter is being independent on the unknown covariance matrix, behaves well in large dimensions, distributional free, and invariant to atmospheric effects. More precisely, this sparse representation approach can alleviate the spectral variability caused by atmospheric effects, and can also better deal with a greater range of noise phenomena.

Our main contribution is to improve the detection performance of the already existing target detectors that exploit this sparse representation approach. In fact, the main drawback of these detectors is that they usually lack a sufficiently universal dictionary, especially for the background $\mathbf{A}_b$; some form of in-scene adaptation would be desirable. Hence, an important problem appears and which is the background dictionary $\mathbf{A}_b$ construction. It is noteworthy to mention that the construction of $\mathbf{A}_b$ is a very challenging problem since a contamination of it by the target pixels can potentially affect the target detection performance.

Our contribution in providing an accurate construction of $\mathbf{A}_b$:
Recall that in the previous contribution for Reason one, we have mentioned that a target and background separation method based on a modification of RPCA will be developed. More precisely, a given HSI can be split into a low-rank background HSI and a sparse target HSI. The latter is considered the object of interest and thus is used directly for the detection. The question is "*Why do we not also benefit from the low-rank component?*". Indeed, since this low-rank component can be pure from the targets, we aim to exploit it for a more accurate construction of $\mathbf{A}_b$, following which various target detectors based on sparse representation can be used to carry out a more elaborate binary hypothesis test. More precisely, we construct $\mathbf{A}_b$ using a specific adaptive method, applied on the low-rank background HSI which is generated from the modified version of RPCA, developed in the contribution of Reason one.

[Reason three] The covariance estimation is very challenging in large dimensions

Due to the fact that in hyperspectral imagery, the number of covariance matrix parameters to estimate grows with the square of the spectral dimension, it becomes impractical to use traditional covariance estimators, failing which the target detection performance can deteriorate significantly. Many a time, the researchers assume that compounding the large dimensionality problem can be alleviated by leveraging on the assumption that the true unknown covariance matrix (of the background surrounding the test pixel) is sparse, namely, many entries are zero. This sparsity assumption can potentially alleviate the large dimensionality challenge.

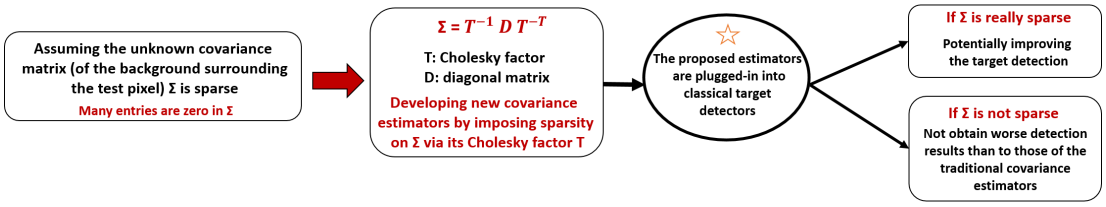


Figure 1.3: Flow Chart of our contribution for Reason three

How is this sparsity exploited?

As our contribution (see Figure 1.3), the covariance matrix $\mathbf{\Sigma}$ (of the background surrounding the test pixel) is first regarded as being made up of $\mathbf{\Sigma} = \mathbf{T}^{-1}\,\mathbf{D}\,\mathbf{T}^{-T}$, where $\mathbf{T}$ is a unit lower triangular matrix (aka Cholesky factor) and $\mathbf{D}$ is a diagonal matrix with positive entries. Then, new covariance estimators are developed by assuming $\mathbf{\Sigma}$ is sparse, namely, many entries are zero, where this sparsity is imposed via the Cholesky factor $\mathbf{T}$. The proposed covariance estimators are then plugged-in into classical target detectors.

Clearly, in real world hyperspectral imagery, the true covariance model is not known, and

hence, there is no prior information on its degree of sparsity. So "*What is the motivation of considering this sparsity constraint on the covariance matrix?*". Answering this question can obviously be like "*For the first intuition, considering this sparsity constraint on the covariance matrix seems to be a very strong assumption, but can contribute in potentially improving the target detection (especially in large dimensions) of the classical target detectors if the true unknown covariance matrix* $\mathbf{\Sigma}$ *(of the background surrounding the test pixel) is indeed sparse*". This answer will directly follow the question "*But what if the true unknown covariance* $\mathbf{\Sigma}$ *is not sparse?*". In this case, our answer is "*Even if the true covariance is not sparse (or not highly sparse), our proposed covariance estimators should at least not achieve worse target detection results (for the classical target detectors) than to those of the traditional covariance estimators*".

1.2.3 The Structure of this Report

The rest of this report is structured along the following lines:

1. **Part II: Automatic Target Detection Based on Low-Rank and Sparse Matrix Decomposition**:
   This part represents the first thesis direction.

   - **Chapter 2** "*A Novel Target Detector for Hyperspectral Imaging*" outlines the main contribution related to Reason one "*In a given HSI, the targets usually occupy a small part of the entire image scene*".
     It starts by providing a general background section that informs the rest of the chapter concerning the RPCA framework. Then, it highlights on our study of exploiting the classical RPCA for separating the targets of interest from the background in hyperspectral imagery. More precisely, we start by evaluating the classical RPCA on four real hyperspectral images to conclude that the direct use of RPCA is inadequate to distinguish the targets from the background. After that, we outline in detail the necessary steps and assumptions for our modification of RPCA by taking into consideration that a prior target information is fully provided to the user. We regard the given HSI as being made up of the sum of a low-rank background HSI, a sparse target HSI that only contains the targets of interest based on a pre-learned target dictionary specified by the user, and a Noise HSI. The sparse target HSI will be directly used for the detection, and hence, constitutes our novel target detector. The chapter ends with a summary of the proposed work.

   - **Chapter 3** "*Improving Background Dictionary Construction for Sparse Representation Methods*" outlines the main contribution related to Reason two "*A hyperspectral test pixel lies in a low-dimensional subspace*". More precisely, this chapter addresses the background dictionary $\mathbf{A}_b$ contamination problem suffered by the dictionary-based

target detection methods which are based on the sparse representation of hyperspectral test pixels. In order to do this, a specific adaptive method is used to construct $\mathbf{A}_b$ on the low-rank background HSI generated in Chapter 2, and hence, avoiding contamination by the target pixels. This improvement for the $\mathbf{A}_b$ construction is implemented into the already existing dictionary-based target detectors (that are based on the sparse representation approach) in order to potentially improve the target detection performance. The chapter ends with a summary of the proposed work.

- **Chapter 4** "*Application to Hyperspectral Target Detection for Chapter 2 and Chapter 3*": Both target detection strategies in Chapter 2 (that is, our novel target detector) and in Chapter 3 (that is, constructing an accurate background dictionary $\mathbf{A}_b$ pure from the target pixels via the low-rank background HSI, and implementing it into the existing dictionary-based target detectors) are evaluated extensively on both synthetic and real experiments for hyperspectral target detection. The chapter ends with a summary of the main contributions and results for both Chapter 2 and Chapter 3, and some directions for future work.

2. **Part III Target Detection Based on Sparse Covariance Matrices**:
   This part is completely independent from Part II
   This part represents the second thesis direction.

   - **Chapter 5** "*General Background*" provides a brief background of all the necessary information required for Chapter 6. The chapter starts by introducing the linear regression analysis. It then overviews the traditional covariance estimators, including a comparative study of them for hyperspectral target detection. Next, it outlines some research works that have been developed in the literature to alleviate the high covariance dimensionality challenge by assuming the covariance matrix is sparse. The chapter ends with a summary.

   - **Chapter 6** "*Imposing Sparsity on the Covariance Matrix via Its Cholesky Factor*" outlines the main contribution related to Reason three "*the covariance estimation is very challenging in large dimensions*". More precisely, this chapter aims to improve the target detection performance of the classical target detectors by imposing sparsity on the covariance matrix based on the assumption that the true unknown covariance matrix (of the background surrounding the test pixel) in a given HSI is sparse, namely, many entries are zero. A new covariance estimators that impose sparsity on the covariance matrix $\mathbf{\Sigma}$ via its Cholesky factor are developed. The chapter also provides some Monte-Carlo simulations as well as experimental data to demonstrate the effectiveness

of our proposed covariance estimators for hyperspectral target detection. The chapter ends with a summary of the whole Part III, and some directions for future work.

3. **Part IV: Concluding remarks**:

   - **Chapter 7** concludes with a summary of all the proposed works and the results presented, and some perspectives, new issues, opportunities and paths, the continuity of this work suggests.

1.2.4 Publications

This dissertation draws heavily on the earlier work and writing in the following papers, written jointly with Prof. Jean-Philippe Ovarlez and Prof. Loong-Fah Cheong:

- **Ahmad W. Bitar**, Jean-Philippe Ovarlez, Loong-Fah Cheong. Sparsity-based Cholesky Factorization and Its Application for Hyperspectral Anomaly Detection [18]. IEEE International Conference on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2017.

- **Ahmad W. Bitar**, Jean-Philippe Ovarlez, Loong-Fah Cheong. Exploitation de la parcimonie par la factorisation de Cholesky et son application pour la détection d'anomalies en imagerie hyperspectrale. GRETSI 2017.

- **Ahmad W. Bitar**, Loong-Fah Cheong, Jean-Philippe Ovarlez. Simultaneous Sparsity-Based Binary Hypothesis Model For Real Hyperspectral Target Detection [16]. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017.

- **Ahmad W. Bitar**, Loong-Fah Cheong, Jean-Philippe Ovarlez. Target and Background Separation in Hyperspectral Imagery for Automatic Target Detection. Accepted to the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018.

- **Ahmad W. Bitar**, Loong-Fah Cheong, Jean-Philippe Ovarlez. Sparse and Low-Rank Matrix Decomposition for Automatic Target Detection in Hyperspectral Imagery [17]. Submitted (it has been accepted, but modifications are required until 25 May 2018) to the IEEE Transactions on Geoscience and Remote Sensing (TGRS). http://arxiv.org/abs/1711.08970, Nov 2017.

Part II

Automatic Target Detection Based on Low-Rank and Sparse Matrix Decomposition

Research is to see what everybody else has seen, and to think what nobody else has thought.

— Albert Szent-Gyorgyi

2

A Novel Target Detector for Hyperspectral Imaging

Contents

► *Synopsis* This chapter introduces a novel target detector for hyperspectral imagery. This detector is independent on the unknown covariance matrix, behaves well in large dimensions, distributional free, and uses a target dictionary to alleviate the atmospheric influence. More precisely, based on a modification of the Robust Principal Component Analysis (RPCA), a given hyperspectral image (HSI) is regarded as being made up of the sum of a low-rank background HSI and a sparse target HSI that contains the targets based on a pre-learned target dictionary specified by the user. The sparse component (the sparse target HSI), will only be the object of interest, and hence, constitutes the novel target detector. This chapter is split into three parts:

1. **The first part 2.1** "*Background and System Overview*" begins by briefly overviewing all necessary information about RPCA. Then, it discusses the great success of RPCA in some applications such as face recognition and foreground detection. It ends by providing a brief system overview of the proposed work.

2. **The second part 2.2** "*Our study on testing RPCA for hyperspectral target detection*" first aims to answer the question "How is RPCA exploited for hyperspectral imagery?". Then, it proves that the direct use of RPCA is inadequate to distinguish the targets from the surrounding background, and hence, a modification of its version is highly recommended.

3. **The third part 2.3** "*Let us modify the RPCA*" outlines the necessary steps for the modification of RPCA by taking into consideration that a prior target information is fully provided to the user. The novel target detector is briefly described and which is simply a sparse HSI automatically generated from the original HSI, but only containing the targets of interest with the background is suppressed.

2.1 Background and System Overview

Suppose a large data matrix $\mathbf{D}$, and we have knowledge that it can be decomposed as the sum of a low-rank matrix $\mathbf{L}_0$ and a sparse matrix $\mathbf{E}_0$, both of arbitrary magnitude. That is,

$$\mathbf{D} = \mathbf{L}_0 + \mathbf{E}_0$$

Here, we do not know any information about the low-dimensional column and row space of $\mathbf{L}_0$, not even their dimension. In addition, we do not know the locations of the nonzero entries of $\mathbf{E}_0$ and not even how many there are.
The question is how can both the low-rank and sparse components be recovered accurately (perhaps even exactly) and efficiently?

A provably correct and scalable solution to the above problem would presumably have an impact on today's data-intensive scientific discovery. The recent explosion of massive amounts of high dimensional data in science, engineering, and society presents a challenge as well as an opportunity to many areas such as image, video, multimedia processing, web relevancy data analysis, search, biomedical imaging and bioinformatics. In such application domains, data now routinely lie in thousands or even billions of dimensions, with a number of samples sometimes of the same order of magnitude.

In many engineering applications such as in pattern analysis and signal processing, an underlying tenet is that the data often contains some type of structure that enables intelligent representation and processing. In real applications, the well-known (linear) subspaces are possibly the most common choice for such a parametric model in well characterizing a given data (e.g. motion, face and texture).

2.1.1 Principal Component Analysis (PCA)

The classical *Principal Component Analysis (PCA)* [49, 80, 87] has become a booming statistical tool for data analysis and dimensionality reduction. Given a set of data vectors, PCA seeks to find a small number of principal components (the orthogonal basis vectors) along which most of the variability of the data lies.
In its simplest form [87], PCA considers that a data matrix $\mathbf{D}$ is formed as a superposition of a low-rank matrix $\mathbf{L}_0$ and a perturbation matrix $\mathbf{N}_0$ (not necessarily sparse), that is:

$$\mathbf{D} = \mathbf{L}_0 + \mathbf{N}_0,$$

and then seeks to recover the low-rank k_L-estimate of $\mathbf{L}_0$ in the l_2 sense:

$$\begin{aligned} &\min_{\mathbf{L}} \left\{ \|\mathbf{D} - \mathbf{L}\|_2 \right\} \\ s.t. \quad &\text{rank}(\mathbf{L}) \leq k_L \,, \end{aligned}$$

which can be efficiently solved via Singular Value Decomposition (SVD) and thresholding. Mathematically, one can think here that the columns of the data matrix are data points and so after plotting them, they may be in a high dimensional space but can be very well distributed around a low-dimensional structure. Hence, PCA deals with the simplest assumption in that the data all lie near some low-dimensional subspace.

Figure 2.1: Corruptions (gross errors) on the face. A total of 58 different illuminations were applied for each face. These face images were used by Candès *et al.* in [25]

Unfortunately, the main disadvantage is that PCA is extremely sensitive to gross errors (Figure 2.1), and hence, the recovery of the low-rank matrix can be arbitrarily so far from the true $\mathbf{L}_0$. Nowadays, one has to always consider the presence of such gross errors which are becoming ubiquitous in modern applications such as image processing, web data analysis, where some measurements may be arbitrarily corrupted (due to occlusions, sensor failures, etc.). For example, in face recognition, one can imagine how it becomes hard to successfully recognizing a face if it is occluded by scarf, or sunglasses. The same problem occurs when the face image is corrupted by specularities (i.e. in the eyes), shadows (i.e. around the nose region), and brightness saturations (i.e. on the face).

Hence, robustifying PCA has become an essential need. In real world, one has to deal with large collection of images, videos, web data, and so having a lot of incorrect entries in huge data matrices should always be taken into consideration.

2.1.2 Robust Principal Component Analysis (RPCA)

A lot of methods aiming to robustify PCA have been proposed in the literature over several decades. For example, the influence function techniques [45, 83], multivariate trimming [67], alternating minimization [89], and random sampling techniques [55].
Unfortunately, none of these PCA variants possessed the strong performance guaranteed by the recent works based on the idea that a given data matrix can be formed by adding an unknown low-rank matrix to an unknown sparse matrix that contains the corruptions with possibly some additional noise errors. This decomposition matrix model can be solved via Robust Principal Component Analysis (RPCA) [21, 25, 141, 149], in which, the low-rank matrix $\mathbf{L}_0$ is being recovered from highly corrupted measurements $\mathbf{D} = \mathbf{L}_0 + \mathbf{E}_0$. Hence, unlike the small noise term $\mathbf{N}_0$ in classical PCA, now the entries in $\mathbf{E}_0$ can have arbitrarily large magnitude, and their support is assumed to be sparse but not known.

More precisely, the RPCA framework aims to estimate both $\mathbf{L}_0$ and $\mathbf{E}_0$ by solving the following minimization problem:

$$\min_{\mathbf{L},\mathbf{E}} \Big\{ \text{rank}(\mathbf{L}) \,+\, \lambda \, \|\mathbf{E}\|_l \Big\} \quad s.t. \quad \mathbf{D} = \mathbf{L} + \mathbf{E}\,, \tag{2.1}$$

where $\lambda > 0$ is a regularizing parameter (it plays a very important role since it trades off between the low-rank and the sparse component) and $\|.\|_l$ indicates certain sparse regularization strategy, such as the l_0 norm adopted in [25], the $l_{2,0}$ norm adopted in [86, 152], and the $l_{0,2}$ norm.

2.1.3 Some recent research in RPCA

Problem (2.1) is NP-HARD to solve due to the presence of the rank term and the $\|.\|_l \in \{\|.\|_0\,, \|.\|_{2,0}\,, \|.\|_{0,2}\}$ norm. In this regard, several recent research works have been done in RPCA and which differ from the decomposition, the loss functions, the optimization problem, and the solvers used. These works in RPCA include:

- *RPCA via Principal Component Pursuit (RPCA-PCP)*: The l_0 norm is used for the sparse component in the RPCA problem in (2.1). However, [25, 30, 149, 160] showed that by relaxing the rank term to the nuclear norm and the l_0 norm to the l_1 norm, i.e.,

$$\min_{\mathbf{L},\mathbf{E}} \Big\{ \|\mathbf{L}\|_* \,+\, \lambda \, \|\mathbf{E}\|_1 \Big\} \quad s.t. \quad \mathbf{D} = \mathbf{L} + \mathbf{E}\,, \tag{2.2}$$

  known as Principal Component Pursuit (PCP), one can recover both $\mathbf{L}_0$ and $\mathbf{E}_0$ exactly with high probability under certain conditions by solving this convex minimization problem.

  why is the nuclear norm used as a surrogation of the rank term?

  Note that: $\|\mathbf{L}\|_* = \text{Tr}\left(\sqrt{\mathbf{L}^T\,\mathbf{L}}\right)$, and if $\mathbf{L}$ is replaced by its singular value decomposition $\mathbf{U}\,\mathbf{S}\,\mathbf{V}^\mathbf{T}$, we obtain $\|\mathbf{L}\|_* = \text{Tr}\{|\mathbf{S}|\}$. Hence, using the nuclear norm is like to apply the l_1 norm on the vector of singular values of $\mathbf{L}$. This demonstrates why is the nuclear norm used

as a convex surrogation of the rank term.

One of the main limitations of RPCA-PCP is that it exactly decomposes $\mathbf{D}$ into $\mathbf{L}$ and $\mathbf{E}$. However, this exact decomposition does not always exist for a real data matrix $\mathbf{D}$ due to the possible presence of a small magnitude dense inlier noise.

- *RPCA via Stable Principal Component Pursuit (RPCA-SPCP)*: Due to the aforementioned limitation of RPCA-PCP, Zhou *et al.* [168] have used the model $\mathbf{D} = \mathbf{L}_0 + \mathbf{E}_0 + \mathbf{N}_0$, and solved a relaxed version of (2.2) known as Stable Principal Component Pursuit (SPCP):

$$\min_{\mathbf{L},\mathbf{E}} \left\{ \|\mathbf{L}\|_* + \lambda \, \|\mathbf{E}\|_1 \right\} \quad s.t. \quad \|\mathbf{D} - \mathbf{L} - \mathbf{E}\|_F \leq \delta \,, \tag{2.3}$$

for some $\delta > 0$. It was shown that the estimation error can be bounded. The Lagrangian form of (2.3) is written as [149, 168]:

$$\min_{\mathbf{L},\mathbf{E}} \left\{ \|\mathbf{L}\|_* + \lambda \, \|\mathbf{E}\|_1 + \frac{1}{2\,\mu} \, \|\mathbf{D} - \mathbf{L} - \mathbf{E}\|_F^2 \right\}, \tag{2.4}$$

where $\mu > 0$.

- *RPCA via Outlier Pursuit (RPCA-OP)*: In contrast to [25] where the l_0 norm is chosen for the sparse component, Xu *et al.* in [152] have been interested to the use of $l_{2,0}$ norm and its surrogation to the $l_{2,1}$ norm. Hence, enforcing sparsity in $\mathbf{E}$ but in columns. The following convex minimization problem is considered:

$$\min_{\mathbf{L},\mathbf{E}} \left\{ \|\mathbf{L}\|_* + \lambda \, \|\mathbf{E}\|_{2,1} \right\} \quad s.t. \quad \mathbf{D} = \mathbf{L} + \mathbf{E} \,. \tag{2.5}$$

- RPCA via Stable Outlier Pursuit (RPCA-SOP): As in [168], Xu *et al.* in [152] further proposed a Stable OP (SOP) that guarantees stable and accurate recovery in the presence of entry-wise noise. The following convex minimization problem is considered:

$$\min_{\mathbf{L},\mathbf{E}} \left\{ \|\mathbf{L}\|_* + \lambda \, \|\mathbf{E}\|_{2,1} + \frac{1}{2\,\mu} \, \|\mathbf{D} - \mathbf{L} - \mathbf{E}\|_F^2 \right\}. \tag{2.6}$$

There have been a lot of other works in RPCA addressing the presence of noise, and modification in the loss functions. These works include the *RPCA via Sparsity Control* [108, 109], *RPCA via Sparse Corruptions* [81], *RPCA via Log-sum heuristic Recovery* [46], *RPCA via Iteratively Reweighted Least Squares* [73, 75] and its improved version in [74], *RPCA via Stochastic Optimization* [54, 68, 85], *RPCA with Dynamic Mode Decomposition* [88], *Bayesian RPCA* [47] and its improved versions in [1, 4, 165], *Approximated RPCA* [167], *Sparse Additive Matrix Factorization* [112, 113], and *Variational Bayesian Sparse Estimator* [36].

2.1.4 Success of RPCA in some applications: face recognition and foreground detection

There are several applications in which the data under study can be modeled as a low-rank plus a sparse decomposition.

Depending on the application and the problems to solve, either the low-rank component or the sparse component can be the object of interest.
We briefly outline two important applications in which the RPCA has obtained encouraging performances, as proved by Candès *et al.* in [25]. These applications include: face recognition and foreground detection.

RPCA for face recognition

In this application, the low-rank component **L** *is the object of interest*

It has been shown that images of human's face can be well approximated by a low-dimensional subspace. Candès *et al.* in [25] have proven (visually) that for a given face image corrupted by some self-shadowing, specularities or saturations in brightness, the RPCA is capable of removing these defects in order to provide better recognition performance. In recognizing a human's face

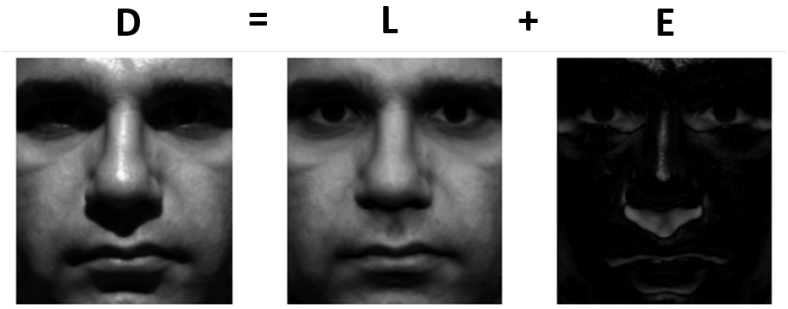


Figure 2.2: Removing shadows, specularities, and saturations from a face image using RPCA-PCP [25]. This example is taken from [25]. The original face image **D** is split into a low-rank face component **L** and a sparse component **E** that corresponds to specularities in the eyes, shadows around the nose region, and brightness saturations on the face.

image under such defect (i.e. self-shadowing), the low-rank component $\mathbf{L}_0$ will correspond to the face image without the defect, and the sparse component $\mathbf{E}_0$ captures the defects present. Figure 2.2 (an example taken from [25]) exhibits a separation evaluation example for applying the RPCA-PCP on a given face image to remove specularities in the eyes, shadows around the nose region, and brightness saturations on the face.

RPCA for foreground detection

In this application, the sparse component **E** *is the object of interest*

Many target detection applications such as in computer vision, image processing, and biomedical image processing, involve an automatic detection of such an object or activity. In a very simple video analysis for example, the targets of interest correspond to the moving objects with respect to a static background image. The term "very simple video" means that the moving objects in the scene only correspond to the true targets (that is, the background is static (i.e. there is no escalator), and there is no a static target (i.e. a person who does not move)).

These targets are also known as “foreground”, and hence, the target detection in video analysis is called “foreground detection”. That is, to detect, distinguish or separate all the moving objects (that correspond to the true targets) from the static information called “background”.
By applying RPCA for foreground detection, the background sequence is modeled by the low-rank subspace that can gradually change over time, while the moving target objects constitute the correlated sparse outliers. If we stack the video frames as columns of a matrix $\mathbf{D}$, then the

Figure 2.3: Detecting the moving targets from a static background using RPCA-PCP

low-rank component $\mathbf{L}_0$ will correspond to the background, and the sparse component $\mathbf{E}_0$ captures the moving objects. Figure 2.3 (an example taken from [25]) exhibits an example of the use of RPCA-PCP on detecting the moving targets from a static background.

Note that in practice, the scene usually has a non static background (i.e. some background objects are in move [19]), e.g., a background that contains a non-target moving object such as an escalator. While these objects should not be considered a part of the foreground, it has been shown that RPCA consider them as targets, and thus, RPCA captures any of the moving objects (that is, the moving objects are being deposited in the sparse component $\mathbf{E}$) which may not necessarily be the targets of interest. Hence, RPCA has achieved encouraging performances for foreground detection in very simple videos analysis, but still inadequate to distinguish the true targets from the surrounding non-static background.

A lot of researchers are aiming today to alleviate the RPCA challenges in realistic environment (i.e. illumination change causing complex intensity variation, background motions (trees, waves) whose magnitude can be greater than the foreground, poor image quality under low light, and camouflage). Interested readers can refer to [62, 63].

The RPCA for foreground detection is not interesting in this dissertation. Instead, we aim to exploit the RPCA for hyperspectral imaging data, as can be seen in the following sections. Note that foreground detection and hyperspectral imagery applications are not directly related.

2.1.5 System Overview of the proposed work

Due to the encouraging performances of RPCA that have been proven by Candès *et al.* in [25] for applications such as face recognition and foreground detection (specifically, for a very simple video analysis), we first aim to study how can RPCA be exploited for hyperspectral imagery [153–155]. More precisely, we regard a given HSI as being made up of the sum of a low-rank background HSI (that constitutes the background) and a sparse target HSI that only contains the targets of interest whose background has been suppressed.

Then, using four real hyperspectral images that differ by the complexity of the targets and their surrounding background, we prove that the direct use of RPCA is inadequate to distinguish the targets of interest from the background.

In this regard, we aim to modify the RPCA by taking into consideration that a prior target information is fully provided to the user. After incorporating this prior information into the sparse component, the latter is being used directly for the detection.

2.2 Our study on testing RPCA for hyperspectral target detection

2.2.1 How can RPCA be exploited for hyperspectral imagery?

How to define both the low-rank component $\mathbf{L}_0$ *and sparse component* $\mathbf{E}_0$*?*

While we do not need to make assumptions about the size, shape, or number of the targets present in a given HSI, it is possible to consider the same constraints as to those used in RPCA. More precisely:

- The total image area of all the target(s) should be small relative to the whole image (i.e. spatially sparse), e.g., several hundred of pixels in a million pixel image, though there is no restriction on target shape or the proximity between targets. This will define the sparse component $\mathbf{E}_0$.
- The background is not too heavily cluttered with many different materials with multiple spectra, so that the background signals should span a low-dimensional subspace, a property that can be expressed as the low-rank condition of a suitably formulated matrix [32, 155, 164]. This will define the low-rank component $\mathbf{L}_0$.

Hence, for any given HSI of size $h \times w \times p$, where h and w are the height and width of the image scene, respectively, and after rearranging it into a two-dimensional matrix $\mathbf{D}$ of size $e \times p$ (by lexicographically ordering the columns), where $e = h \times w$, the HSI can be modeled as $\mathbf{D} = \mathbf{L}_0 + \mathbf{E}_0 + \mathbf{N}_0$.

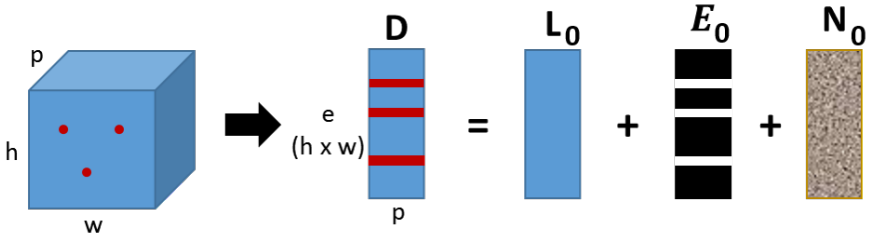


Figure 2.4: A given HSI that contains three targets (the red circles). The HSI is split into a low-rank background HSI $\mathbf{L}_0$, a sparse target HSI $\mathbf{E}_0$, and the noise HSI $\mathbf{N}_0$.

In this regard, we aim to solve the following minimization problem:

$$\min_{\mathbf{L},\mathbf{E}} \left\{ \tau \, \mathrm{rank}(\mathbf{L}) + \lambda \, \|\mathbf{E}\|_{0,2} + \|\mathbf{D} - \mathbf{L} - \mathbf{E}\|_F^2 \right\}, \tag{2.7}$$

where τ controls the rank of $\mathbf{L}$, and λ the sparsity level in $\mathbf{E}$. Note that in our case, the matrix $\mathbf{E}_0$ will be sparse in rows, and hence, the $l_{0,2}$ norm is used for the sparse component.
Again, problem (2.7) is NP-HARD, and thus is solved by the following convex minimization problem (after surrogating the rank term to the nuclear norm and the $l_{0,2}$ norm to the $l_{1,2}$ norm):

$$\min_{\mathbf{L},\mathbf{E}} \left\{ \tau \, \|\mathbf{L}\|_* + \lambda \, \|\mathbf{E}\|_{1,2} + \|\mathbf{D} - \mathbf{L} - \mathbf{E}\|_F^2 \right\}, \tag{2.8}$$

We can observe that unlike the "RPCA-SOP" in (2.6), we have now an $l_{1,2}$ norm for the sparse component (since in our case, the matrix $\mathbf{E}$ should be sparse in rows). In addition, since weights are only relative to one another, we only need two parameters and can fix the remaining one as 1. In our case, we have just replaced the two parameters λ and μ (applied on the second and third term as in problems (2.4) and (2.6)) by τ and λ (applied on the first and second term). The main reason why we do that is just because in our modified RPCA version later, it will be easier to explain the parameters settings with τ and λ instead of λ and μ.

2.2.2 Is the direct use of classical RPCA adequate to distinguish the targets of interest from the background?

The answer is no

In the aforementioned section 2.1.4, we have briefly outlined the success of RPCA in applications such as face recognition and foreground detection. In foreground detection for example, the RPCA searches for the moving objects (but that may not necessarily be the true targets). But if one is interested to test the effect of using RPCA for target detection in hyperspectral imagery, thus, "*what can be the RPCA search?*" and "*is RPCA adequate to identify the targets of interest from the background?*".

Our findings

RPCA searches for small heterogeneous and high contrast objects. RPCA is not adequate to distinguish the targets $\cdots$

It is very important to note that in our hyperspectral imagery application, there are often other small, heterogeneous, high contrast regions that are non-targets. In our extensive evaluations later, we prove that these small (heterogeneous) regions will be deemed as targets under the general RPCA framework. Compounding the decomposition is also the often uniform material present in most targets, which means that they would contribute only a small increase in the rank of the background HSI $\mathbf{L}$ if they were to be grouped in the background HSI. Indeed, some other heterogeneous non-target objects or specular highlights may contribute a larger increase in rank and thus they are more liable to be treated as targets in the decomposition under the general RPCA. In other words, there is a substantial overlap between $\mathbf{L}$ and $\mathbf{E}$ for the general RPCA to be well-posed or work well.

Let us prove our findings

Let us now prove our aforementioned findings by evaluating the RPCA model solved via problem (2.8) on four real hyperspectral images with different background and target complexity. As can be seen from the results (see Figure 2.5 to 2.8), despite the effort to individually tune the parameters for best separation for each of the four images, it was not possible to obtain a clean target and background separation. From all the evaluations below, we prove that RPCA searches for small heterogeneous and high contrast regions that may not necessarily be the true targets of interest.

It is important to mention that from our proven findings in the evaluations below, the failure of the general RPCA is not related to which surrogation penalty has been used for the sparse term, but to the fact that using the general RPCA model without the incorporation of any prior information (either on the low-rank component or on the sparse component) is not capable to disambiguate the true targets from other small heterogeneous and/or high contrast objects.

Hence, a modification of RPCA to distinguish the targets of interest from the background is highly recommended !!!

Evaluation on the Nuance Cri HSI ✓ The Nuance Cri HSI [164, 166] is acquired by the Nuance Cri hyperspectral sensor. It covers an area of 400×400 pixels with 46 spectral bands in wavelengths ranging from 650 to 1100nm. It contains ten rocks targets in a simple background and thus we can obviously observe from Figure 2.5 that this HSI poses no problem for the general RPCA. However the other images below represent more complex background.

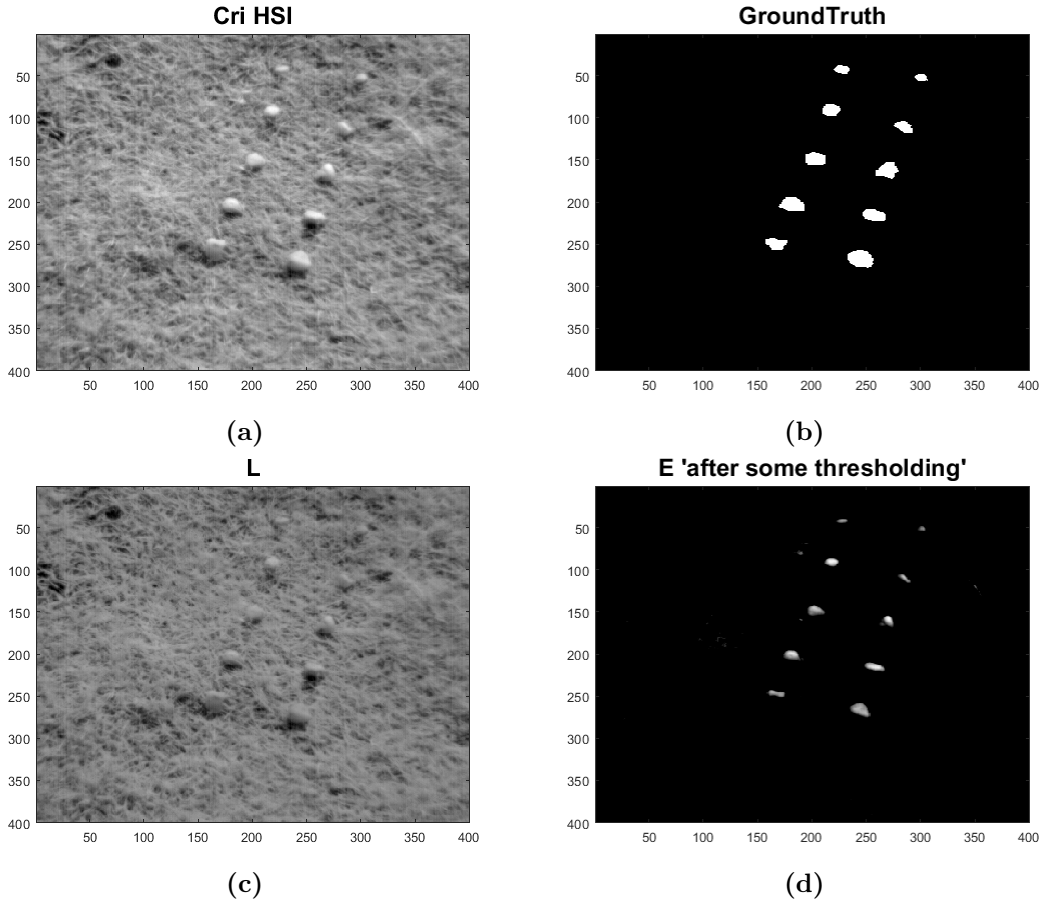


Figure 2.5: Evaluation of the classical RPCA on the Nuance Cri HSI. **(a)**: the original Nuance Cri HSI (we exhibit the mean power in db over the 46 spectral bands). **(b)**: the GroundTruth image for the targets of interest. **(c)**: the low-rank background HSI **L** (mean power in db). **(d)**: the sparse target HSI **E** after some thresholding (mean power in db)

Evaluation on PaviaC HSI ✗ The PaviaC HSI [164] is a selected small zone from Pavia Center City. It is a 100×126 image and consists of 102 bands in wavelengths ranging from 430 to 860nm. The main background materials are bridge and water. There are some vehicles on the bridge and bare soil near the bridge pier, and hence, are considered as the targets to detect. We observe from Figure 2.6 that both the vehicles, bare soil and the bridge pier are being deposited in the sparse image.

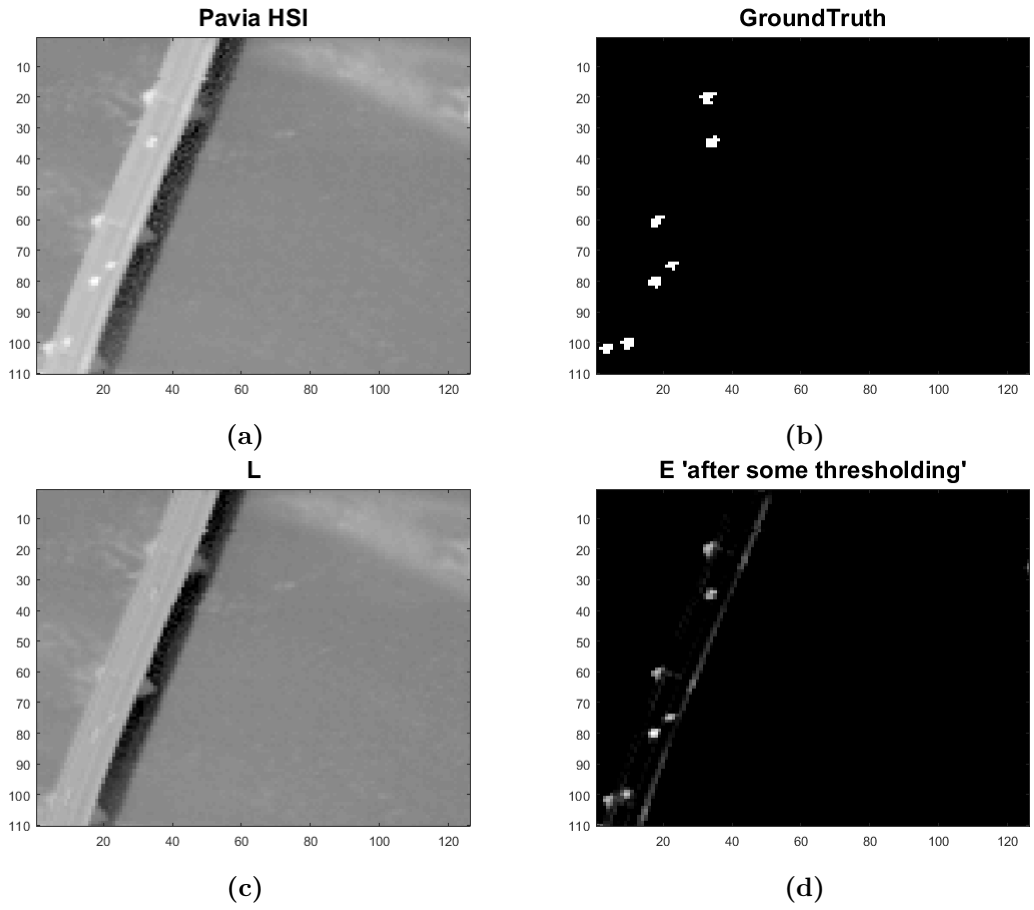


Figure 2.6: Evaluation of the classical RPCA on PaviaC HSI. **(a)**: the original PaviaC HSI (we exhibit the mean power in db over the 102 spectral bands). **(b)**: the GroundTruth image for the targets of interest. **(c)**: the low-rank background HSI $\mathbf{L}$ (mean power in db). **(d)**: the sparse target HSI $\mathbf{E}$ after some thresholding (mean power in db)

Evaluation on DATA HSI ✗ The DATA HSI [135] is a 201×200 image and consisting of 167 spectral bands. It depicts a scrubby terrain with small heterogeneous regions comprised of trees and one vehicle, the latter is being the target of interest in this case. We observe from Figure 2.7 that both the vehicle and trees are being deposited in the sparse target image. Even with a lot of false alarms in the sparse target image, the background is still not completely cleansed of the target.

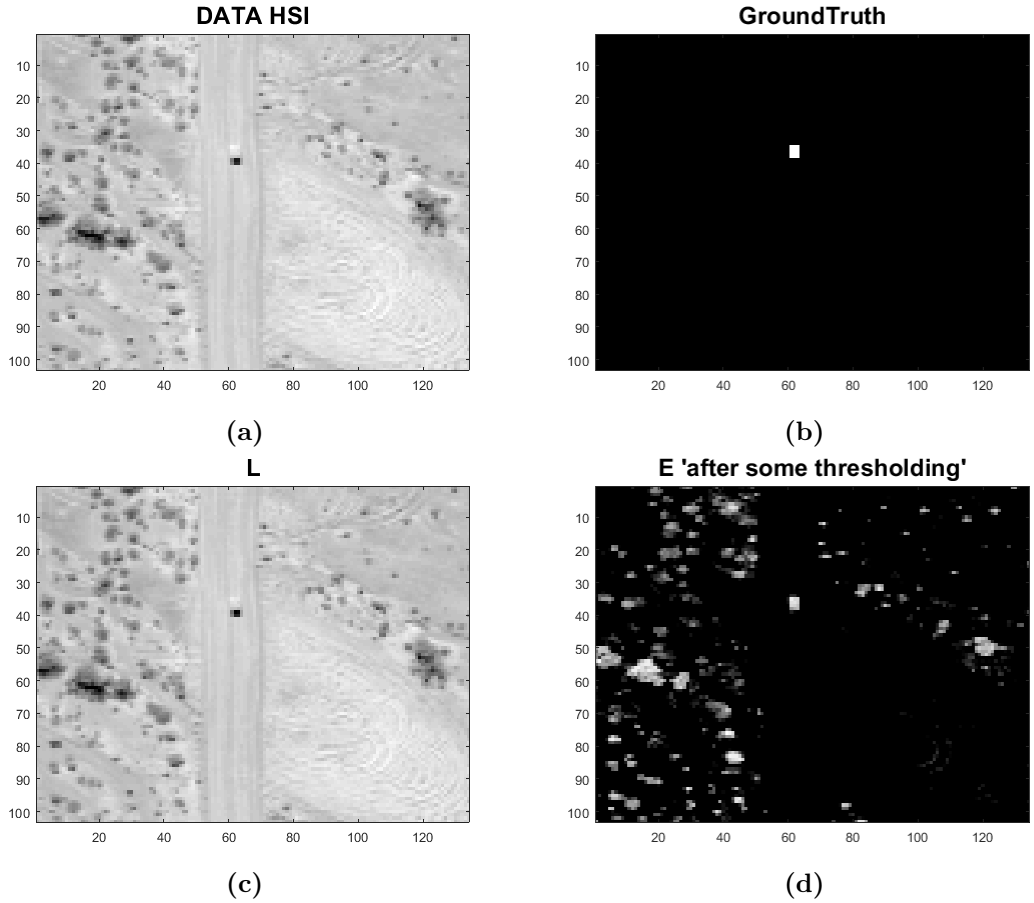


Figure 2.7: Evaluation of the classical RPCA on DATA HSI. **(a)**: the original HSI (we exhibit the mean power in db over the 167 spectral bands). **(b)**: the GroundTruth image for the targets of interest. **(c)**: the low-rank background HSI $\mathbf{L}$ (mean power in db). **(d)**: the sparse target HSI $\mathbf{E}$ after some thresholding (mean power in db)

Evaluation on Cuprite HSI ✗ This HSI [133, 134] is a region of the Cuprite mining district area, of size 250×291 pixels and consisting of 186 spectral bands in wavelengths ranging from 0.4046 to $2.4573\mu m$. In this small zone area, three buddingtonite outcrops (spectrally dominated by buddingtonite) are considered as targets, and their locations are shown in the GroundTruth[1]. It has been noted by Gregg et. al. in [134] that the ammonia in buddingtonite has a distinct N-H

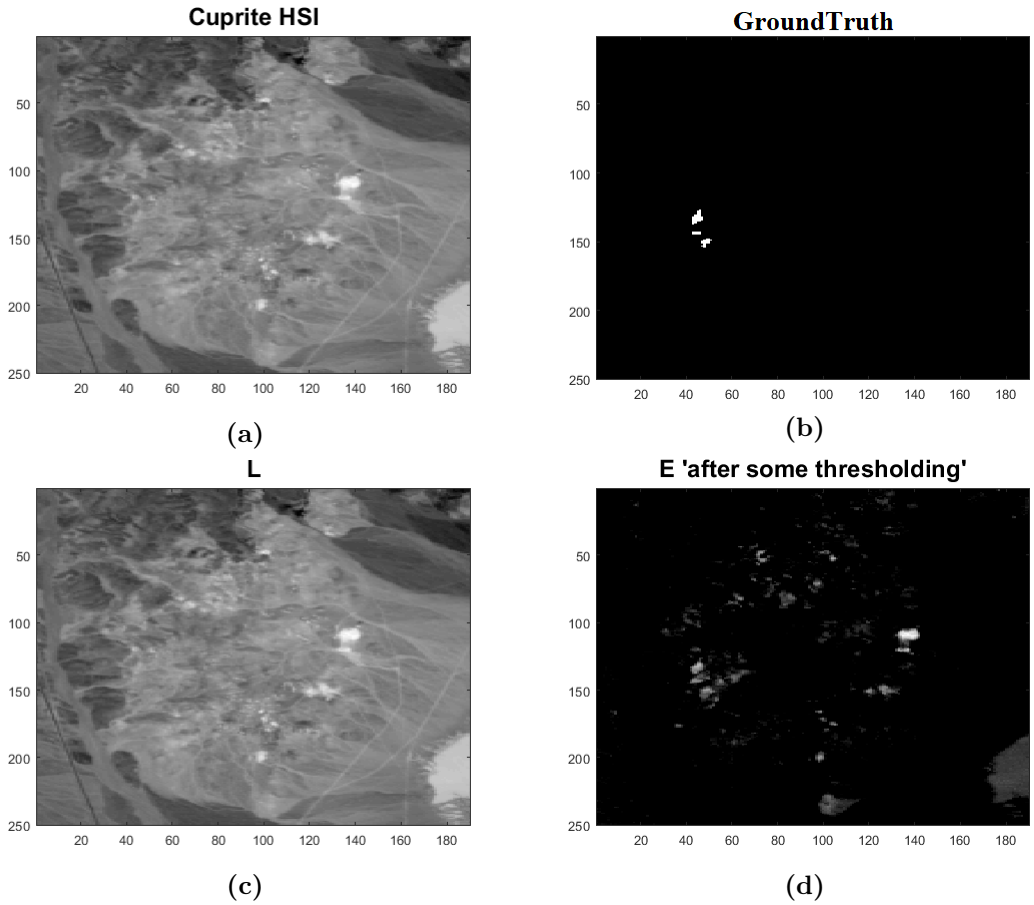


Figure 2.8: Evaluation of classical RPCA on Cuprite HSI. **(a)**: the original HSI (we exhibit the mean power in db over the 186 spectral bands). **(b)**: the GroundTruth image for the targets of interest. **(c)**: the low-rank background HSI **L** (mean power in db). **(d)**: the sparse target HSI **E** after some thresholding (mean power in db)

combination absorption at 2.12μm, a position similar to that of the cellulose absorption in dried vegetation, from which it can be distinguished based on its narrower band width and asymmetry. Hence, the buddingtonite 2.12μm combination absorption is unique in wavelength location relative to those of most other minerals in the image (that is, it is easily recognized based on its unique 2.12μm absorption band). This might be a reason (but we do not think so) of why the general RPCA is able to find those Buddingtonite outcrops in addition to the small heterogeneous and high contrast regions which are also deposited in the sparse target image.

[1]Note that there may also be smaller buddingtonite outcrops in the NE quadrant of the eastern alteration center, but they are spectrally dominated by alunite's absorption.

In addition, one can imagine that the lignin N-H absorption in vegetation would look somewhat like the N-H combination absorption in buddingtonite, but that using more spectral bands better differentiates buddingtonite from lignin in plants. Both are relatively broad absorptions isolated in wavelength space from other absorptions. However, there is actually vegetation at Cuprite – probably between 10 to 15% ground cover, though the buddingtonite areas in this image zone are relatively vegetation free. Thus, the three buddingtonite outcrops in this image zone can be considered to be homogeneous surrounded by areas with more vegetation on the western side of the eastern alteration center.

2.3 Let us modify the RPCA!

2.3.1 Some related works and issues

Whatever the real application may be, somehow the general RPCA model needs to be subject to further assumptions for successfully distinguishing the true target from the background. There have been some recent works that are developed to introduce a subspace basis dictionary in the general RPCA framework, either in the low-rank or sparse component.

Introducing a subspace basis dictionary into the low-rank component

The generalized model of RPCA, named as the Low-Rank Representation (LRR) [97], allows the use of a subspace basis as a dictionary or just uses self-representation to obtain the LRR. More precisely, given a set of data samples each of which can be represented as a linear combination of the bases in a dictionary, LRR aims at finding the lowest-rank representation of all data jointly. The main advantage of LRR over RPCA is that it can handle well the data drawn from a union of multiple subspaces.
The LRR model is defined as follows:

$$\min_{\mathbf{Y},\mathbf{E}} \left\{ \text{rank}(\mathbf{Y}) + \lambda \|\mathbf{E}\|_l \right\} \quad s.t. \quad \mathbf{D} = \mathbf{G}\,\mathbf{Y} + \mathbf{E}\,, \tag{2.9}$$

where again as in (2.1), the $\|.\|_l$ indicates certain sparse regularization strategy such as l_0 norm, $l_{2,0}$ norm, or $l_{0,2}$ norm. $\mathbf{G}$ is a "dictionary" that linearly spans the data space. By setting $\mathbf{G} = \mathbf{I}$, the problem (2.9) falls back to (2.1). Hence, LRR is regarded as a generalization of RPCA that essentially uses the standard bases as the dictionary.

The fact is that LRR is more challenging than RPCA due to the presence of the dictionary matrix $\mathbf{G}$ where its construction needs a very careful attention. More precisely, $\mathbf{G}$ has to be constructed from the background and should not contain the target samples. This is one of the main reasons that limits the use of LRR unless one develops an efficient method to construct an accurate dictionary $\mathbf{G}$. Note that from [97], the traditional method was to set $\mathbf{G}$ to $\mathbf{D}$, which is not an accurate method due to the huge targets contamination in $\mathbf{G}$ and the very high computational complexity to estimate $\mathbf{Y}$ and $\mathbf{E}$.

In this chapter, we will discard the use of LRR due to the big challenge on constructing the dictionary $\mathbf{G}$, and we will only be interested on the use of RPCA and its proposed modified version.

Introducing a subspace basis dictionary in the sparse component

In the earliest models using low-rank matrix to represent the background [25, 149, 168], no prior knowledge on the target was considered. In some applications such as Speech enhancement and hyperspectral imagery, we may expect some prior information about the target of interest, which

can be provided to the user. Thus, incorporating this prior information into the separation scheme in the general RPCA model should allow the user to potentially improve the target extraction performance.
In [37, 132], the authors proposed a speech enhancement system by exploiting the knowledge about the likely form of the targeted speech. This was accomplished by factorizing the sparse component from RPCA into the product of a dictionary of target speech templates and a sparse activation matrix. The proposed methods in [37] and [132] typically differ on how the fixed target dictionary of speech spectral templates is constructed.

2.3.2 Our modification of RPCA

In hyperspectral imagery, as in speech enhancement system, a subsequent prior information about the targets to detect can be provided to the user.
But this raises the question "*what the target prior information can be and how its usage should be dealt with ?*". In real world hyperspectral imagery, this prior target information may not be only related to its spatial properties (e.g. size, shape, texture), which is usually not at our disposal, but to its spectral signature. The latter usually hinges on the nature of the given HSI where the spectra of the targets of interest present have been already measured by some laboratories or with some hand-held spectrometers. For our work, we consider that information is available and we have obtained it from some online spectral libraries.

Our proposed modification method of RPCA is very related to [37, 132], that is, we introduce a subspace basis dictionary in the sparse component by taking into consideration that a target prior information is available to the user. More precisely, our method assumes that the target spectra is known and that the atmospheric influence can be accounted for by the target dictionary $\mathbf{A}_t$. This pre-learned target dictionary $\mathbf{A}_t$ is used to cast the general RPCA into a more specific form, specifically, we further factorize the sparse component $\mathbf{E}$ from RPCA into the product of $\mathbf{A}_t$ and a sparse activation matrix $\mathbf{C}$. This modification is essential to disambiguate the true targets from other small heterogeneous and high contrast regions, and hence, the overlap problem illustrated in the evaluations examples in section 2.2.2 can be much relieved.

Formulating the problem

Suppose an HSI of size $h \times w \times p$, where h and w are the height and width of the image scene, respectively, and p is the number of spectral bands. Our proposed modification of RPCA is mainly based on the following steps:

1. Let us consider that the given HSI contains q pixels $\{\mathbf{x}_i\}_{i\in[1,\,q]}$ of the form:

$$\mathbf{x}_i = \alpha_i\,\mathbf{t}_i + (1-\alpha_i)\,\mathbf{b}_i, \quad 0 < \alpha_i \leq 1\,,$$

where $\mathbf{t}_i$ represents the known target that replaces a fraction α_i of the background $\mathbf{b}_i$ (i.e. at the same spatial location). The remaining $(e-q)$ pixels in the given HSI, with $e = h \times w$, are thus only background ($\alpha = 0$).

2. We assume that all $\{\mathbf{t}_i\}_{i\in[1,\,q]}$ consist of similar materials, thus they should be represented by a linear combination of N_t common target samples $\{\mathbf{a}_j^t\}_{j\in[1,\,N_t]}$, where $\mathbf{a}_j^t \in \mathbb{R}^p$ (the superscript t is for target), but weighted with different set of coefficients $\{\beta_{i,j}\}_{j\in[1,N_t]}$. Thus, each of the q targets is represented as:

$$\mathbf{x}_i = \alpha_i \sum_{j=1}^{N_t} \left(\beta_{i,j}\, \mathbf{a}_j^t\right) + (1-\alpha_i)\, \mathbf{b}_i \qquad i \in [1, q]\,.$$

3. We rearrange the given HSI into a two-dimensional matrix $\mathbf{D} \in \mathbb{R}^{e\times p}$, with $e = h \times w$ (by lexicographically ordering the columns). This matrix $\mathbf{D}$, can be decomposed into a low-rank matrix $\mathbf{L}_0$ representing the pure background, a sparse matrix capturing any spatially small signal residing in the known target subspace, and a noise matrix $\mathbf{N}_0$. More precisely, the model is:

$$\mathbf{D} = \mathbf{L}_0 + (\mathbf{A}_t\, \mathbf{C}_0)^T + \mathbf{N}_0\,,$$

where $(\mathbf{A}_t\mathbf{C}_0)^T$ is the sparse target matrix, ideally with q non-zero rows representing $\alpha_i\{\mathbf{t}_i^T\}_{i\in[1,q]}$, with target dictionary $\mathbf{A}_t \in \mathbb{R}^{p\times N_t}$ having columns representing target samples $\{\mathbf{a}_j^t\}_{j\in[1,N_t]}$, and a coefficient matrix $\mathbf{C}_0 \in \mathbb{R}^{N_t\times e}$ that should be a sparse column matrix, again ideally containing q non-zero columns each representing $\alpha_i[\beta_{i,1}, \cdots, \beta_{i,N_t}]^T$, $i \in [1, q]$. $\mathbf{N}_0$ is assumed to be independent and identically distributed Gaussian noise with zero mean and unknown standard deviation.

4. After reshaping $\mathbf{L}_0$, $(\mathbf{A}_t\, \mathbf{C}_0)^T$ and $\mathbf{N}_0$ back to a cube of size $h \times w \times p$, we call these entities the "low-rank background HSI", "sparse target HSI", and "noise HSI", respectively.

In order to recover the low-rank matrix $\mathbf{L}_0$ and sparse target matrix $(\mathbf{A}_t\mathbf{C}_0)^T$, we consider the following minimization problem:

$$\min_{\mathbf{L},\mathbf{C}} \left\{ \tau\, \mathrm{rank}(\mathbf{L}) + \lambda\, \|\mathbf{C}\|_{2,0} + \left\|\mathbf{D} - \mathbf{L} - (\mathbf{A}_t\mathbf{C})^T\right\|_F^2 \right\}, \tag{2.10}$$

where τ controls the rank of $\mathbf{L}$, and λ the sparsity level in $\mathbf{C}$.

Recovering a low-rank background matrix and a sparse target matrix by convex optimization

As in RPCA-PCP [25], RPCA-SPCP [168], RPCA-OP [152], and RPCA-SOP [152], we relax the rank and the $||.||_{2,0}$ term to their convex proxies. More precisely, we use the nuclear norm $||\mathbf{L}||_*$

as a surrogate for the rank($\mathbf{L}$) term, and the $l_{2,1}$ norm for the $l_{2,0}$ norm.

We now need to solve the following convex minimization problem:

$$\min_{\mathbf{L},\mathbf{C}} \left\{ \tau \, \|\mathbf{L}\|_* + \lambda \, \|\mathbf{C}\|_{2,1} + \left\| \mathbf{D} - \mathbf{L} - (\mathbf{A}_t \mathbf{C})^T \right\|_F^2 \right\} . \tag{2.11}$$

Problem (2.11) is solved via an alternating minimization of two sub-problems. Specifically, at each iteration k:

$$\mathbf{L}^{(k)} = \underset{\mathbf{L}}{\operatorname{argmin}} \left\{ \left\| \mathbf{L} - \left(\mathbf{D} - \left(\mathbf{A}_t \, \mathbf{C}^{(k-1)} \right)^T \right) \right\|_F^2 + \tau \, \|\mathbf{L}\|_* \right\} , \tag{2.12a}$$

$$\mathbf{C}^{(k)} = \underset{\mathbf{C}}{\operatorname{argmin}} \left\{ \left\| \left(\mathbf{D} - \mathbf{L}^{(k)} \right)^T - \mathbf{A}_t \, \mathbf{C} \right\|_F^2 + \lambda \, \|\mathbf{C}\|_{2,1} \right\} . \tag{2.12b}$$

The minimization sub-problems (2.12a) (2.12b) are convex and each can be solved optimally.

Solving sub-problem (2.12a)**:** We solve sub-problem (2.12a) via the Singular Value Thresholding operator [23]. We assume that $\left(\mathbf{D} - \left(\mathbf{A}_t \, \mathbf{C}^{(k-1)} \right)^T \right)$ has a rank equal to r. According to theorem 2.1 in [23], sub-problem (2.12a) admits the following closed-form solution:

$$\boxed{\mathbf{L}^{(k)} = D_{\tau/2} \left(\mathbf{D} - \left(\mathbf{A}_t \, \mathbf{C}^{(k-1)} \right)^T \right) = \mathbf{U}^{(k)} \, D_{\tau/2} \left(\mathbf{S}^{(k)} \right) \, \mathbf{V}^{(k)T} = \mathbf{U}^{(k)} \, \operatorname{diag} \left(\left\{ \left(s_t^{(k)} - \tfrac{\tau}{2} \right)_+ \right\} \right) \mathbf{V}^{(k)T}}$$

where $\mathbf{S}^{(k)} = \operatorname{diag} \left(\left\{ s_t^{(k)} \right\}_{1 \le t \le r} \right)$, and $D_{\tau/2}(.)$ is the singular value shrinkage operator.
The matrices $\mathbf{U}^{(k)} \in \mathbb{R}^{e \times r}$, $\mathbf{S}^{(k)} \in \mathbb{R}^{r \times r}$ and $\mathbf{V}^{(k)} \in \mathbb{R}^{p \times r}$ are generated by the singular value decomposition of $\left(\mathbf{D} - \left(\mathbf{A}_t \, \mathbf{C}^{(k-1)} \right)^T \right)$.

Proof. Since the function $\left\{ \left\| \mathbf{L} - \left(\mathbf{D} - \left(\mathbf{A}_t \, \mathbf{C}^{(k-1)} \right)^T \right) \right\|_F^2 + \tau \, \|\mathbf{L}\|_* \right\}$ is strictly convex, it is easy to see that there exists a unique minimizer, and we thus need to prove that it is equal to $D_{\tau/2} \left(\mathbf{D} - \left(\mathbf{A}_t \, \mathbf{C}^{(k-1)} \right)^T \right)$. Note that to understand how the aforementioned closed-form solution has been obtained, we provide in detail the proof steps that have been given in [23].
To do this, let us first find the derivative of sub-problem (2.12a) w.r.t. $\mathbf{L}$ and set it to zero. We obtain:

$$\left(\mathbf{D} - \left(\mathbf{A}_t \, \mathbf{C}^{(k-1)} \right)^T \right) - \hat{\mathbf{L}} = \frac{\tau}{2} \, \partial \left\| \hat{\mathbf{L}} \right\|_* , \tag{2.13}$$

where $\partial \left\| \hat{\mathbf{L}} \right\|_*$ is the set of subgradients of the nuclear norm. Let $\mathbf{U}_L \, \mathbf{S}_L \, \mathbf{V}_L^T$ be the Singular Value Decomposition (SVD) of $\mathbf{L}$, it is known [26, 96, 147] that

$$\partial \, \|\mathbf{L}\|_* = \left\{ \mathbf{U}_L \, \mathbf{V}_L^T + \mathbf{W} \, : \, \mathbf{W} \in \mathbb{R}^{e \times p}, \, \mathbf{U}_L^T \, \mathbf{W} = \mathbf{0}, \, \mathbf{W} \, \mathbf{V}_L = \mathbf{0}, \, \|\mathbf{W}\|_2 \le 1 \right\} .$$

Set $\hat{\mathbf{L}} = D_{\tau/2} \left(\mathbf{D} - \left(\mathbf{A}_t \, \mathbf{C}^{(k-1)} \right)^T \right)$ for short. In order to show that $\hat{\mathbf{L}}$ obeys eq. (2.13), suppose the SVD of $\left(\mathbf{D} - \left(\mathbf{A}_t \, \mathbf{C}^{(k-1)} \right)^T \right)$ is given by:

$$\left(\mathbf{D} - \left(\mathbf{A}_t \, \mathbf{C}^{(k-1)} \right)^T \right) = \mathbf{U}_0 \, \mathbf{S}_0 \, \mathbf{V}_0^T + \mathbf{U}_1 \, \mathbf{S}_1 \, \mathbf{V}_1^T ,$$

where $\mathbf{U}_0$, $\mathbf{V}_0$ (resp. $\mathbf{U}_1$, $\mathbf{V}_1$) are the singular vectors associated with singular values larger than $\tau/2$ (resp. inferior than or equal to $\tau/2$). With these notations, we have:

$$\hat{\mathbf{L}} = D_{\tau/2}\left(\mathbf{U}_0\,\mathbf{S}_0\,\mathbf{V}_0^T\right) = \left(\mathbf{U}_0\,\left(\mathbf{S}_0 - \frac{\tau}{2}\,\mathbf{I}\right)\,\mathbf{V}_0^T\right).$$

Thus, if we return back to eq. (2.13), we obtain:

$$\begin{aligned}
\mathbf{U}_0\,\mathbf{S}_0\,\mathbf{V}_0^T + \mathbf{U}_1\,\mathbf{S}_1\,\mathbf{V}_1^T - \mathbf{U}_0\,\left(\mathbf{S}_0 - \frac{\tau}{2}\,\mathbf{I}\right)\,\mathbf{V}_0^T &= \frac{\tau}{2}\,\partial\left\|\hat{\mathbf{L}}\right\|_*, \\
\Rightarrow \mathbf{U}_1\,\mathbf{S}_1\,\mathbf{V}_1^T + \mathbf{U}_0\,\frac{\tau}{2}\,\mathbf{V}_0^T &= \frac{\tau}{2}\,\partial\left\|\hat{\mathbf{L}}\right\|_*, \\
\Rightarrow \left(\mathbf{U}_0\,\mathbf{V}_0^T + \mathbf{W}\right) &= \partial\left\|\hat{\mathbf{L}}\right\|_*,
\end{aligned}$$

where $\mathbf{W} = \dfrac{2}{\tau}\,\mathbf{U}_1\,\mathbf{S}_1\,\mathbf{V}_1^T$.

By definition, $\mathbf{U}_0^T\,\mathbf{W} = \mathbf{0}$, $\mathbf{W}\,\mathbf{V}_0 = \mathbf{0}$, and we also have $\|\mathbf{W}\|_2 \leq 1$.
Hence, $\left(\mathbf{D} - \left(\mathbf{A}_t\,\mathbf{C}^{(k-1)}\right)^T\right) - \hat{\mathbf{L}} = \dfrac{\tau}{2}\,\partial\left\|\hat{\mathbf{L}}\right\|_*$, which concludes the proof. ■

Solving sub-problem (2.12b)**:** (2.12b) can be solved by various methods, among which we adopt the Alternating Direction Method of Multipliers (ADMM) [20]. More precisely, we introduce an auxiliary variable $\mathbf{F}$ into sub-problem (2.12b) and recast it into the following form:

$$\left(\mathbf{C}^{(k)}, \mathbf{F}^{(k)}\right) = \underset{s.t.\ \ \mathbf{C}=\mathbf{F}}{\operatorname{argmin}} \left\{ \left\| \left(\mathbf{D} - \mathbf{L}^{(k)}\right)^T - \mathbf{A}_t\,\mathbf{C} \right\|_F^2 + \lambda\,\|\mathbf{F}\|_{2,1} \right\}. \tag{2.14}$$

Problem (2.14) is then solved as follows (scaled form of ADMM):

$$\mathbf{C}^{(k)} = \underset{\mathbf{C}}{\operatorname{argmin}} \left\{ \left\| \left(\mathbf{D} - \mathbf{L}^{(k)}\right)^T - \mathbf{A}_t\,\mathbf{C} \right\|_F^2 + \frac{\rho^{(k-1)}}{2} \left\| \mathbf{C} - \mathbf{F}^{(k-1)} + \frac{1}{\rho^{(k-1)}}\,\mathbf{Z}^{(k-1)} \right\|_F^2 \right\}, \tag{2.15a}$$

$$\mathbf{F}^{(k)} = \underset{\mathbf{F}}{\operatorname{argmin}} \left\{ \lambda\,\|\mathbf{F}\|_{2,1} + \frac{\rho^{(k-1)}}{2} \left\| \mathbf{C}^{(k)} - \mathbf{F} + \frac{1}{\rho^{(k-1)}}\,\mathbf{Z}^{(k-1)} \right\|_F^2 \right\}, \tag{2.15b}$$

$$\mathbf{Z}^{(k)} = \mathbf{Z}^{(k-1)} + \rho^{(k-1)}\,\left(\mathbf{C}^{(k)} - \mathbf{F}^{(k)}\right), \tag{2.15c}$$

where $\mathbf{Z} \in \mathbb{R}^{N_t \times e}$ is the Lagrangian multiplier matrix, and ρ is a positive scalar.

Solving sub-problem (2.15a):

$$\begin{aligned}
-2\,\mathbf{A}_t^T\left(\left(\mathbf{D} - \mathbf{L}^{(k)}\right)^T - \mathbf{A}_t\,\mathbf{C}\right) + \rho^{(k-1)}\,\left(\mathbf{C} - \mathbf{F}^{(k-1)} + \frac{1}{\rho^{(k-1)}}\,\mathbf{Z}^{(k-1)}\right) &= \mathbf{0}, \\
\Rightarrow \left(2\,\mathbf{A}_t^T\,\mathbf{A}_t + \rho^{(k-1)}\,\mathbf{I}\right)\,\mathbf{C} = \rho^{(k-1)}\,\mathbf{F}^{(k-1)} - \mathbf{Z}^{(k-1)} + 2\,\mathbf{A}_t^T\,\left(\mathbf{D} - \mathbf{L}^{(k)}\right)^T&.
\end{aligned}$$

This implies:

$$\boxed{\mathbf{C}^{(k)} = \left(2\,\mathbf{A}_t^T\,\mathbf{A}_t + \rho^{(k-1)}\,\mathbf{I}\right)^{-1}\,\left(\rho^{(k-1)}\,\mathbf{F}^{(k-1)} - \mathbf{Z}^{(k-1)} + 2\,\mathbf{A}_t^T\,\left(\mathbf{D} - \mathbf{L}^{(k)}\right)^T\right)}$$

Solving sub-problem (2.15b):

According to Lemma 3.3 in [156] and Lemma 4.1 in [97], sub-problem (2.15b) admits the following closed form solution:

$$\boxed{[\mathbf{F}]_{:,j}^{(k)} = \max\left(\left\|[\mathbf{C}]_{:,j}^{(k)} + \frac{1}{\rho^{(k-1)}}\,[\mathbf{Z}]_{:,j}^{(k-1)}\right\|_2 - \frac{\lambda}{\rho^{(k-1)}},\, 0\right)\left(\frac{[\mathbf{C}]_{:,j}^{(k)} + \frac{1}{\rho^{(k-1)}}\,[\mathbf{Z}]_{:,j}^{(k-1)}}{\left\|[\mathbf{C}]_{:,j}^{(k)} + \frac{1}{\rho^{(k-1)}}\,[\mathbf{Z}]_{:,j}^{(k-1)}\right\|_2}\right)}$$

Proof. At the j^{th} column, sub-problem (2.15b) refers to:

$$[\mathbf{F}]_{:,j}^{(k)} = \underset{[\mathbf{F}]_{:,j}}{\operatorname{argmin}}\left\{\lambda\left\|[\mathbf{F}]_{:,j}\right\|_2 + \frac{\rho^{(k-1)}}{2}\left\|[\mathbf{C}]_{:,j}^{(k)} - [\mathbf{F}]_{:,j} + \frac{1}{\rho^{(k-1)}}\,[\mathbf{Z}]_{:,j}^{(k-1)}\right\|_2^2\right\}.$$

By finding the derivative w.r.t $[\mathbf{F}]_{:,j}$ and setting it to zero, we obtain:

$$-\rho^{(k-1)}\left([\mathbf{C}]_{:,j}^{(k)} - [\mathbf{F}]_{:,j} + \frac{1}{\rho^{(k-1)}}\,[\mathbf{Z}]_{:,j}^{(k-1)}\right) + \frac{\lambda\,[\mathbf{F}]_{:,j}}{\left\|[\mathbf{F}]_{:,j}\right\|_2} = \mathbf{0}$$

$$\Rightarrow [\mathbf{C}]_{:,j}^{(k)} + \frac{1}{\rho^{(k-1)}}\,[\mathbf{Z}]_{:,j}^{(k-1)} = [\mathbf{F}]_{:,j} + \frac{\lambda\,[\mathbf{F}]_{:,j}}{\rho^{(k-1)}\left\|[\mathbf{F}]_{:,j}\right\|_2}. \tag{2.16}$$

By computing the l_2 norm of (2.16), we obtain:

$$\left\|[\mathbf{C}]_{:,j}^{(k)} + \frac{1}{\rho^{(k-1)}}\,[\mathbf{Z}]_{:,j}^{(k-1)}\right\|_2 = \left\|[\mathbf{F}]_{:,j}\right\|_2 + \frac{\lambda}{\rho^{(k-1)}}. \tag{2.17}$$

From equation (2.16) and equation (2.17), we have:

$$\frac{[\mathbf{C}]_{:,j}^{(k)} + \frac{1}{\rho^{(k-1)}}\,[\mathbf{Z}]_{:,j}^{(k-1)}}{\left\|[\mathbf{C}]_{:,j}^{(k)} + \frac{1}{\rho^{(k-1)}}\,[\mathbf{Z}]_{:,j}^{(k-1)}\right\|_2} = \frac{[\mathbf{F}]_{:,j}}{\left\|[\mathbf{F}]_{:,j}\right\|_2}. \tag{2.18}$$

Consider that:

$$[\mathbf{F}]_{:,j} = \|[\mathbf{F}]_{:,j}\|_2 \times \frac{[\mathbf{F}]_{:,j}}{\|[\mathbf{F}]_{:,j}\|_2}. \tag{2.19}$$

By replacing $\|[\mathbf{F}]_{:,j}\|_2$ from (2.17) into (2.19), and $\frac{[\mathbf{F}]_{:,j}}{\|[\mathbf{F}]_{:,j}\|_2}$ from (2.18) into (2.19), we conclude the proof. ■

Some initializations and convergence criterion

We initialize $\mathbf{L}^{(0)} = \mathbf{C}^{(0)} = \mathbf{F}^{(0)} = \mathbf{Z}^{(0)} = \mathbf{0}$, $\rho^{(0)} = 10^{-4}$ and update $\rho^{(k)} = 1.1\,\rho^{(k-1)}$. The criteria for convergence of sub-problem (2.12b) is $\left\|\mathbf{C}^{(k)} - \mathbf{F}^{(k)}\right\|_F^2 \leq 10^{-6}$.

For Problem (2.11), we stop the iteration when the following convergence criterion is satisfied:

$$\frac{\left\|\mathbf{L}^{(k)} - \mathbf{L}^{(k-1)}\right\|_F}{\|\mathbf{D}\|_F} \leq \epsilon \quad \text{and} \quad \frac{\left\|\left(\mathbf{A}_t\,\mathbf{C}^{(k)}\right)^T - \left(\mathbf{A}_t\,\mathbf{C}^{(k-1)}\right)^T\right\|_F}{\|\mathbf{D}\|_F} \leq \epsilon$$

where $\epsilon > 0$ is a precision tolerance parameter. We set $\epsilon = 10^{-4}$.

2.3.3 Our target detection strategy in this chapter: let us introduce our novel target detector $(\mathbf{A}_t\mathbf{C})^T$

We use $(\mathbf{A}_t\mathbf{C})^T$ directly as a detector. Note that for this detection strategy, we require as few false alarms as possible to be deposited in the target image, but we do not need the target fraction to be entirely removed from the background (that is, a very weak target separation can suffice). As long as enough of the target fractions are moved to the target image such that non-zero support is detected at the corresponding pixel location, it will be adequate for our detection scheme. From this standpoint, we should choose a λ that is relatively large, so that the target image is really sparse with zero or little false alarms, and only signals that reside in the target subspace specified by $\mathbf{A}_t$ will be deposited there.

2.3.4 Construction of the target dictionary $\mathbf{A}_t$

An important problem that requires a very careful attention is the construction of an appropriate dictionary $\mathbf{A}_t$ in order to well capture the target to separate from the background. In reality, the target present in the HSI can be highly affected by the atmospheric conditions, sensor noise, and material composition that may produce huge variations on the target spectra. In view of these real effects, it is very difficult to model the target dictionary ($\mathbf{A}_t$) well. But this raises the question on "*how these effects should be dealt with?*".

Some scenarios for modelling the target dictionary have been followed over several decades. For example, by using physical models and the MODTRAN atmospheric modeling program [9], target spectral signatures can be generated under various atmospheric conditions. For simplicity, we handle this problem by exploiting target samples that are available in some online spectral libraries. More precisely, $\mathbf{A}_t$ can be constructed via the United States Geological Survey (USGS - Reston) Spectral Library [38]. However, the user can also deal with the Advanced Spaceborne Thermal Emission and Reflection (ASTER) spectral library [5] that includes data from the USGS Spectral Library, the Johns Hopkins University (JHU) Spectral Library, and the Jet Propulsion Laboratory (JPL) Spectral Library.

For instance, it is usually difficult to find, for a specific given target, a sufficient number of available samples in the online spectral libraries. Hence, the dictionary $\mathbf{A}_t$ can still not be sufficiently selective and accurate. This is the most reason why problem (2.11) can fail to well capture the target.

Summary

The classical Robust Principal Component Analysis (RPCA) has shown encouraging performances is some applications such as face recognition and foreground detection. In this chapter, we aimed to exploit the RPCA for hyperspectral target detection. More precisely, the given HSI is regarded as being made up of the sum of a low-rank background HSI $\mathbf{L}$ and a sparse target HSI $\mathbf{E}$. By evaluating the RPCA on four real hyperspectral images, we proved that its direct

use is inadequate to distinguish the targets of interest from their surrounding background. In order to address this issue, we have assumed that the spectra of the target is provided to the user. Hence, we highly expect that by incorporating this target prior information into the RPCA model can potentially help to identify the true targets. To do this, we have factorized the sparse component $\mathbf{E}$ from RPCA into the product of a target dictionary $\mathbf{A}_t$ and a sparse activation matrix $\mathbf{C}$. As we will observe from the experiments in Chapter 4, this modification is very essential to disambiguate the true targets from the background.

All intelligent thoughts have already been thought;
what is necessary is only to try to think them again.

— Johann Wolfgang von Goethe

3

Improving Background Dictionary Construction for Sparse Representation Methods

Contents

► *Synopsis* Recall that in Chapter 2, a modified version of the Robust Principal Component Analysis (RPCA) has been developed, and by which the sparse component $(\mathbf{A}_t\mathbf{C})^T$ was the object of interest and was directly used for the detection. But "*can the low-rank component also be exploited to provide another way of an automatic target detection in hyperspectral imagery?*". Indeed, this chapter reuses the same modified version of RPCA that has been developed in Chapter 2, but now the low-rank component $\mathbf{L}$ is being the object of interest. More precisely, the low-rank component $\mathbf{L}$ is exploited to improve the background dictionary $\mathbf{A}_b$ construction for sparse representation methods, and hence, potentially improving the target detection performance. This chapter is split into two parts:

1. **The first part 3.1** "*Background and System Overview*" first overviews the sparse representation methods for hyperspectral test pixels. Then, it outlines one of the famous greedy algorithms called "Orthogonal Matching Pursuit", as well as the usual method for the $\mathbf{A}_b$ construction. Next, it describes the dictionary-based target detection called "SRBBH

detector" that will be used in our dissertation. It ends with a system overview of the proposed work.

2. **The second part 3.2** "*Let us improve the usual* $\mathbf{A}_b$ *construction method*" outlines how the $\mathbf{A}_b$ construction is improved using the low-rank component from the modified version of RPCA that has been developed in Chapter 2.

3.1 Background and System Overview

Recent years have witnessed a growing interest on the notion of sparsity as a way to model signals. The basic assumption of this model is that natural signals can be represented as a "sparse" linear combination of atom signals taken from a dictionary. In this regard, two main issues need to be addressed:

- How to represent a signal in the sparsest way, for a given dictionary,
- How to construct an accurate dictionary in order to successfully representing the signal.

3.1.1 Sparse Representation for hyperspectral test pixels

Recently, a signal classification technique via sparse representation was developed for face recognition application [148]. It is observed that aligned faces of the same object with varying lighting conditions approximately lie in a low-dimensional subspace [7]. Hence, a test face image can be sparsely represented by atom signals from all classes. This representation approach has also been exploited in several other signal classification problems such as iris recognition [122], tumor classification [76], and hyperspectral imagery unmixing [72].

In this context, Chen *et al.* [33] have been inspired by the work done for face recognition application in [148], and developed an approach for sparse representation of hyperspectral test pixels. In particular, each test pixel $\mathbf{x} \in \mathbb{R}^p$ (target or background) in a given HSI, is assumed to lie in a low-dimensional subspace of the p-dimensional spectral-measurement space, and thus can be represented by a very few atom signals taken from the dictionaries. For example, if a test pixel $\mathbf{x}$ contains the target (that is, $\mathbf{x} = \alpha\,\mathbf{t} + (1-\alpha)\,\mathbf{b}$, with $0 < \alpha \leq 1$), it can be sparsely represented by atom signals taken from the target dictionary (denoted as $\mathbf{A}_t$); whereas, if $\mathbf{x}$ is only a background pixel (it does not contain the target, e.g., $\alpha = 0$), it can be sparsely represented by atom signals taken from the background dictionary (denoted as $\mathbf{A}_b$).
Very recently, Zhang *et al.* [163] have extended the work done by Chen *et al.* [33] by combining the idea of binary hypothesis and sparse representation together, obtaining a more complete and realistic sparsity model than in [33]. More precisely, Zhang *et al.* [163] have assumed that if the test pixel $\mathbf{x}$ belongs to hypothesis H_0 (target absent), it will be modeled by the $\mathbf{A}_b$ only; otherwise, it will be modeled by the union of $\mathbf{A}_b$ and $\mathbf{A}_t$. This in fact yields a competition between

the two hypotheses corresponding to the different pixel class label. The sparse representation model of [163] will be considered throughout this chapter.

The sparsity model of [163]

Using an overcomplete dictionary, either for the background (denoted as $\mathbf{A}_b$) or target (denoted as $\mathbf{A}_t$), the test signal $\mathbf{x} \in \mathbb{R}^p$ can be approximately represented by a very few atom signals taken either from $\mathbf{A}_b$ (if $\mathbf{x}$ belongs to H_0) or from the uinon of $\mathbf{A}_b$ and $\mathbf{A}_t$ (if $\mathbf{x}$ belongs to H_1).

More precisely, if $\mathbf{x} \in H_0$:

$$\begin{aligned} \mathbf{x} &= \varrho_1 \, \mathbf{a}_1^b + \varrho_2 \, \mathbf{a}_2^b + \cdots + \varrho_{N_b} \, \mathbf{a}_{N_b}^b \, , \\ &= \left[\mathbf{a}_1^b, \, \mathbf{a}_2^b, \, \cdots \, , \, \mathbf{a}_{N_b}^b\right] \, \left[\varrho_1, \, \varrho_2, \, \cdots \, , \, \varrho_{N_b}\right]^T \, , \\ &= \mathbf{A}_b \, \boldsymbol{\varrho} \, , \end{aligned}$$

where N_b is the number of background samples (number of columns in $\mathbf{A}_b$), $\mathbf{A}_b$ is a $p \times N_b$ background dictionary whose columns are the background atom signals $\mathbf{a}_1^b, \, \mathbf{a}_2^b, \, \cdots \, , \, \mathbf{a}_{N_b}^b$, and $\boldsymbol{\varrho} \in \mathbb{R}^{N_b}$ is an unknown vector (and that should be sparse) whose entries are the abundances of the corresponding samples in $\mathbf{A}_b$.

If $\mathbf{x} \in H_1$:

$$\begin{aligned} \mathbf{x} &= \beta_1 \, \mathbf{a}_1^b + \beta_2 \, \mathbf{a}_2^b + \cdots + \beta_{N_b} \, \mathbf{a}_{N_b}^b + \theta_1 \, \mathbf{a}_1^t + \theta_2 \, \mathbf{a}_2^t + \cdots + \theta_{N_t} \, \mathbf{a}_{N_t}^t \, , \\ &= \left[\mathbf{A_b} \, \mathbf{A}_t\right] \, \left[\boldsymbol{\beta}^T \, \boldsymbol{\theta}^T\right]^T \, , \\ &= \mathbf{A} \, \boldsymbol{\gamma}, \end{aligned}$$

where N_t is the number of target samples (number of columns in $\mathbf{A}_t$), $\mathbf{A}_t$ is a $p \times N_t$ target dictionary whose columns are the target samples $\mathbf{a}_1^t, \, \mathbf{a}_2^t, \, \cdots \, , \, \mathbf{a}_{N_t}^t$, $\mathbf{A} \in \mathbb{R}^{p \times (N_b + N_t)}$ is the union of $\mathbf{A}_b$ and $\mathbf{A}_t$, $\boldsymbol{\gamma} \in \mathbb{R}^{N_b + N_t}$ is an unknown vector (and that should be sparse) whose entries are the abundances of the corresponding samples in $\mathbf{A}$.

Note that both $\boldsymbol{\varrho}$ and $\boldsymbol{\gamma}$ turn out to be sparse vectors (i.e., a vector with only few nonzero entries).

Sparsity reconstruction

This section considers the reconstruction problem of finding the sparse vectors $\boldsymbol{\varrho}$ and $\boldsymbol{\gamma}$ for a test pixel $\mathbf{x}$ given the dictionaries $\mathbf{A}_b$ and $\mathbf{A}_t$. Both the representation $\boldsymbol{\varrho}$ satisfying $\mathbf{A}_b \, \boldsymbol{\varrho} = \mathbf{x}$, and $\boldsymbol{\gamma}$ satisfying $\mathbf{A} \, \boldsymbol{\gamma} = \mathbf{x}$, can be obtained by solving the following (non-convex and NP-HARD) optimization problems for the sparsest vector:

$$\hat{\boldsymbol{\varrho}} = \underset{\boldsymbol{\varrho}}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{A}_b\boldsymbol{\varrho}\|_2 \quad \text{s.t.} \quad \|\boldsymbol{\varrho}\|_0 \leq k_0\,, \tag{3.1a}$$

$$\hat{\boldsymbol{\gamma}} = \underset{\boldsymbol{\gamma}}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{A}\boldsymbol{\gamma}\|_2 \quad \text{s.t.} \quad \|\boldsymbol{\gamma}\|_0 \leq k_0'\,. \tag{3.1b}$$

In fact, k_0 and k_0' are a given upper bound on the sparsity level [144]. In fact, if k_0 and k_0' are set differently, this can lead to significantly weakened competition between hypotheses H_0 and H_1. That is why, and to greatly reduce the complexity of the parameter adjustment, k_0 and k_0' are set equally to each other [163].

Considering the optimization tasks for the synthesis and analysis models, it is unclear how to solve efficiently these problems, given that they are neither smooth nor convex. Indeed, the two sub-problems posed in equation (3.1a) and (3.1b) are known to be NP-HARD [115]. This limitation leads the researchers to seek approximation algorithms. There are two main approaches to approximate the solution to the sub-problems posed in (3.1a) and (3.1b):

- The first approach is the greedy family of methods (see [22, 44, 52, 99, 100, 121] for more details). In this approach, we find algorithms such as Thresholding, Matching Pursuit (MP) [100], and Orthogonal Matching Pursuit (OMP) [121].
- The second approach is the relaxation methods, which attempt to solve the problem by smoothing the l_0-norm and using continuous optimization techniques. All these techniques are commonly referred to as pursuit algorithms. For example, the popular relaxation techniques is to substitute the problem with a simpler, one by replacing the l_0 norm by an l_g norm with $g = 1$ or $g \leq 1$ as has been done in [31] and [71].

3.1.2 The Orthogonal Matching Pursuit (OMP) greedy algorithm

In this section, we give a detailed description of the Orthogonal Matching Pursuit (OMP) algorithm as given in [3, 24, 142]. Note that we only detail the OMP to solve sub-problem (3.1a). Obviously, the same concept can be applied to solve sub-problem (3.1b).

The OMP algorithm

The OMP is a stepwise forward selection algorithm and is easy to implement. For any subset $S \subseteq \{1, \cdots, N_b\}$, denote by $\mathbf{A}_b(S)$ a submatrix of $\mathbf{A}_b$ containing the columns $\mathbf{a}_h^b$ with $h \in S$. To identify $\boldsymbol{\varrho}$ in sub-problem (3.1a), we need to determine which columns of $\mathbf{A}_b$ participate in the measurement vector $\mathbf{x}$. The idea behind the OMP algorithm is to pick columns from $\mathbf{A}_b$ in a *greedy fashion*. In each iteration, we choose a column from $\mathbf{A}_b$ that is most strongly correlated with the remaining part of $\mathbf{x}$. Then, we substract off its contribution to $\mathbf{x}$ and iterate on the residual. The outputs are $\hat{\boldsymbol{\varrho}}$ which is the estimate for the ideal signal $\boldsymbol{\varrho}$. The basic OMP algorithm is described as follows:

1. *Step 1*: Initialize the residual $\mathbf{r}_0 = \mathbf{x}$, the index set $\Lambda_0 = \phi$. Let the iteration counter $s = 1$.

2. *Step 2*: Find the index Ω_s that solves the maximization problem

$$\Omega_s = \underset{h=1,\cdots,N_b}{\operatorname{argmax}} \left| < \mathbf{r}_{s-1}, \mathbf{a}_h^b > \right| \tag{3.2}$$

   and add the index to the set of selected indices. That is, augment the index set $\Lambda_s = \Lambda_{s-1} \cup \{\Omega_s\}$.

3. *Step 3*: Note that if the maximum occurs for multiple indices, the tie is broken deterministically.

4. *Step 4*: Let $\mathbf{P}_s = \mathbf{A}_b(\Lambda_s) \left((\mathbf{A}_b(\Lambda_s))^T \mathbf{A}_b(\Lambda_s) \right)^{-1} (\mathbf{A}_b(\Lambda_s))^T$ denote the projection onto the linear space spanned by the elements of $\mathbf{A}_b(\Lambda_s)$.

5. *Step 5*: Update $\mathbf{r}_s = (\mathbf{I} - \mathbf{P}_s)\mathbf{x}$.

6. *Step 6*: Increment s, and return to Step 2 if $s < k_0$.

7. *Step 7*: The estimate $\hat{\boldsymbol{\varrho}}$ for the ideal signal $\boldsymbol{\varrho}$ has nonzero indices at the components listed in $\Lambda_{k_0} = \{\Omega_s\}_{s=1}^{k_0}$. The value of the estimate $\hat{\boldsymbol{\varrho}}$ in component Ω_s equals the s^{th} component of $\mathbf{P}_s$.

3.1.3 The usual adaptive method for the $\mathbf{A}_b$ construction

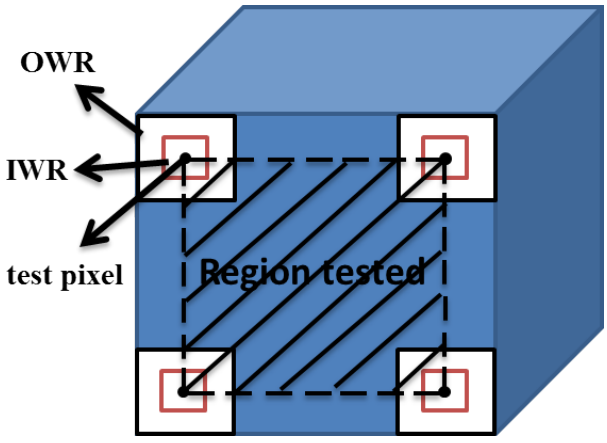


Figure 3.1: The sliding dual concentric window across the HSI

The construction of a locally adaptive $\mathbf{A}_b$ is a very challenging problem since a contamination of it by the target pixels can potentially affect the target detection performance. Usually, the adaptive scheme (and which is used by Chen *et al.* in [33] and Zhang *et al.* in [163]) is based on a dual concentric window centered on the test pixel (Figure 3.1), with an Inner Window Region (IWR) centered within an Outer Window Region (OWR), and only the pixels in the OWR will constitute

the samples for $\mathbf{A}_b$. In other words, if the size of OWR is $m_1 \times m_1$ and the size of IWR is $m_2 \times m_2$, with $m_2 < m_1$, then the total number of pixels in the OWR that will form $\mathbf{A}_b$ is $m_1^2 - m_2^2$.

3.1.4 The SRBBH detector

From the aforementioned sparsity model in 3.1.1, Zhang *et al.* in [163] have considered the following target detector, called "SRBBH detector":

$$D_{SRBBH}(\mathbf{x}) = \|\mathbf{x} - \mathbf{A}_b\,\hat{\boldsymbol{\varrho}}\|_2 - \|\mathbf{x} - \mathbf{A}\,\hat{\boldsymbol{\gamma}}\|_2 \ , \tag{3.3}$$

If $D_{SRBBH}(\mathbf{x}) > \eta$ with η being a prescribed threshold value, then $\mathbf{x}$ is declared as target; otherwise, $\mathbf{x}$ will be labeled as background.

3.1.5 System Overview of the proposed work

In the usual adaptive method for the $\mathbf{A}_b$ construction in section 3.1.3, the dimension of IWR is very important and has a strong impact on the target detection performance since it aims to enclose the targets of interest to be detected. It should be set larger than or equal to the size of all the desired targets of interest in the corresponding HSI, so as to exclude the target pixels from erroneously appearing in the background dictionary $\mathbf{A}_b$. However, information about the target size in the image is usually not at our disposal. It is also very unwieldy to set this size parameter when the target could be of irregular shape (e.g. searching for lost plane parts of a missing aircraft (see the first column of Figure 3.2)). Another tricky situation is when there are multiple targets in close proximity in the image (e.g. military vehicles in long convoy formation (see the second column of Figure 3.2)).

Figure 3.2: From left to right: lost plane parts of a missing aircraft, military vehicles in long convoy.

In this regard, we handle the aforementioned challenges in constructing $\mathbf{A}_b$ by providing a method which is capable of removing the targets from the background, and hence, avoiding the use of an IWR to construct $\mathbf{A}_b$ as well as dealing with a larger range of target size, shape, number, and placement in the image. To do this, we reuse the same target and background separation model that has been developed in Chapter 2 (which is based on the same modification of RPCA), but now, the low-rank component is being the object of interest.

More precisely, from the proposed modified RPCA model that has been discussed in the previous Chapter 2, we aim to use the background HSI $\mathbf{L}$ (since it is pure from the target pixels) for a more accurate construction of $\mathbf{A}_b$, following which various dictionary-based methods can be used to carry out a more elaborate binary hypothesis test (in this dissertation, we make use of the SRBBH detector which was described in section 3.1.4). Via the background HSI $\mathbf{L}$, the locally adaptive $\mathbf{A}_b$ is constructed as in section 3.1.3 but now without the need of using an IWR, and also avoiding contamination by the target pixels.

3.2 Let us improve the usual construction method of $\mathbf{A}_b$

From our proposed modification of RPCA in the previous Chapter 2, the low-rank component $\mathbf{L}$ is now being the object of interest. Hence, we use the background HSI $\mathbf{L}$ for a more accurate construction of $\mathbf{A}_b$, following which various dictionary-based-methods can be used to carry out a more elaborate binary hypothesis test. More precisely, via the background HSI $\mathbf{L}$, a locally adaptive $\mathbf{A}_b$ can be constructed without the need of using an IWR, and also avoiding contamination by the target pixels.

3.2.1 Alleviating the target contamination in $\mathbf{A}_b$

For each test pixel in the original HSI, we create a concentric window of size $m \times m$ on the background HSI $\mathbf{L}$, and all the pixels within the window (except the center pixel) will each contribute to one column in $\mathbf{A}_b$ (see Figure 3.3). Note that this concentric window amounts to an OWR of size $m \times m$ with IWR of size 1×1.

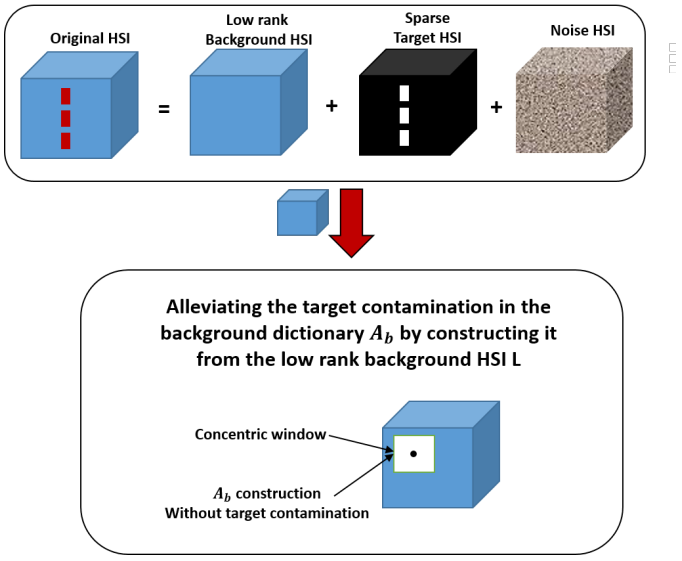


Figure 3.3: Addressing the background dictionary $\mathbf{A}_b$ construction challenge

3.2.2 Our target detection strategy in this chapter

We make use of the SRBBH detector [163] (see equation (3.3) in section 3.1.4 [1]), but with the background dictionary $\mathbf{A}_b$ constructed in the preceding manner (in section 3.2.1).

It is noteworthy to mention that for this scheme to work, and in contrast to the target detection strategy in Chapter 2 (see section 2.3.3 of Chapter 2), we do not need now a clean separation (by clean separation, we mean that all targets are present in $(\mathbf{A}_t\mathbf{C})^T$ with no false alarms); specifically, we require the entire target fraction to be separated from the background and deposited in the target image, but some of the background objects can also be deposited in the target image. As long as enough signatures of these background objects remain in the background HSI $\mathbf{L}$, the $\mathbf{A}_b$ constructed will be adequately representative of the background.

It is important to mention that the edges of the HSI are not processed and so the images are trimmed depending on the window size. As a result, we shall call each of the trimmed image as "the region tested". In fact, by taking a large concentric window, a lot of pixels in the image (near the edges) will not be tested. One can imagine how this can become problematic if these excluded pixels from testing contain some or all the targets of interest. In this regard, after removing the targets from the background by our problem (2.11) in Chapter 2, a small concentric window will be sufficient to construct an accurate background dictionary $\mathbf{A}_b$, and hence, almost the entire image will be tested.

Note also that we could have constructed $\mathbf{A}_b$ directly from all the pixels in the low-rank background HSI $\mathbf{L}$ (except the pixel that corresponds to the test pixel in the original HSI) without the use of any sliding concentric window. This has the advantage of testing the entire image for the detection (that is, the region tested = the original image). However, we choose not to do this as this would result in a substantially larger $\mathbf{A}_b$ and therefore a much increased computational cost in solving sub-problems (3.1a)(3.1b) for the SRBBH detector.

Summary

To handle the $\mathbf{A}_b$ contamination by target pixels and which potentially affect the target detection performance for the SRBBH detector, we aimed to reuse the same proposed modification of RPCA in Chapter 2, and the low-rank component (the low-rank background HSI $\mathbf{L}$) is being the object of interest. More precisely, we aim to contruct $\mathbf{A}_b$ from the low-rank background HSI $\mathbf{L}$. That is, for each test pixel in the given HSI, $\mathbf{A}_b$ is constructed from the low-rank background HSI using a concentric window (an OWR with IWR of size 1×1), and all the pixels within the window

[1] We recall that the reason why we choose the SRBBH detector instead of [33] is because it combines the idea of binary hypothesis and sparse representation, obtaining a more complete and realistic model than [33].

(except the center pixel) will each contribute to one column in $\mathbf{A}_b$.

Do research. Feed your talent. Research not only wins the war on cliche, it's the key to victory over fear and it's cousin, depression.

— Robert McKee

4

Application to Hyperspectral Target Detection for Chapter 2 and Chapter 3

Contents

► *Synopsis* This chapter provides both synthetic as well as real experiments to gauge the effectiveness of our proposed target detection strategies in Chapter 2 (see section 2.3.3 in Chapter 2) and Chapter 3 (see section 3.2.2 in Chapter 3). It is noteworthy to mention that the hyperspectral image datasets in Figure 2.5 (that contains 10 rocks targets), Figure 2.6 (containing targets such as vehicles and bare soil) and Figure 2.7 (that contains one vehicle target) are excluded from our experiments due to the fact that there is no available samples for the targets of interest in the online spectral libraries, and hence, it was impossible to construct the target dictionary $\mathbf{A}_t$.
Recall that the detection strategy in Chapter 2 is outlined in section 2.3.3 of Chapter 2; and the detection strategy in Chapter 3 is outlined in section 3.2.2 in Chapter 3.
This chapter is split into five parts.

1. **The first part 4.1** "*The hyperspectral dataset*" describes the hyperspectral dataset used for our synthetic and real experiments. Note that only two small zones of the acquired dataset

are used. More precisely, one small HSI zone of size $101 \times 101 \times 186$ is considered for the synthetic experiments; whereas another small HSI zone of size $250 \times 291 \times 186$ is used for the real experiments.

2. **The second part 4.2** "*General discussion about τ and λ*" provides a brief discussion on the selection of parameters τ and λ for both detection strategies in Chapter 2 and Chapter 3. More precisely, the parameters will be set manually depending on the HSI used and the target to detect. An automatic selection of the parameters is not yet developed, but should be in the future.

3. **The third part 4.3** "*Synthetic Experiments*" presents the synthetic experiments for both detection strategies in Chapter 2 and Chapter 3. The experiments corroborate our claim that targets with low fill-fraction and targets in convoy formation can be handled.

4. **The forth part 4.4** "*Real Experiments*" presents the real experiments for both detection strategies in Chapter 2 and Chapter 3. The experiments demonstrate the effectiveness of our modified version of RPCA where the target of interest is detected with very little false alarms for both detection strategies.

5. **The fifth part 4.5** "*Summary of the whole Part II and some future directions*" outlines a brief summary of the proposed detection strategies in both Chapter 2 and Chapter 3. It also provides some directions for future work.

4.1 The hyperspectral dataset

The HSI that is used in our experiments is the Cuprite mining district area which is well understood mineralogically [133, 134]. It contains well exposed zones of advanced argillic alteration, consisting principally of kiolinite, alunite, and hydrothermal silica. The Cuprite HSI was acquired by the Airborne Visible / Infrared Imaging Spectrometer (AVIRIS) in 23 June 1995 at local noon, and under high visibility conditions by an NASAER-2 aircraft flying at an altitude of 20 km. The scene we consider is a concatenation of two sectors labeled as "f970619t01p02_r02_sc03.a.rfl" and "f970619t01p02_r02_sc04.a.rfl" in the online data [146] (see Figure 4.1). The resulting image is a 1024×614 - pixel and consisting of 224 spectral (color) bands in contiguous (of about 0.01μm) wavelengths ranging exactly from 0.4046 to 2.4573μm. Prior to some analysis of the Cuprite HSI, the spectral bands 1-4, 104-113, 148-167, and 221-224 are removed due to the water absorption and low SNR in those bands. As a result, a total of 186 bands are used in our experiments. For details about the type of minerals (and their locations) present in this image area, please refer to Figure 5a in [134].

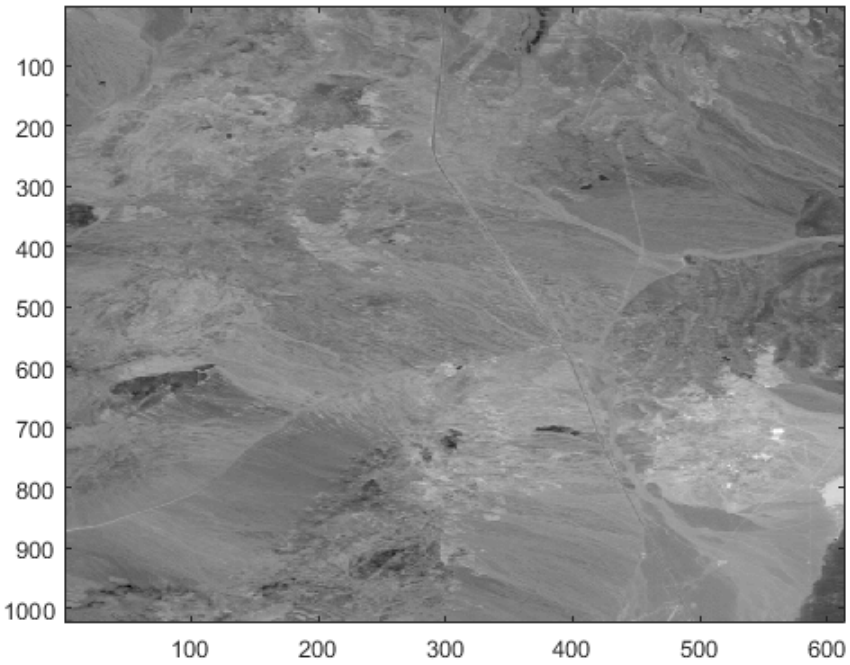


Figure 4.1: The hyperspectral dataset: a concatenation of "f970619t01p02_r02_sc03.a.rfl" and "f970619t01p02_r02_sc04.a.rfl".

4.2 General discussion about τ and λ

In this section, we discuss the main difficulties that face our problem (2.11) in accurately choosing the values for τ and λ bor both detection strategies in Chapter 2 and Chapter 3.
It is important to note that the main limitation of our work is the lack of a method that can automatically selects the parameters τ and λ. Developing such a method is undoubtedly a very interesting step to achieve. Unfortunately, this still seems a very challenging task to us, but will be of our main interest in the future.

For instance, τ and λ are set manually (to achieve the best target detection performance) for both detection strategies in Chapter 2 and Chapter 3. However, this manual selection depends on the HSI used, the spatial and spectral dimensions of the given HSI, on the targets present, and even on how accurate is the target dictionary $\mathbf{A}_t$. All these challenges strongly encourage us to alleviate the manual selection problem of τ and λ in the future.

How can we play with τ and λ ?

We have found that the smarter way to set τ and λ would be to decide on the ratio of τ and λ, respectively for both detection strategies in Chapter 2 and Chapter 3.
In fact, the different requirements imposed by the two detection strategies in both Chapter 2 and

Chapter 3 that can lead to our particular choice setting of the τ to λ ratio also dictate how we should set the relative values of the weights between the first two terms and the third term in (2.11):

1. A lower penalty associated with the third term (that is, by raising the absolute levels of τ and λ) would tolerate more deviation and thus encourage more noise or image clutters (by image clutters we mean the small heterogeneous objects and specular highlights) to be absorbed by this term. This is particularly important for the detection strategy in Chapter 2 when there are a lot of image clutters that do not exactly conform to a low-rank background model: since these clutters do not satisfy the low-rank property, they have a propensity to show up in the second term if we do not sufficiently lower the penalty for the third term, and thus, contribute to a lot of false alarms for the detection strategy in Chapter 2.

2. On the other hand, such a low-penalty setting for the third term may not be a good idea for the detection strategy in Chapter 3 as the third term absorbs too much of the image clutters that actually form the background, causing the background dictionary $\mathbf{A}_b$ so constructed to lose representative power.

Parameters settings

For the detection strategy in Chapter 2, we found that the ratio of τ to λ must be equal to $\frac{5}{2}$ in both synthetic and real experiments. For the detection strategy in Chapter 3, we found that the ratios of τ to λ should be high to make sure that all of the targets are removed to the target image. We set this ratio to approximately 6 for the synthetic experiments and 10 for the real experiments. The ratio for the latter case must be higher because for the real experiments, we do not really have a comprehensive enough target dictionary to represent the target well and thus we need extra incentive for the target fractions to go to the target image.

In sum, for the detection strategy in Chapter 2, we set τ and λ at 0.05 and 0.02 in the synthetic experiments, whereas at 0.5 and 0.2 in the real experiments. For the detection strategy in Chapter 3, we set τ and λ at 0.8 and 0.133 in the synthetic experiments, whereas at 3 and 0.3 in the real experiments.

4.3 Synthetic Experiments

The experiments are done on a 101×101 zone (pixels in rows 389 to 489 and columns 379 to 479) from the acquired Cuprite scene (see Figure 4.2). We incorporate in this zone, 7 target blocks (each of size 6×3) with $\alpha \in [0.01, 1]$ (all have the same α), placed in long convoy formation all formed by the same synthetic (perfect) target $\mathbf{t}$ consisting of a sulfate mineral type known as "Jarosite". We make sure by referring to Figure 5a in [134] that the small zone we consider does not already contain any Jarosite patch. The target $\mathbf{t}$ that we created actually consists of the mean of the first six Jarosite mineral samples taken from the online United States Geological Survey (USGS - Reston)

Spectral Library [38] (see Figure 4.3). The target $\mathbf{t}$ replaces a fraction $\alpha \in [0.01, 1]$ from the background; specifically, the following values of α are considered: 0.01, 0.02, 0.05, 0.1, 0.3, 0.5, 0.8, and 1. As for the target dictionary $\mathbf{A}_t$, it is constructed from the six acquired Jarosite samples [1].

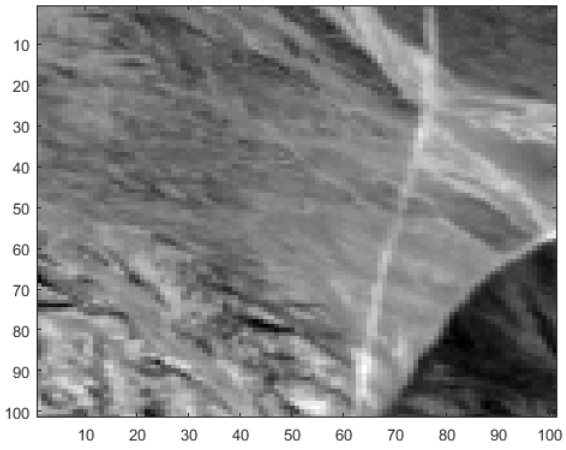


Figure 4.2: The HSI zone used for the synthetic experiments
.

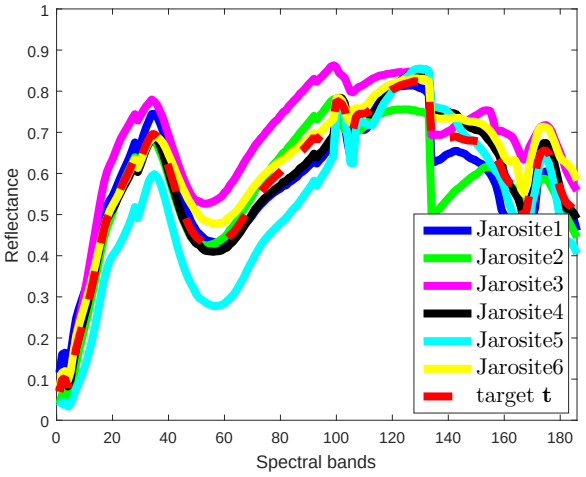


Figure 4.3: Plot of the six Jarosite samples taken from the online USGS Spectral library (which will constitue the target dicitonary $\mathbf{A}_t$), and the target of interest $\mathbf{t}$ consisting of the mean of the six Jarosite samples.

[1] Note that both the HSI and the Jarosite target samples are normalized to values between 0 and 1.

4.3.1 Synthetic experiments for the detection strategy in Chapter 2

Figure 4.4 to 4.11 depict the detection results of $(\mathbf{A}_t\ \mathbf{C})^T$ for different values of α. The plots correspond to the mean power in dB over the 186 spectral bands. As can be seen, the detection strategy in Chapter 2 (that is, our novel target detector) detects all the targets with little false alarms until $\alpha \leq 0.1$ when a lot of false alarms appear.

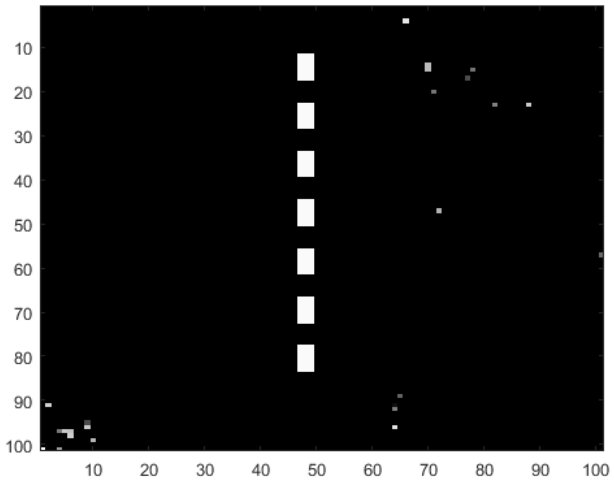


Figure 4.4: Visual detection of $(\mathbf{A}_t\ \mathbf{C})^T$ for the 7 target blocks for $\alpha = 1$

.

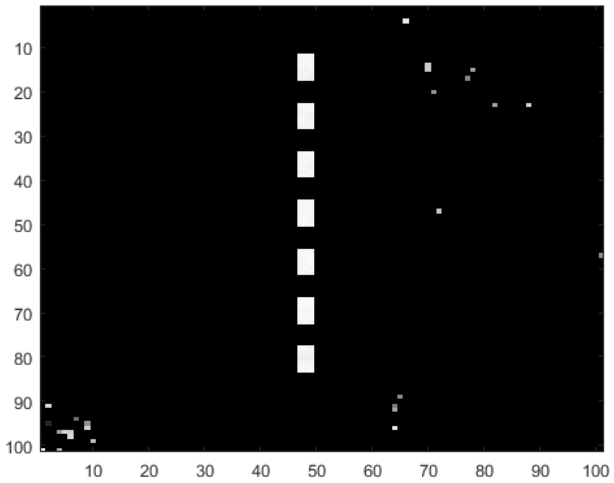


Figure 4.5: Visual detection of $(\mathbf{A}_t\ \mathbf{C})^T$ for the 7 target blocks for $\alpha = 0.8$

.

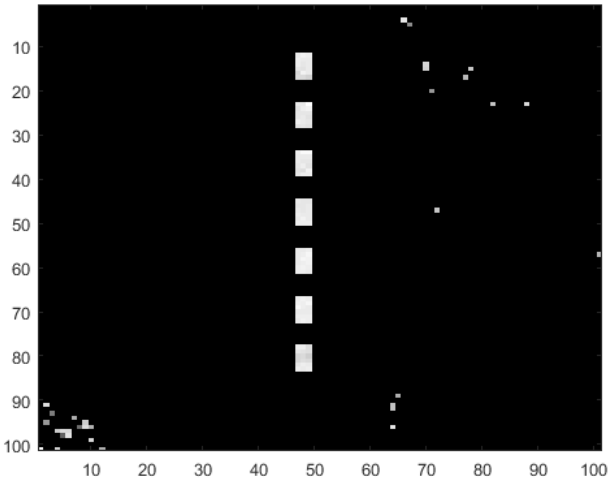


Figure 4.6: Visual detection of $(\mathbf{A}_t\,\mathbf{C})^T$ for the 7 target blocks for $\alpha = 0.5$

.

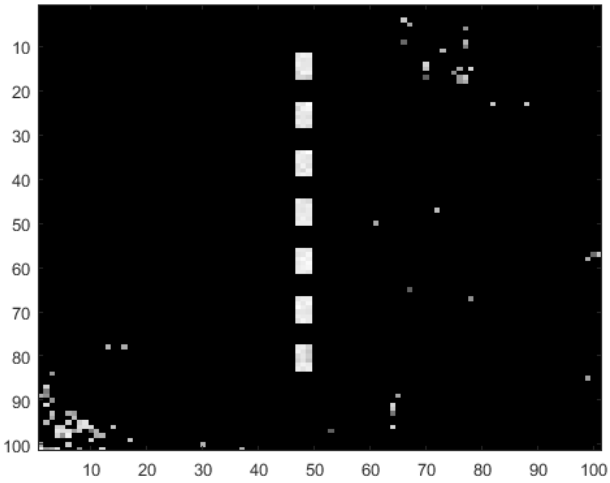


Figure 4.7: Visual detection of $(\mathbf{A}_t\,\mathbf{C})^T$ for the 7 target blocks for $\alpha = 0.3$

.

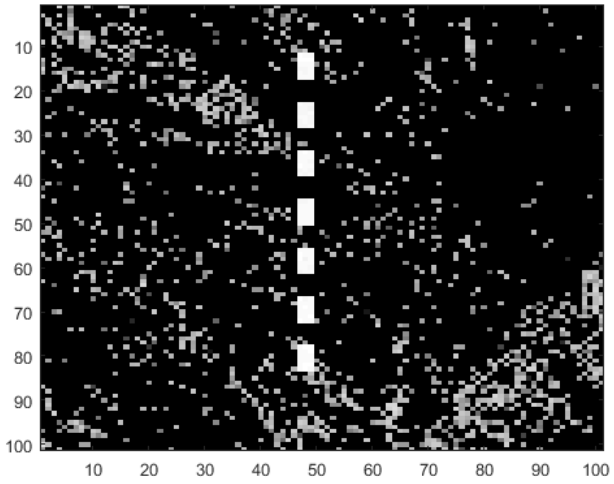


Figure 4.8: Visual detection of $(\mathbf{A}_t\,\mathbf{C})^T$ for the 7 target blocks for $\alpha = 0.1$

.

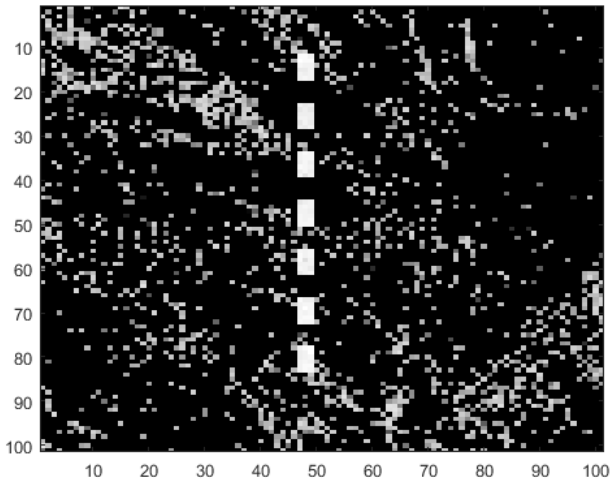


Figure 4.9: Visual detection of $(\mathbf{A}_t\,\mathbf{C})^T$ for the 7 target blocks for $\alpha = 0.05$

.

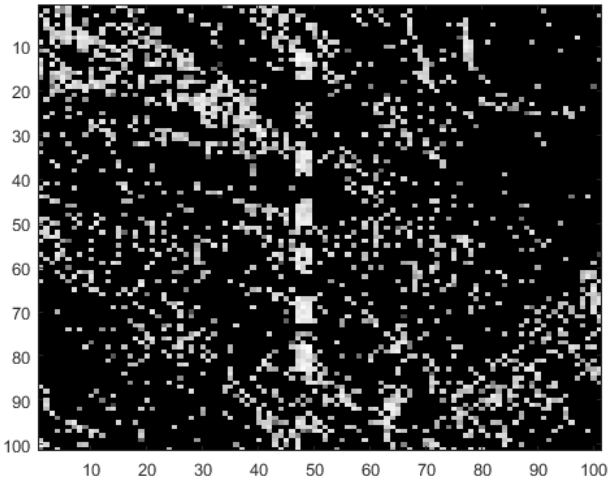


Figure 4.10: Visual detection of $(\mathbf{A}_t\ \mathbf{C})^T$ for the 7 target blocks for $\alpha = 0.02$

.

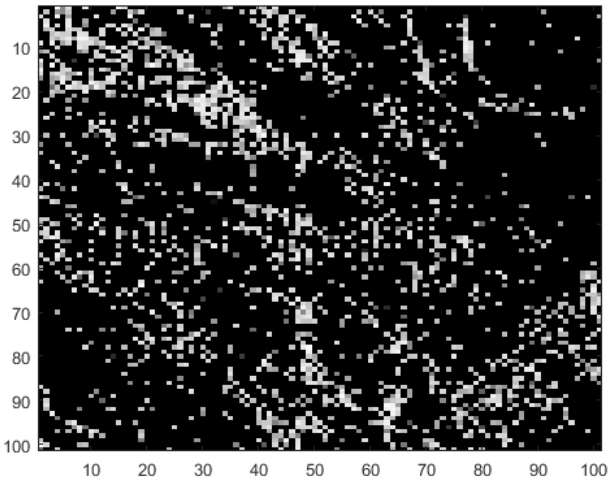


Figure 4.11: Visual detection of $(\mathbf{A}_t\ \mathbf{C})^T$ for the 7 target blocks for $\alpha = 0.01$

.

4.3.2 Synthetic experiments for the detection strategy in Chapter 3

We first provide in Figure 4.12 a visual evaluation of the separation of the 7 target blocks for low $\alpha = 0.1$. We can observe that our problem (2.11) successfully discriminates these perceptually invisible targets from the background in $\mathbf{D}$ and separate them. The 7 darker blocks that appear in $\mathbf{L}$ correspond to the dimmer fraction of the background that remains after the targets have been removed at the corresponding spatial locations. Having qualitatively inspect the separation, we now aim to quantitatively evaluate the target detection performances of the SRBBH detector [163] when $\mathbf{A}_b$ is constructed using a small concentric window of size 5×5. That is, $\mathbf{A}_b \in \mathbb{R}^{p \times 24}$ (after excluding the center pixel) and the region tested consists of an image of size 97×97.

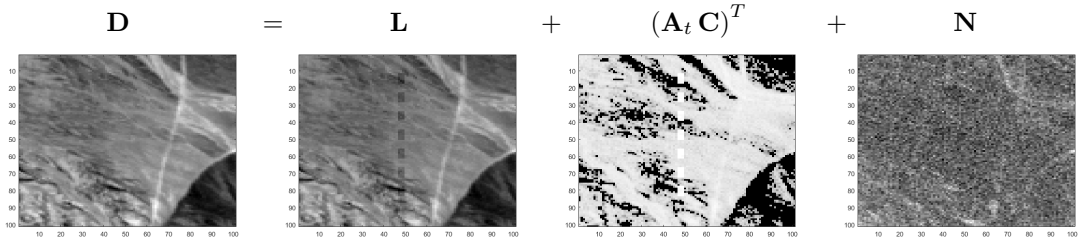


Figure 4.12: Visual separation of the 7 target blocks for $\alpha = 0.1$: We exhibit the mean power in dB over the 186 bands. Columns from left to right: the original HSI containing the 7 target blocks ($\alpha = 0.1$), low-rank background HSI $\mathbf{L}$, target HSI $(\mathbf{A}_t\,\mathbf{C})^T$, noise HSI.

In what follows, we shall use $\mathbf{D}_b$ to represent the HSI that does not contain the 7 target blocks (that is, the pure background image), and $\mathbf{D}$ to represent the HSI after incorporating the 7 target blocks (that is, it contains the targets) for $\alpha \in [0.01, 1]$.
We now consider three scenarios to form the columns in $\mathbf{A}_b$:

1. For each test pixel in $\mathbf{D}$, we create the concentric window on $\mathbf{D}_b$. This represents the ideal case since $\mathbf{A}_b$ is free from the targets.

2. For each test pixel in $\mathbf{D}$, we create the concentric window on $\mathbf{D}$.

3. For each test pixel in $\mathbf{D}$, we create the concentric window on the low-rank background HSI $\mathbf{L}$.

The target detection performances are evaluated quantitatively specifically by the Receiver Operating Characteristics (ROC) curves which describe the probability of detection (P_d) against the probability of false alarm (P_{fa}) as we vary the threshold η between the minimal and maximal values of each detector output. A good detector presents high P_d values at low P_{fa}, i.e., the curve is closer to the upper left corner. More particularly, the P_d can be determined as the ratio of the number of the target pixels determined as target (the detector output at each pixel on the target region exceeds the threshold value) and the total number of true target pixels. Whereas the P_{fa} can be calculated by the ratio of the number of false alarms (the detector output at each pixel on the background region that is outside the target region exceeds the threshold value) and the total number of pixels in the region tested.

Figure 4.13 to 4.20 depict the quantitative detection results. Clearly, increasing α should render the target detection less challenging, and thus, better detection results are being expected. However, this fact can not always be the case for the SRBBH detector when $\mathbf{A}_b$ is constructed from $\mathbf{D}$ (blue curves): It is true that the increase in α helps to improve the detection, but at the same time leads to more target contamination in $\mathbf{A}_b$ which in turn suppresses the detection improvement that ought be had. That is why the SRBBH detector (blue curves) does not reap

full benefits from the increase in α, and thus, presents poor detection results even for large α values.

By constructing $\mathbf{A}_b$ from $\mathbf{L}$ (which only contains the background with the targets are removed after applying problem (2.11)), the SRBBH detector (dashed green curves) can better detect the targets especially for $\alpha \geq 0.1$, and has competitive detection results when compared to the ideal case when $\mathbf{A}_b$ is constructed from $\mathbf{D}_b$. The detection performances start to deteriorate progressively for very small α values and degenerate to the SRBBH level (blue curves) for $\alpha \leq 0.02$.

To sum up, the obtained target detection results corroborate our claim that we can handle targets with low fill-fraction and in convoy formation.

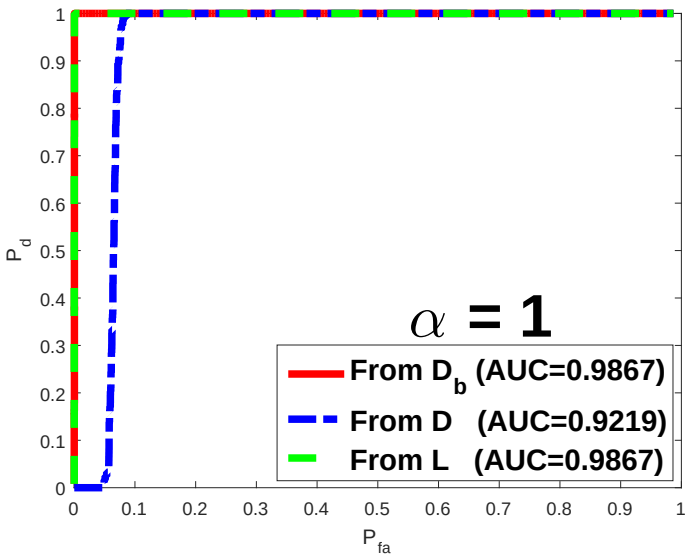


Figure 4.13: ROC curves (with their Area Under Curves (AUC) values) for $\alpha = 1$ of the SRBBH detector when $\mathbf{A}_b$ is constructed from $\mathbf{D}_b$, $\mathbf{D}$ and $\mathbf{L}$.

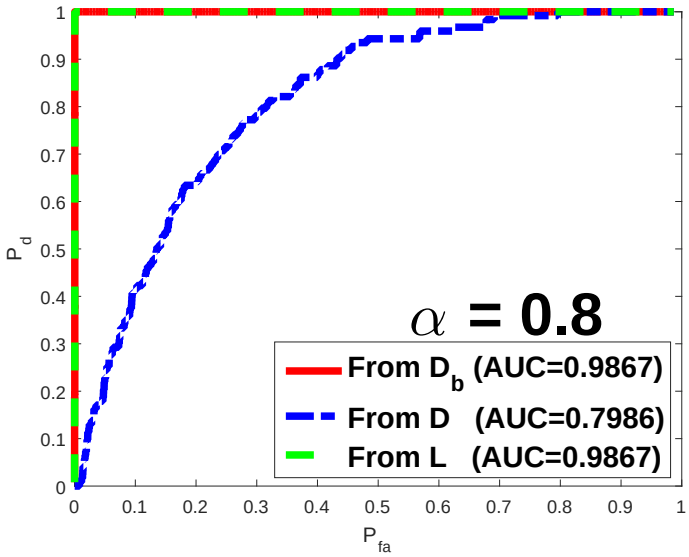


Figure 4.14: ROC curves (with their Area Under Curves (AUC) values) for $\alpha = 0.8$ of the SRBBH detector when $\mathbf{A}_b$ is constructed from $\mathbf{D}_b$, $\mathbf{D}$ and $\mathbf{L}$.

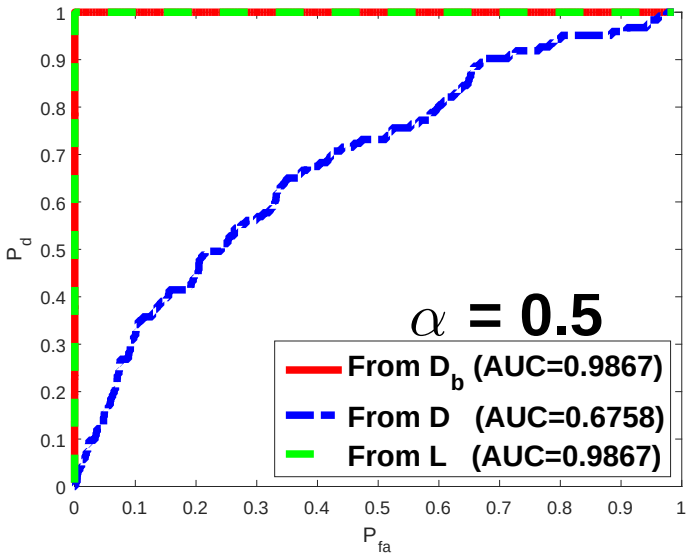


Figure 4.15: ROC curves (with their Area Under Curves (AUC) values) for $\alpha = 0.5$ of the SRBBH detector when $\mathbf{A}_b$ is constructed from $\mathbf{D}_b$, $\mathbf{D}$ and $\mathbf{L}$.

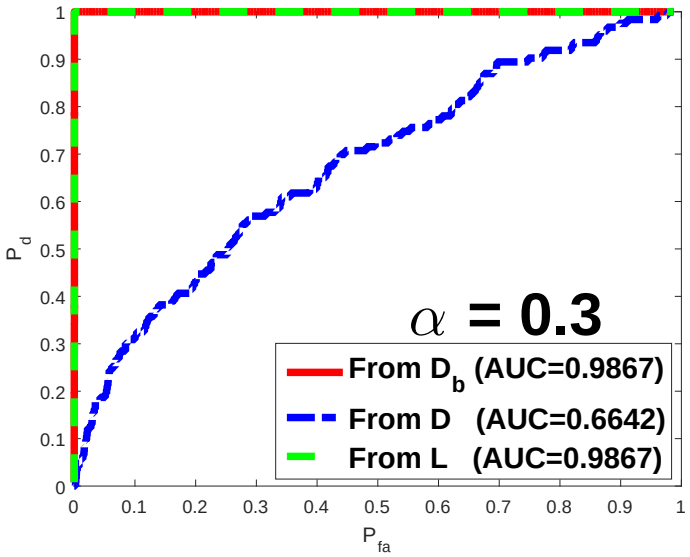


Figure 4.16: ROC curves (with their Area Under Curves (AUC) values) for $\alpha = 0.3$ of the SRBBH detector when $\mathbf{A}_b$ is constructed from $\mathbf{D}_b$, $\mathbf{D}$ and $\mathbf{L}$.

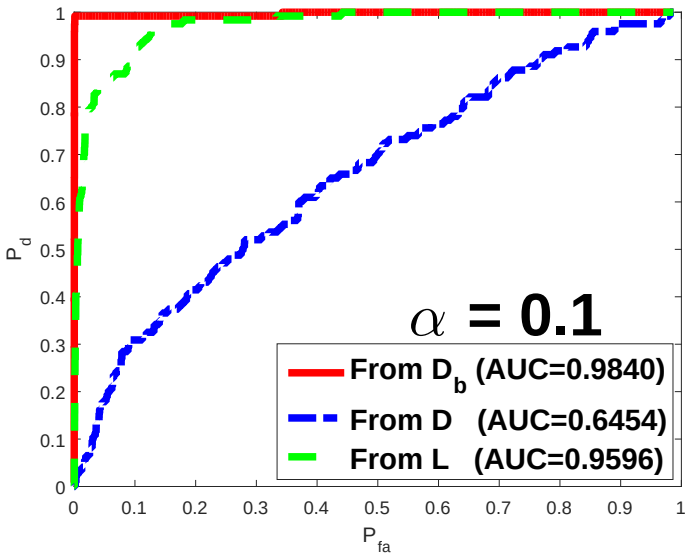


Figure 4.17: ROC curves (with their Area Under Curves (AUC) values) for $\alpha = 0.1$ of the SRBBH detector when $\mathbf{A}_b$ is constructed from $\mathbf{D}_b$, $\mathbf{D}$ and $\mathbf{L}$.

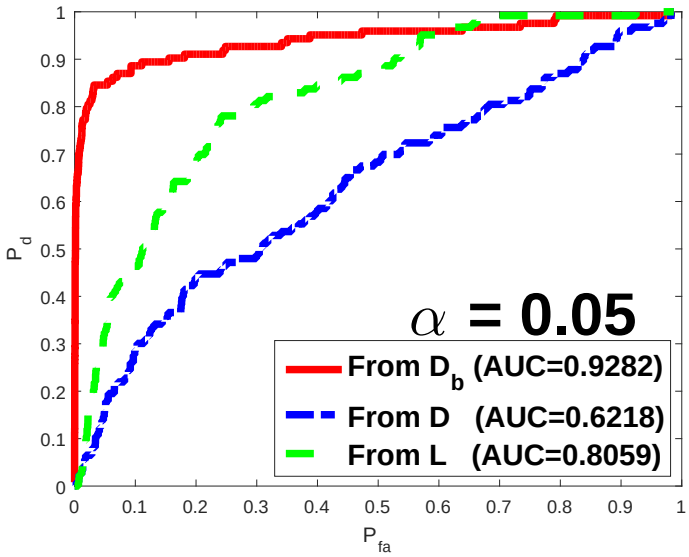


Figure 4.18: ROC curves (with their Area Under Curves (AUC) values) for $\alpha = 0.05$ of the SRBBH detector when $\mathbf{A}_b$ is constructed from $\mathbf{D}_b$, $\mathbf{D}$ and $\mathbf{L}$.

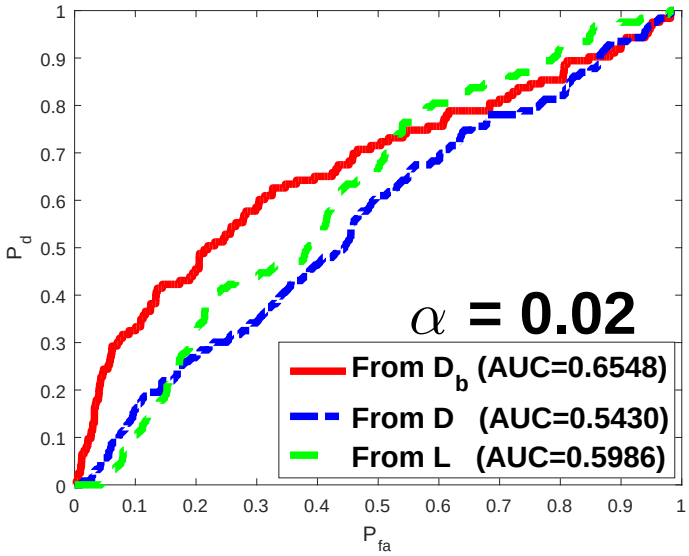


Figure 4.19: ROC curves (with their Area Under Curves (AUC) values) for $\alpha = 0.02$ of the SRBBH detector when $\mathbf{A}_b$ is constructed from $\mathbf{D}_b$, $\mathbf{D}$ and $\mathbf{L}$.

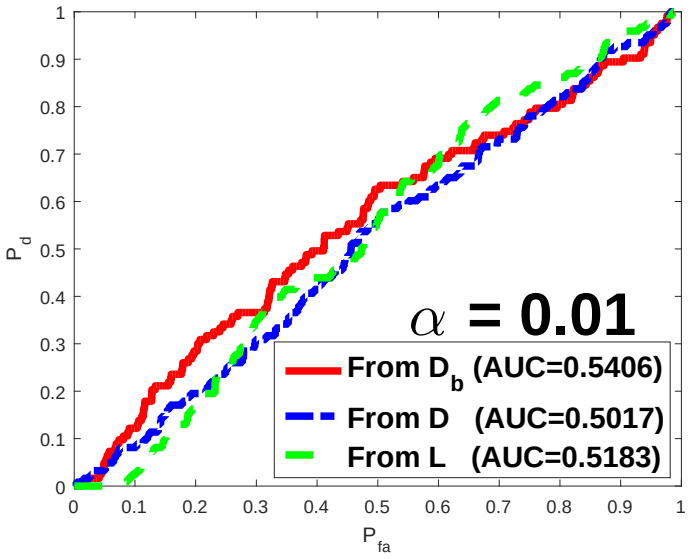


Figure 4.20: ROC curves (with their Area Under Curves (AUC) values) for $\alpha = 0.01$ of the SRBBH detector when $\mathbf{A}_b$ is constructed from $\mathbf{D}_b$, $\mathbf{D}$ and $\mathbf{L}$.

4.4 Real Experiments

This part evaluates qualitatively the target detection performances of the SRBBH detector, using a concentric window of size 5×5 on a region of size 250×291 pixels taken from the acquired Cuprite HSI. We consider this zone specifically to detect the Tectosilicate mineral type target pixels known as Buddingtonite. The mean power in dB over the 186 spectral bands of this zone and the Buddingtonite GroundTruth are shown in Figure 4.21.

There are three Buddingtonite samples available in the online Advanced Spaceborne Thermal Emission and Reflection (ASTER) spectral library - Version 2.0 [5] and they will form our target dictionary $\mathbf{A}_t$. The ASTER Spectral library was released on December 2008 to include data from the USGS Spectral Library, the Johns Hopkins University Spectral Library, and the Jet Propulsion Laboratory Spectral Library.

Both the HSI and the Buddingtonite target samples are normalized to values between 0 and 1. Figure 4.22 depicts the Buddingtonite target samples taken from the online ASTER Spectral library.

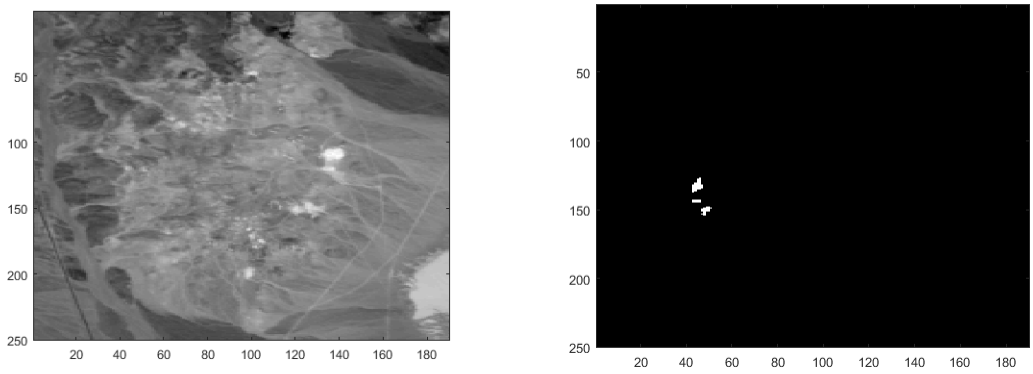


Figure 4.21: Columns from left to right: The HSI dataset used for the real experiments, the GroundTruth of the Buddingtonite target pixels.

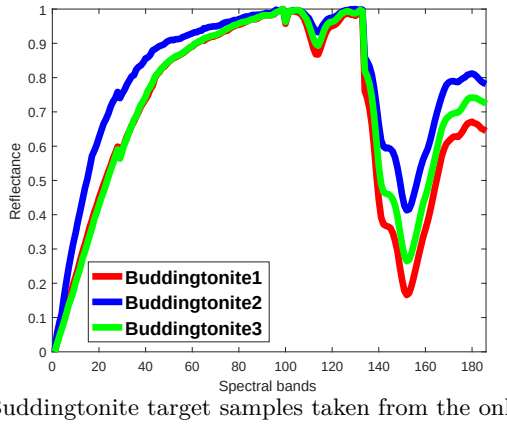


Figure 4.22: Plot of the Buddingtonite target samples taken from the online ASTER Spectral library.

4.4.1 Real experiments for the detection strategy in Chapter 2

Figure 4.23 depicts the detection of the Buddingtonite targets in $(\mathbf{A}_t\,\mathbf{C})^T$. The Buddingtonite targets are detected with very little false alarms.

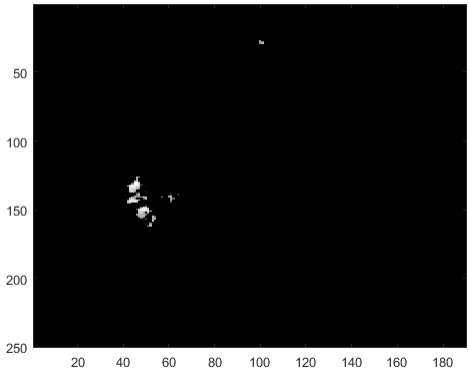


Figure 4.23: The detection in $(\mathbf{A}_t\,\mathbf{C})^T$ (we exhibit the mean power in dB over the 186 bands) for the detection strategy in Chapter 2.

4.4.2 Experiments for the detection strategy in chapter 3

As a consequence of the decomposition depicted in Figure 4.24, the subspace overlap problem illustrated in Figure 2.8 is now much relieved, as can be seen from Figure 4.24. Figure 4.25

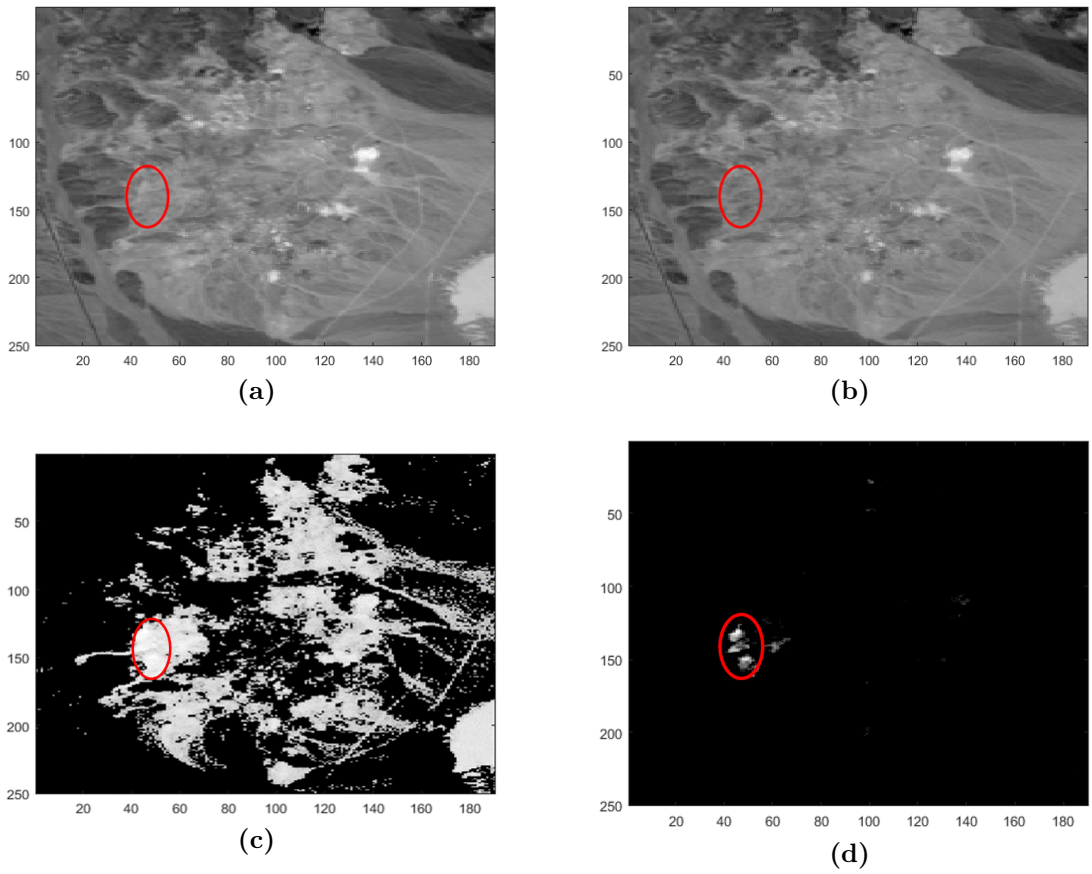


Figure 4.24: Visual separation (we exhibit the mean power in dB over the 186 bands) of the Buddingtonite targets using the target dictionary $\mathbf{A}_t$ constructed from the ASTER Spectral library. **(a)**: original HSI $\mathbf{D}$, **(b)**: low-rank background HSI $\mathbf{L}$, **(c)**: sparse target HSI $(\mathbf{A}_t\,\mathbf{C})^T$, **(d)**: sparse target HSI $(\mathbf{A}_t\,\mathbf{C})^T$ after some thresholding.

and 4.26 evaluate qualitatively the SRBBH detection results when $\mathbf{A}_b$ is constructed from $\mathbf{D}$ and $\mathbf{L}$, respectively, using a concentric window of size 5×5. The effectiveness of problem (2.11) in improving the target detection is evident.

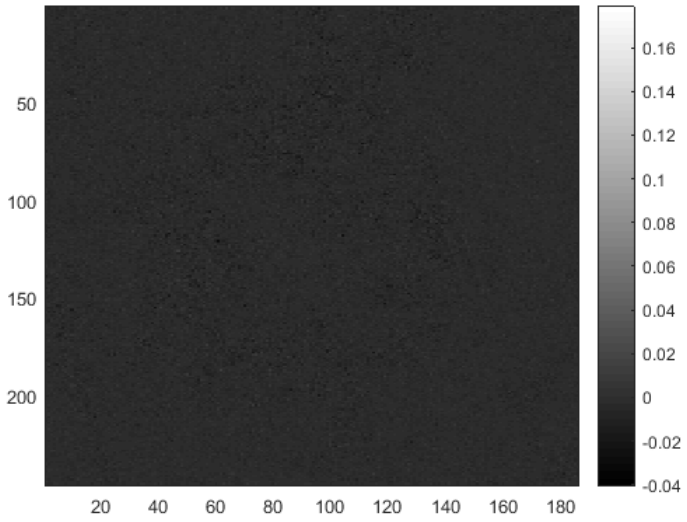


Figure 4.25: The SRBBH detector (2D Visualization of the Buddingtonite target pixels detection) when $\mathbf{A}_b$ is constructed from $\mathbf{D}$.

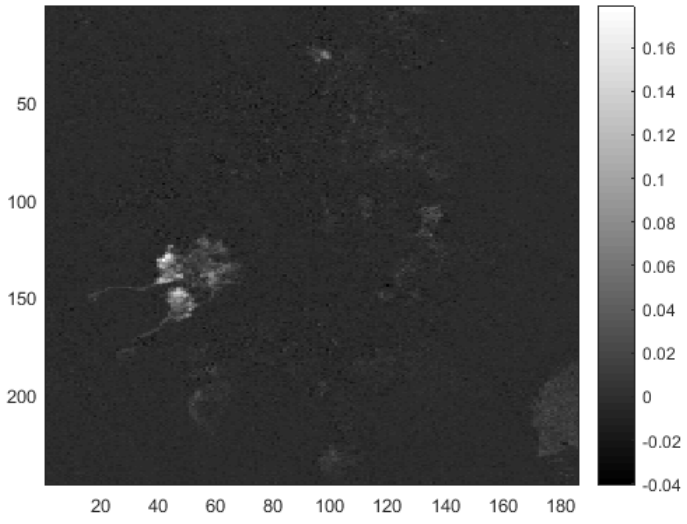


Figure 4.26: The SRBBH detector (2D Visualization of the Buddingtonite target pixels detection) when $\mathbf{A}_b$ is constructed from $\mathbf{L}$.

4.5 Summary of the whole Part II and some future directions

4.5.1 Summary

A method based on a modification of RPCA is proposed to separate known targets of interest from the background in hyperspectral imagery. More precisely, we regard the given HSI as being made up of the sum of a low-rank background HSI $\mathbf{L}$ and a sparse target HSI $\mathbf{E}$ that should contain the

targets of interest. Based on a pre-learned target dicitonary $\mathbf{A}_t$ that is constructed from some online spectral libraries, we customize the general RPCA by factorizing the sparse component $\mathbf{E}$ into the product of $\mathbf{A}_t$ and a sparse activation matrix $\mathbf{C}$. This modification was essential to disambiguate the true targets from other small heterogeneous and high contrast regions. Following the decomposition:

1. The first outlined target detection strategy in chapter 2 was to directly use the component $(\mathbf{A}_t\mathbf{C})^T$ as a detector. Only the signals that reside in the target subspace specified by $\mathbf{A}_t$ are deposited at the non-zero entries of $(\mathbf{A}_t\mathbf{C})^T$.

2. The second outlined target detection strategy in Chapter 3 addresses the background dictionary contamination problem suffered by dictionary-based methods such as SRBBH. To do this, the low-rank background HSI $\mathbf{L}$ was exploited to construct $\mathbf{A}_b$. More precisely, for each test pixel in the original HSI, the $\mathbf{A}_b$ is constructed from $\mathbf{L}$ using a small concentric window, and all the pixels within the window (except the center pixel) will each contribute to one column in $\mathbf{A}_b$.

Both detection strategies in Chapter 2 and Chapter 3 are evaluated independently to each other on both synthetic and real experiments, and the results of which demonstrate their effectiveness for hyperspectral target detection. In particular, they can deal with targets of any shapes or targets that occur in close proximity, and are resilient to most values of fill-fractions unless they are too small.

4.5.2 Some directions for future work

As for future enhancements, a likely first step would be to evaluate the proposed modified RPCA model on more real datasets. Other promising avenues for further research include:

1. We have mentioned in part 4.2 of this chapter that the selection of parameters τ and λ strongly depends on the HSI used, on the spatial and spectral dimension of the given HSI, on the target of interest to detect, on the location of the target in the image scene, and on the target dictionary $\mathbf{A}_t$. This encourages us to work hard in the future to develop such an automatic selection method for the parameters (i.e. develop a method that takes the aforementioned causes as input).

2. Obviouly, we can observe that the τ to λ ratios as well as the settings of τ and λ (see part 4.2 of this chapter) for the detection strategy of Chapter 2 are not similar to those for the detection strategy of Chapter 3. We highly expect that if one could use directly the $l_{2,0}$ norm (that is, without surrogating it towards the convex $l_{2,1}$ norm), both detection strategies might have the same parameters settings.

3. Interestingly, what we have not mentioned in part 4.2 of this chapter is that the selection of parameters τ and λ depends on the target fill-fraction α. During this work (omitted here), we did a lot of experiments on the HSI zone used in the synthetic experiments by replacing

a fraction α from the background pixel at location $(34, 50)$ by the target $\mathbf{t}$ corresponding to the mean of the six jarosite target samples. We have observed that if one needs to separate the $\alpha\mathbf{t}$ from $(1-\alpha)\mathbf{b}$ using our problem in (2.11) (that is, we need that $\alpha\mathbf{t}$ to be deposited in the sparse component, and the $(1-\alpha)\mathbf{b}$ in the low-rank component), the selection of λ will not be unique for all α values. More precisely, the higher α is, the more need to decrease λ. This is due to the fact that a higher value of α implies a more important target fraction to separate from the background. But we highly expect that if one could use directly the $l_{2,0}$ norm, a unique value of λ might be chosen for all α values.

In this regard, our future works will mainly focus on the use of other proxies than the $l_{2,1}$ norm (closer to the $l_{2,0}$ norm) which can help to alleviate the $l_{2,1}$ artifact and probably the manual selection problem of τ and λ.

****Some remarks**

In 20 February 2018, we have received an important feedback (by mail) from Prof. Dimitris Manolakis (working in the Massachusetts Institute of Technology (MIT)) concerning our work:

Prof. Manolakis's feedback (by mail) Your approach seems interesting, but I am not sure how the low-rank versus sparse model approach can handle targets versus background discrimination. All material spectra have some kind of structure. Your approach is interesting and it is within the scope of academic research or a PhD thesis.

Our answer was Our approach strongly depends on how is the target dictionary constructed. If the latter does not well represent the target of interest, our approach can fail on discriminating the targets from the background. This was the big challenge in the real application due to the variations on the target spectra (due to the atmospheric conditions, sensor noise, and material composition). There is also a lack of signatures present in the online spectral libraries that represent the target that we want to detect. This renders the construction of the target dictionary very challenging and not well accurate.
In addition, the Buddingtonite in the real application can be considered as an easy "target" to detect or to separate from the background. This is because it does not look like any other mineral with its distinct 2.1μm absorption[2]. However, we have observed that the paradigm in military applications of HSI usually center on finding the target but ignoring all the rest. Sometimes that rest is important especially if the target is well matched to the surroundings. Perhaps our approach should be tested specifically on that challenge in the future.

[2]We got this remark (after a discussion by mail) from Dr. Gregg A. Swayze who is working in the USGS Spectroscopy Lab. We really greatly thank this person for his time in providing us helpful remarks and suggestions specially about the Cuprite data.

Part III

Target Detection Based on Sparse Covariance Matrices

Imagination is more important than knowledge. Knowledge is limited. Imagination encircles the world.

— Albert Einstein

5 General Background

Contents

► *Synopsis* This chapter first gives a brief introduction to the linear regression analysis. Then, it outlines in detail the traditional covariance estimators, including a comparative study of them on some Monte-Carlo simulations for hyperspectral target detection. The obtained results demonstrate the poor behavior of the traditional covariance estimators when the spectral dimension p is large compared to the sample size n. Next, the chapter provides some of the research works that have been developed in the literature and in which sparsity has been imposed on the covariance matrix in order to alleviate the challenges brought by the high covariance dimensionality.

5.1 Introduction to linear regression analysis

Given a sample $\{\mathbf{x}_i\}_{i\in[1,\,n]}$, consider the following regression model:

$$y_i = f_{\boldsymbol{\beta}}(\mathbf{x}_i^T) + e_i\,, \tag{5.1}$$

where y_i is the dependent or response variable, $\mathbf{x}_i = [x_{i,0},\, x_{i,1},\, \cdots,\, x_{i,p}]^T \in \mathbb{R}^{p+1}$ designates the vector of independent or explanatory variables, $\boldsymbol{\beta} = [\beta_0,\, \beta_1,\, \cdots,\, \beta_p]^T \in \mathbb{R}^{p+1}$ is a vector of the unknown regression coefficients, and e_i is the residual error variable. Hereafter, we will consider that $\beta_0 = 0$, so that $\boldsymbol{\beta} \in \mathbb{R}^p$ and $\mathbf{x}_i \in \mathbb{R}^p$. The case of one explanatory variable is called "simple linear regression". Whereas for more than one explanatory variable, the process is called "multiple linear regression".

By assuming a simple linear regression function f, equation (5.1) can be re-written as follows:

$$y_i = x_{i,1}\beta_1 \,+\, x_{i,2}\beta_2 \,+\, \cdots \,+\, x_{i,p}\beta_p \,+\, e_i\,. \tag{5.2}$$

By writing equation (5.2) in vector-matrix form for any $i \in [1,\, n]$, one obtains: $\mathbf{y} = \mathbf{A}\boldsymbol{\beta} + \mathbf{e}$, where $\mathbf{y} \in \mathbb{R}^n$, $\mathbf{A} = [x_{i,t}]_{n\times p}$, $\boldsymbol{\beta} \in \mathbb{R}^p$, and $\mathbf{e} \in \mathbb{R}^n$.

5.1.1 The Ordinary Least Squares (OLS) method

In case when $n > p$, more equations than unknowns are thus present and this leads to an overdetermined linear system of equations. Hence, there is enough data in order to estimate the unknown regression coefficients $\beta_1, \cdots, \beta_p$, and usual methods such as the Ordinary Least Squares (OLS) can be implemented:

$$\hat{\boldsymbol{\beta}}^{OLS} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \frac{1}{2}\sum_{i=1}^{n}\left(y_i - \sum_{t=1}^{p} x_{i,t}\,\beta_t\right)^2 = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \frac{1}{2}\,\|\mathbf{y} - \mathbf{A}\,\boldsymbol{\beta}\|_2^2 = \left(\mathbf{A}^T\,\mathbf{A}\right)^{-1}\,\mathbf{A}^T\,\mathbf{y}\,. \tag{5.3}$$

Hence, the least squares method aims to estimate the regression coefficients by minimizing the squared discrepancies between observed data, on the one hand, and their expected values on the other.

5.1.2 Penalized Least Squares

The whole machinery of least-squares fails or does not work well for the high dimensional data where the $\mathbf{A}^T\mathbf{A}$ matrix may be difficult to be inverted. In this regard, a regularized version of the least squares solution may be preferable.
We consider the penalized least squares method:

$$\underset{\boldsymbol{\beta}}{\operatorname{argmin}}\ \frac{1}{2}\ \|\mathbf{y} - \mathbf{A}\,\boldsymbol{\beta}\|_2^2 + \sum_{t=1}^{p} p_{\varphi}\left(|\beta_t|\right)\,, \tag{5.4}$$

where $p_{\varphi}(.)$ is a penalty function indexed by the regularized parameter φ.

Some commonly used penalty functions (see Figure 5.1):

- l_0 penalty;
- l_2 penalty (Ridge regression);
- l_1 penalty (LASSO regression) [140];
- $SCAD$ penalty [53];
- etc.

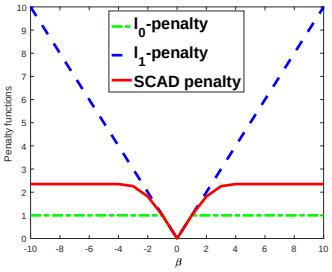


Figure 5.1: Plots of the l_0, l_1 and $SCAD$ penalty functions: we set $\boldsymbol{\beta} = [10 : 10]$, $\varphi = 1$ and $a = 3.7$

5.1.3 Penalized regression using l_2 norm: The ridge regression

As has been mentioned in section 5.1.2, a regularized version of the least squares solution may be preferable. The traditional remedy is the ridge regression (aka Tikhonov regularization) [78], which replaces the residual sum of squares of errors by its penalized version:

$$\underset{\boldsymbol{\beta}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\,\boldsymbol{\beta}\|_2^2 + \varphi \sum_{t=1}^{p} |\beta_t|^2 \,, \tag{5.5}$$

where $\varphi > 0$ is a penalty controlling the length of the vector of regression parameters. The ridge regression has a very nice closed form solution that is easily interpreted, and this can be helpful in practice. That is, $\hat{\boldsymbol{\beta}}^{Ridge} = \left(\mathbf{A}^T\,\mathbf{A} + \varphi\,\mathbf{I}\right)^{-1}\,\mathbf{A}^T\,\mathbf{y}$.
The ridge solution works rather well when p is not too large; however, in general it does not induce sparsity in the model (this is fine when sparsity is not important).

5.1.4 Penalized regression using l_1 norm: The LASSO regression

The Least Absolute Shrinkage and Selection Operator (LASSO) is one of the most popular approaches for selecting significant variables and estimating regression coefficients simultaneously. The main difference between LASSO and ridge is the penalty term they use. The LASSO uses the l_1 penalty which imposes sparsity (this sounds good when sparsity is important) among the regression coefficients, that is, it minimizes the sum of squares of residuals subject to a constraint on the sum of absolute values of the regression coefficients:

$$\underset{\boldsymbol{\beta}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\,\boldsymbol{\beta}\|_2^2 + \varphi \sum_{t=1}^{p} |\beta_t| \,, \tag{5.6}$$

where $\varphi > 0$ is a tuning parameter controlling the sparsity of the model.
One of the prime differences between LASSO and ridge regression is that in ridge regression, as the penalty is increased, all regression coefficients are reduced while still remaining non-zero; whereas in LASSO, increasing the penalty will cause more and more of the regression coefficients to be driven to zero. Thus, LASSO automatically selects more relevant features and discards the others, whereas ridge regression never fully discards any feature.

Unlike the closed-form ridge solution, and due to the nature of the constraint in (5.6), the LASSO solution is non-linear in the responses y_i's. Fundamental to understand and compute the LASSO solution is the *Soft thresholding operator* [48].

5.1.5 Orthogonal Design in Penalized Least Squares

Insight about the nature of the penalization methods can be gleaned from the orthogonal design case. Usually, one assumes that the predictors are orthogonal, i.e., $\mathbf{A}^T\,\mathbf{A} = \mathbf{I}$, and which is of course a restrictive assumption.

Hence, by expanding out the first term in (5.4), we get:

$$\begin{aligned}
&\frac{1}{2}\,\|\mathbf{y}\|_2^2 - \mathbf{y}^T\,\mathbf{A}\,\boldsymbol{\beta} + \frac{1}{2}\,\|\mathbf{A}\boldsymbol{\beta}\|_2^2 \\
&= \frac{1}{2}\,\mathbf{y}^T\,\mathbf{y} - \mathbf{y}^T\,\mathbf{A}\,\boldsymbol{\beta} + \frac{1}{2}\,(\mathbf{A}\,\boldsymbol{\beta})^T\,(\mathbf{A}\,\boldsymbol{\beta}) \\
&= \frac{1}{2}\,\mathbf{y}^T\,\mathbf{y} - \mathbf{y}^T\,\mathbf{A}\,\boldsymbol{\beta} + \frac{1}{2}\,\boldsymbol{\beta}^T\,\left(\mathbf{A}^T\mathbf{A}\right)\,\boldsymbol{\beta} \\
&= \frac{1}{2}\,\mathbf{y}^T\,\mathbf{y} - \mathbf{y}^T\,\mathbf{A}\,\boldsymbol{\beta} + \frac{1}{2}\,\boldsymbol{\beta}^T\,\boldsymbol{\beta}\,.
\end{aligned}$$

By discarding $\mathbf{y}^T\,\mathbf{y}$ since it does not contain any of the variables of interest (that is, $\boldsymbol{\beta}$), problem (5.4) can be reformulated as:

$$\underset{\boldsymbol{\beta}}{\operatorname{argmin}}\ \left(-\mathbf{y}^T\,\mathbf{A}\,\boldsymbol{\beta} + \frac{1}{2}\,\boldsymbol{\beta}^T\,\boldsymbol{\beta}\right) + \sum_{t=1}^{p} p_\varphi\left(|\beta_t|\right)\,. \tag{5.7}$$

As we have previously considered, the columns of $\mathbf{A}$ are orthonormal, so that, $\hat{\boldsymbol{\beta}}^{OLS} = \mathbf{A}^T\,\mathbf{y}$ (since $\mathbf{A}^T\,\mathbf{A} = \mathbf{I}$). Problem (5.7) can thus be re-written as:

$$\underset{\boldsymbol{\beta}}{\operatorname{argmin}}\ \frac{1}{2}\,\left\|\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}^{OLS}\right\|_2^2 + \sum_{t=1}^{p} p_\varphi\left(|\beta_t|\right)\,. \tag{5.8}$$

Now the optimization problem is separable in β_t's. It suffices to consider the univariate Penalized Least Squares problem:

$$\hat{\beta}_t = \underset{\beta_t}{\operatorname{argmin}}\ \frac{1}{2}\left(\beta_t - \hat{\beta}_t^{OLS}\right)^2 + p_\varphi(|\beta_t|)\,. \tag{5.9}$$

According to Antoniadis and Fan in [2], the solution to the minimization problem (5.9) exists and is unique.

5.1.6 LASSO: Orthogonal Design and the Soft thresholding rule

Under the orthogonal design, i.e., $\mathbf{A}^T\,\mathbf{A} = \mathbf{I}$, the LASSO estimation can be greatly simplified. The problem becomes:

$$\hat{\beta}_t = \underset{\beta_t}{\operatorname{argmin}}\ \frac{1}{2}\,\left(\beta_t - \hat{\beta}_t^{OLS}\right)^2 + \varphi\,|\beta_t|\,. \tag{5.10}$$

The Lasso estimator is given by (the Soft thresholding rule):

$$\hat{\beta}_t^{Soft} = \mathcal{S}\left(\hat{\beta}_t^{OLS};\,\varphi\right)\,,$$

where $\mathcal{S}(.;\varphi)$ is the Soft-thresholding operator:

$$\mathcal{S}\left(\hat{\beta}_t^{OLS};\varphi\right) = \operatorname{sgn}\left(\hat{\beta}_t^{OLS}\right)\left(|\hat{\beta}_t^{OLS}| - \varphi\right)_+ .$$

5.1.7 The Smoothly Clipped Absolute Deviation (SCAD) thresholding

In the search for an ideal penalty function, Fan and Li [53] advocate the use of regularization methods leading to estimators having the following desirable properties:

- *Sparsity*: the resulting estimator should be a thresholding rule, that is, the estimated regression coefficients that are lower than a specified threshold φ are set to zero, which establishes thresholding.
- *Unbiasedness*: The resulting estimator should have a low bias, particularly for large parameter values.
- *Continuity*: The resulting estimator should be continuous in the data to reduce instability in the model prediction.

In a sense, the sparsity, unbiasedness, and continuity properties of the penalized least squares force the penalty function to be non-differentiable at the origin and non-convex over $(0,\infty)$. Thus, to enhance the desirable properties of the l_1 penalty function, one may consider the Smoothly Clipped Absolute Deviation (SCAD) penalty, defined as:

$$p_{\varphi,a}^{SCAD}(|\beta_t|) = \begin{cases} \varphi|\beta_t| & \text{if } |\beta_t| \le \varphi, \\ -\dfrac{|\beta_t^2| - 2\,a\,\varphi\,|\beta_t| + \varphi^2}{2\,(a-1)} & \text{if } \varphi < |\beta_t| \le a\,\varphi \\ \dfrac{(a+1)\,\varphi^2}{2} & \text{if } |\beta_t| > a\,\varphi \end{cases} \text{, with } a > 2 \ ,$$

The SCAD penalty corresponds to a quadratic spline function with knots at φ and $a\varphi$ (see Figure 5.1). The value $a = 3.7$ was recommended by Fan and Li [53]. The objective function with the SCAD penalty is:

$$\hat{\beta}_t = \underset{\beta_t}{\operatorname{argmin}} \frac{1}{2}\left(\beta_t - \hat{\beta}_t^{OLS}\right)^2 + p_{\varphi,a}^{SCAD}(|\beta_t|)\,. \tag{5.11}$$

The resulting minimization of (5.11) has the following closed-form SCAD thresholding rule:

$$\hat{\beta}_t^{SCAD} = \begin{cases} \left(|\hat{\beta}_t^{OLS}| - \varphi\right)_+ \operatorname{sgn}(\hat{\beta}_t^{OLS}) & \text{if } |\hat{\beta}_t^{OLS}| \le 2\,\varphi, \\ \dfrac{(a-1)\,\hat{\beta}_t^{OLS} - \operatorname{sgn}(\hat{\beta}_t^{OLS})a\varphi}{a-2} & \text{if } 2\,\varphi < |\hat{\beta}_t^{OLS}| \le a\,\varphi \\ \hat{\beta}_t^{OLS} & \text{if } |\hat{\beta}_t^{OLS}| > a\,\varphi, \end{cases} \ .$$

5.1.8 Geometric interpretation of the Soft and SCAD thresholding rules

For a geometric interpretation of both the Soft and SCAD thresholding rules, we have considered $\hat{\boldsymbol{\beta}}^{OLS} = [0,\cdots,5]$ by steps of 0.001, to be the least squares estimate of $\boldsymbol{\beta}$. Then we have applied

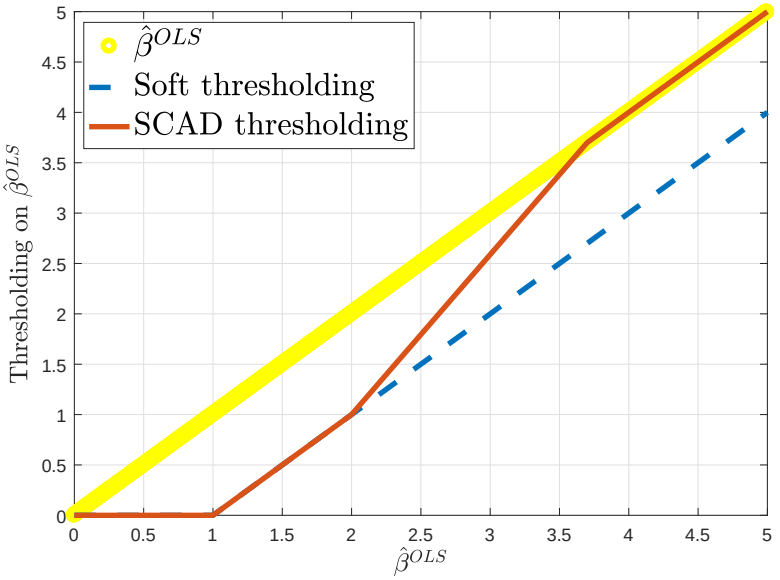


Figure 5.2: Soft and SCAD thresholding rules for $\varphi = 1$ and $a = 3.7$

the Soft and SCAD thresholding on $\hat{\boldsymbol{\beta}}^{OLS}$ with $\varphi = 1$, and $a = 3.7$ for SCAD. Figure 5.2 depicts the results of the two thresholding rules. Both the Soft and SCAD thresholding rules establish sparsity, that is, automatically set estimated coefficients that are below φ to zero. Recall from [53] that besides the sparsity property, choosing a good penalty function should result in an estimator with two further important properties: unbiasedness and continuity. The Soft thresholding rule is continuous but shifts the thresholded values by $\pm\varphi$. So that, if $|\hat{\beta}_t^{OLS}| > \varphi$ and $\hat{\beta}_t^{OLS} > 0$, the result will be $\hat{\beta}_t^{Soft} = \hat{\beta}_t^{OLS} - \varphi$, while $\hat{\beta}_t^{OLS} + \varphi$ if $\hat{\beta}_t^{OLS} < 0$. The SCAD thresholding rule is continuous and results in less bias than Soft as can be seen in Figure 5.2.

5.2 Traditional covariance estimation approaches

5.2.1 The Sample Covariance Matrix: $\hat{\boldsymbol{\Sigma}}_{SCM}$

In practice it is rare to perfectly know the distribution of the data. The Gaussian model assumption is considered the common widely hypothesis used in several applications. Suppose that we observe n independent and identically distributed (i.i.d) p-random vectors $\mathbf{x}_1, \cdots, \mathbf{x}_n$, distributed according to a multivariate Gaussian distribution with zero mean and unknown covariance matrix $\boldsymbol{\Sigma}$. That is, $\mathbf{x}_i \sim \mathcal{N}(\mathbf{0}_p, \boldsymbol{\Sigma})$, $i = 1, 2, \cdots, n$.

In this regard, the likelihood function of the matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n] \in \mathbb{R}^{p \times n}$ is:

$$L(\boldsymbol{\Sigma}, \mathbf{X}) = (2\pi)^{-np/2} \det(\boldsymbol{\Sigma})^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} \mathbf{x}_i^T \boldsymbol{\Sigma}^{-1} \mathbf{x}_i\right). \tag{5.12}$$

The log-likelihood function of $\mathbf{X}$ is:

$$\begin{aligned}
\Lambda = \log\left(L(\mathbf{\Sigma}, \mathbf{X})\right) &= \frac{-np}{2}\log(2\pi) - \frac{n}{2}\log\left(\det(\mathbf{\Sigma})\right) - \frac{1}{2}\sum_{i=1}^{n}\mathbf{x}_i^T\,\mathbf{\Sigma}^{-1}\,\mathbf{x}_i\,, \\
&= \frac{-np}{2}\log(2\pi) - \frac{n}{2}\log\left(\det(\mathbf{\Sigma})\right) - \frac{1}{2}\sum_{i=1}^{n}\mathrm{Tr}\left(\mathbf{\Sigma}^{-1}\,\mathbf{x}_i\,\mathbf{x}_i^T\right)\,, \\
&= \frac{-np}{2}\log(2\pi) - \frac{n}{2}\log\left(\det(\mathbf{\Sigma})\right) - \frac{1}{2}\mathrm{Tr}\left(\mathbf{\Sigma}^{-1}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^T\right)\,, \\
&= \frac{-np}{2}\log(2\pi) + \frac{n}{2}\log\left(\det(\mathbf{\Sigma})\right)^{-1} - \frac{1}{2}\mathrm{Tr}\left(\mathbf{\Sigma}^{-1}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^T\right)\,, \\
&= \frac{-np}{2}\log(2\pi) + \frac{n}{2}\log\left(\det(\mathbf{\Sigma}^{-1})\right) - \frac{1}{2}\mathrm{Tr}\left(\mathbf{\Sigma}^{-1}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^T\right)\,. \quad (5.13)
\end{aligned}$$

Note that the trace trick $\mathbf{x}_i^T\,\mathbf{\Sigma}^{-1}\,\mathbf{x}_i = \mathrm{Tr}\left(\mathbf{\Sigma}^{-1}\,\mathbf{x}_i\,\mathbf{x}_i^T\right)$ came from two neat properties of the trace:

- $\mathrm{Tr}(c) = c$ when c is a constant (i.e. $\mathbf{x}_i^T\,\mathbf{\Sigma}^{-1}\,\mathbf{x}_i \in \mathbb{R}$) ,
- $\mathrm{Tr}\left(\mathbf{U}\,\mathbf{V}\,\mathbf{W}\right) = \mathrm{Tr}\left(\mathbf{W}\,\mathbf{U}\,\mathbf{V}\right) = \mathrm{Tr}\left(\mathbf{V}\,\mathbf{W}\,\mathbf{U}\right)$, with $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$ are matrices.

Taking now the derivative w.r.t. $\mathbf{\Sigma}^{-1}$ and setting it to zero, we have:

$$\frac{n}{2\,\det\left(\mathbf{\Sigma}^{-1}\right)}\det\left(\mathbf{\Sigma}^{-1}\right)\mathbf{\Sigma}^T - \frac{1}{2}\sum_{i=1}^{n}\mathbf{x}_i^T\,\mathbf{x}_i = \frac{n}{2}\mathbf{\Sigma}^T - \frac{1}{2}\sum_{i=1}^{n}\mathbf{x}_i^T\,\mathbf{x}_i = \mathbf{0}\,.$$

Hence,

$$\boxed{\hat{\mathbf{\Sigma}}_{SCM} = [\hat{\sigma}_{g,l}]_{p\times p} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\,\mathbf{x}_i^T}$$

**Some remarks

Note that $\dfrac{\partial}{\partial\mathbf{\Sigma}^{-1}}\log\left(\det(\mathbf{\Sigma}^{-1})\right) = \dfrac{1}{\det\left(\mathbf{\Sigma}^{-1}\right)}\dfrac{\partial\det\left(\mathbf{\Sigma}^{-1}\right)}{\partial\mathbf{\Sigma}^{-1}} = \dfrac{1}{\det\left(\mathbf{\Sigma}^{-1}\right)}\mathrm{adj}\left(\mathbf{\Sigma}^{-T}\right)$,
where $\mathrm{adj}\left(\mathbf{\Sigma}^{-T}\right)$ is the adjugate of matrix $\mathbf{\Sigma}^{-T}$, defined as: $\mathrm{adj}\left(\mathbf{\Sigma}^{-T}\right) = \det\left(\mathbf{\Sigma}^{-T}\right)\left(\mathbf{\Sigma}^{-T}\right)^{-1} = \det\left(\mathbf{\Sigma}^{-1}\right)\mathbf{\Sigma}^T$.

5.2.2 Covariance estimation based on the OLS technique: $\hat{\mathbf{\Sigma}}_{OLS}$

The estimation of covariance matrices through optimization of an objective function (e.g. a log-likelihood function) is usually a difficult numerical problem, since the resulting estimates should be positive definite matrices.

Covariance estimation via linear regression

In order to address the positivity definiteness constraint problem of $\hat{\mathbf{\Sigma}}_{SCM}$, Pourahmadi [124] has modeled the covariance matrices via linear regressions. This is done by letting $\hat{\mathbf{x}} = [\hat{x}_1, \ldots, \hat{x}_p]^T \in \mathbb{R}^p$, and consider each element $\hat{x}_t$, $t \in [1, p]$, as the linear least squares predictor of x_t based on

its $t-1$ predecessors $\{x_j\}_{j\in[1,\,t-1]}$. In particular, for $t \in [1,\ p]$, let

$$\hat{x}_t = \sum_{j=1}^{t-1} C_{t,j}\, x_j\,.$$

For each value of $t \geq 1$, we get:

$$\begin{aligned}
x_1 &= \epsilon_1\,,\\
x_2 &= C_{2,1}\, x_1 + \epsilon_2\,,\\
x_3 &= C_{3,1}\, x_1 + C_{3,2}\, x_2 + \epsilon_3\,,\\
&\vdots\\
x_p &= C_{p,1}\, x_1 + C_{p,2}\, x_2 + \cdots + C_{p,p-1}\, x_{p-1} + \epsilon_p\,,
\end{aligned}$$

where $\epsilon_t = x_t - \hat{x}_t$ for $t \in [1,\, p]$ denotes the prediction error with variance $\mathrm{var}\,(\epsilon_t) = \mathbb{E}\left[(\epsilon_t)^2\right] = \theta_t^2$. Note that for $t = 1$, let $\hat{x}_1 = \mathbb{E}\,(x_1) = 0$, and hence, $\mathrm{var}(\epsilon_1) = \theta_1^2 = \mathbb{E}\left[(x_1)^2\right]$.
This is equivalent to:

$$\boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_p \end{pmatrix} = \begin{bmatrix} 1 & & & & \\ -C_{2,1} & 1 & & & \\ -C_{3,1} & -C_{3,2} & 1 & & \\ \vdots & \vdots & \cdots & \ddots & \\ -C_{p,1} & -C_{p,2} & \cdots & -C_{p,p-1} & 1 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} = \mathbf{T}\,\mathbf{x}\,,$$

where $\mathbf{T}$ is a unit lower triangular matrix with $-C_{t,j}$ in the (t,j)th position for $t \in [2,\, p]$ and $j \in [1,\, t-1]$, and $\mathbf{x} = [x_1,\, x_2,\, \cdots,\, x_p]^T \in \mathbb{R}^p$.
One has $\mathrm{cov}\,(\boldsymbol{\epsilon}) = \mathbf{T}\,\boldsymbol{\Sigma}\,\mathbf{T}^T = \mathbf{D}$. Thus, $\mathbf{D}^{-1} = \mathbf{T}^{-T}\,\boldsymbol{\Sigma}^{-1}\,\mathbf{T}^{-1} \Rightarrow \boldsymbol{\Sigma}^{-1} = \mathbf{T}^T\,\mathbf{D}^{-1}\,\mathbf{T}$, where $\mathbf{D}$ is a diagonal matrix with entries $\theta_1^2,\, \theta_2^2,\, \cdots,\, \theta_p^2$. A very interesting consequence is that for any $\left(\hat{\mathbf{T}},\, \hat{\mathbf{D}}\right)$, $\hat{\boldsymbol{\Sigma}} = \hat{\mathbf{T}}^{-1}\,\hat{\mathbf{D}}\,\hat{\mathbf{T}}^{-T}$ is always guaranteed to be positive definite.

The estimator $\hat{\boldsymbol{\Sigma}}_{OLS}$

Given a sample $\{\mathbf{x}_i\}_{i\in[1,\,n]}$, we have:

$$x_{i,t} = \sum_{j=1}^{t-1} C_{t,j}\, x_{i,j} + \epsilon_{i,t} \quad ,t \in [2,\, p], \quad i \in [1,\, n]. \tag{5.14}$$

Hence, for any $t \in [2, p]$, we have:

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \\ \vdots \\ x_{n,t} \end{bmatrix} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,t-1} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,t-1} \\ \vdots & & & \\ x_{n,1} & x_{n,2} & \cdots & x_{n,t-1} \end{bmatrix} \begin{bmatrix} C_{t,1} \\ C_{t,2} \\ \vdots \\ C_{t,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \vdots \\ \epsilon_{n,t} \end{bmatrix} \tag{5.15}$$

(5.15) is similar to a simple linear regression model:

$$\mathbf{y}_t = \mathbf{A}_{n,t}\,\boldsymbol{\beta}_t + \mathbf{e}_t\,, \tag{5.16}$$

where $\mathbf{y}_t = [x_{1,t}, \cdots, x_{n,t}]^T \in \mathbb{R}^n$, $\mathbf{A}_{n,t} = [x_{i,j}]_{n\times(t-1)}$, $\boldsymbol{\beta}_t = [C_{t,1}, \cdots, C_{t,t-1}]^T \in \mathbb{R}^{(t-1)}$, and $\mathbf{e}_t = [\epsilon_{1,t}, \cdots, \epsilon_{n,t}]^T \in \mathbb{R}^n$.

When $n > p$, the OLS estimate of $\boldsymbol{\beta}_t$ and the corresponding residual variance are plugged in $\mathbf{T}$ and $\mathbf{D}$ for each $t \in [2, p]$, respectively. At the end, one obtains the estimator:

$$\boxed{\hat{\boldsymbol{\Sigma}}_{OLS} = \hat{\mathbf{T}}_{OLS}^{-1}\,\hat{\mathbf{D}}_{OLS}\,\hat{\mathbf{T}}_{OLS}^{-T}}$$

Note that $\hat{\mathbf{T}}_{OLS}$ has -$\hat{C}_{t,j}^{OLS}$ in the (t,j)th position for $t \in [2, p]$ and $j \in [1, t-1]$.

5.2.3 The Tyler estimator (also known as Fixed-Point (FP)): $\hat{\boldsymbol{\Sigma}}_{FP}$

In real world hyperspectral imagery, the environment is heterogeneous and thus the true model assumption might be so far from Gaussianity. However, it has been shown that the Gaussian distribution is not always a good model assumption for background characterization in hyperspectral imaging. In that case, both $\hat{\boldsymbol{\Sigma}}_{SCM}$ and $\hat{\boldsymbol{\Sigma}}_{OLS}$ are not robust and so they will face difficulties in estimating the true covariance without an extreme amount of errors. One of the most general and acknowledged models for background statistics characterization is the family of Elliptically Contoured Distributions, originally introduced by Kelker in [90].

The Compound Gaussian (CG) distributions represent an important subclass of the elliptical contoured distributions that are widely used in signal processing applications, e.g. for wireless radio propagation problems [157], radar clutter echoes modeling [65], and hyperspectral background characterization [102, 116].

The main idea behind this model is to suppose that the background, locally Gaussian, presents spatially variable power. More precisely, a random vector $\mathbf{c}$ has a Compound Gaussian distribution if it can be written as:

$$\mathbf{c} = \tau^{1/2}\,\mathbf{z}\,, \tag{5.17}$$

where $\tau > 0$ is called "texture" whose distribution is not known, whereas $\mathbf{z}$ is a random vector with a multivariate Gaussian distribution with zero mean and unknown covariance matrix $\boldsymbol{\Sigma}$.

The probability density function (PDF) of $\mathbf{c}$ can be expressed as:

$$f_{\mathbf{c}}(\mathbf{c}) = \frac{1}{\pi^p\,\det(\boldsymbol{\Sigma})}\int_0^{+\infty}\frac{1}{\tau^p}\,\exp\left(-\frac{\mathbf{c}^T\boldsymbol{\Sigma}^{-1}\mathbf{c}}{\tau}\right)f_\tau(\tau)\,d\tau\,, \tag{5.18}$$

where $f_\tau(.)$ is the texture PDF. In the case when $\tau \sim \mathrm{Gam}(\nu, 1/\nu)$ with $\nu > 0$, each vector $\mathbf{x}_i$ is said to have a p-variate K-distribution with shape parameter ν. In the limit $\nu \to \infty$, the K-distribution reduces to the multivariate normal distribution.

Given that τ is unknown, most of the time, the MLE associated with this modeling can not be obtained. However, many authors such as [66], [119] and [41] have obtained an approached MLE known as the Tyler estimator (aka Fixed-Point (FP) estimator).

Let $(\mathbf{c}_1, \cdots, \mathbf{c}_n)$ be a n-sample with the same distribution as $\mathbf{c}$ in equation (5.17). The Fixed-Point estimator is defined as the unique solution, up to a scale factor, of the equation:

$$\boxed{\hat{\mathbf{\Sigma}}_{FP} = \frac{p}{n} \sum_{i=1}^{n} \frac{\mathbf{x}_i \, \mathbf{x}_i^T}{\mathbf{x}_i^T \, \hat{\mathbf{\Sigma}}_{FP}^{-1} \, \mathbf{x}_i}}$$

5.2.4 Comparative study of the traditional covariance estimators using the ANMF target detector

A brief overview of the Adaptive Normalized Matched Filter (ANMF)

In performing sub-pixel detection, one can assume that the background is homogeneous and follows a multivariate normal distribution. However, the target spectrum can be simply considered to be superimposed with the background spectrum, so that they interact in an additive manner. The detection scheme related to an observed spectrum $\mathbf{x}$ is given by:

$$\begin{cases} H_0 : \mathbf{x} = \mathbf{b} \sim \mathcal{N}(\mathbf{0}_p, \mathbf{\Sigma}) \\ H_1 : \mathbf{x} = \delta \, \mathbf{t} + \beta \, \mathbf{b} \sim \mathcal{N}(\delta \, \mathbf{t}, \beta^2 \, \mathbf{\Sigma}), \; \delta, \beta > 0 \, , \end{cases} \tag{5.19}$$

which is known as the additive signal model for sub-pixel targets. Since for unresolved targets, the area covered by the background is different under the two hypothesis, it is more reasonable to consider that the covariance under hypothesis H_1 is not equal to $\mathbf{\Sigma}$, but to $\beta^2 \, \mathbf{\Sigma}$. In this case, the background has the same structure but different variances. The Likelihood Ratio Test (LRT) approach for the detection problem (5.19) leads to the following invariant Normalized Matched Filter (NMF) detector:

$$D_{NMF} = \frac{\left| \mathbf{t}^T \, \mathbf{\Sigma}^{-1} \, \mathbf{x} \right|^2}{\left(\mathbf{t}^T \, \mathbf{\Sigma}^{-1} \, \mathbf{t} \right) \left(\mathbf{x}^T \, \mathbf{\Sigma}^{-1} \, \mathbf{x} \right)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \, , \tag{5.20}$$

where η is the decision threshold to yield the desired probability of false alarm P_{fa}. The two-step Adaptive Normalized Matched Filter (ANMF) is obtained when $\mathbf{\Sigma}$ is replaced by its estimate. In this context, if D_{ANMF} exceeds η, the hypothesis H_1 is chosen; whereas H_0 otherwise.

Evaluation performance of ANMF using the traditional covariance estimators

To measure the performance of the ANMF target detector, it is classical to draw the Receiver Operating Characteristics (ROC) curves. More particularly, the curve depicts the probability of detection (P_d) as a function of the Signal to Noise Ratio (SNR). the SNR for the problem (5.19) is:

$$SNR = \frac{\delta^2}{\beta^2} \, \mathbf{t}^T \, \mathbf{\Sigma^{-1}} \, \mathbf{t} \, . \tag{5.21}$$

For the evaluations, we specifically choose β=1 in (5.21).

The performance of each of the estimators $\hat{\mathbf{\Sigma}}_{SCM}$, $\hat{\mathbf{\Sigma}}_{OLS}$ and $\hat{\mathbf{\Sigma}}_{FP}$, when they are plugged-in into the Normalized Matched Filter (NMF) target detector, is evaluated in terms of probability of detection P_d versus SNR for a probability of false alarm $P_{fa} = 10^{-2}$. The experiments are

conducted for the Autoregressive model of order 1, AR(1), $\mathbf{\Sigma} = [\sigma_{gl}]_{p\times p}$, where $\sigma_{gl} = c^{|g-l|}$, for $c = 0.3$.
We selected $p = \{10, 30, 60, 80\}$, $\nu \in \{\infty, 0.5, 0.1\}$, and $n = 80$. All computations were performed using 10^5 Monte-Carlo trials. The artificial target we consider is a vector containing normally distributed pseudorandom values. Figure 5.3, 5.4 and 5.5 depict the evaluation outputs for $\nu = \infty$, $\nu = 0.5$ and $\nu = 0.1$, respectively.

- From all the figures (5.3, 5.4, 5.5), we can observe that $\hat{\mathbf{\Sigma}}_{OLS}$ always achieve (slightly) higher detection results than to those of $\hat{\mathbf{\Sigma}}_{SCM}$.
- When $\nu = \infty$ The Fixed-Point estimator $\hat{\mathbf{\Sigma}}_{FP}$ always achieve slightly lower target detection results than to those of $\hat{\mathbf{\Sigma}}_{SCM}$ and $\hat{\mathbf{\Sigma}}_{OLS}$ (see Figure 5.3). This is to be very expected since for $\nu = \infty$, the assumption on the data is completely Gaussian.
- For $\nu = 0.5$ and $\nu = 0.1$ The Fixed-Point $\hat{\mathbf{\Sigma}}_{FP}$ achieves higher detection results than to those of $\hat{\mathbf{\Sigma}}_{SCM}$ and $\hat{\mathbf{\Sigma}}_{OLS}$ (see Figure 5.4, 5.5). This is because $\hat{\mathbf{\Sigma}}_{FP}$ is a robust estimator.
- The traditional covariance estimators are ideal only when n tends to ∞ Obviously, when the spectral dimension p is considered large (i.e. $p = 60$ and $p = 80$) compared to the number of observed data n, $\hat{\mathbf{\Sigma}}_{SCM}$, $\hat{\mathbf{\Sigma}}_{OLS}$ and $\hat{\mathbf{\Sigma}}_{FP}$ face difficulties in estimating $\mathbf{\Sigma}$ without an extreme amount of errors. Thus, they behave poorly in large dimensions, unless some regularization techniques (i.e. Shrinkage [34, 35, 84, 94, 95, 117, 118, 130], Random Matrix Theory (RMT) [39, 40, 135–137], Sparsity [13, 29, 82, 127, 151]) are applied.

5.3 Some Research works for alleviating the high dimensional covariance estimation challenge via sparsity

The covariance matrix estimation problem plays an essential role in time series analysis [61], spatial data analysis [42], longitudinal data analysis [60], STAP and MIMO STAP radar applications [39], and hyperspectral imagery. Obviously, when the spectral dimension p is considered large compared to the number of observed data n, traditional covariance estimators such as $\hat{\mathbf{\Sigma}}_{SCM}$, $\hat{\mathbf{\Sigma}}_{OLS}$ and $\hat{\mathbf{\Sigma}}_{FP}$, face difficulties in estimating $\mathbf{\Sigma}$ without an extreme amount of errors. However, the problem of estimating high-dimensional covariance matrices has been extensively studied. Realizing the challenges brought by the high covariance dimensionality, researchers have proposed various regularization techniques to consistently estimate $\mathbf{\Sigma}$ based on the assumption that the covariance matrix is sparse, namely, many entries are zero.
Bickel et al. [13] proposed a banded version of $\hat{\mathbf{\Sigma}}_{SCM}$. Unfortunately, this kind of regularization does not always guarantee positive definiteness of the estimator.
In [127], a class of generalized thresholding operators applied to the off-diagonal entries of $\hat{\mathbf{\Sigma}}_{SCM}$ have been discussed. These operators combine shrinkage with thresholding and have the advantage

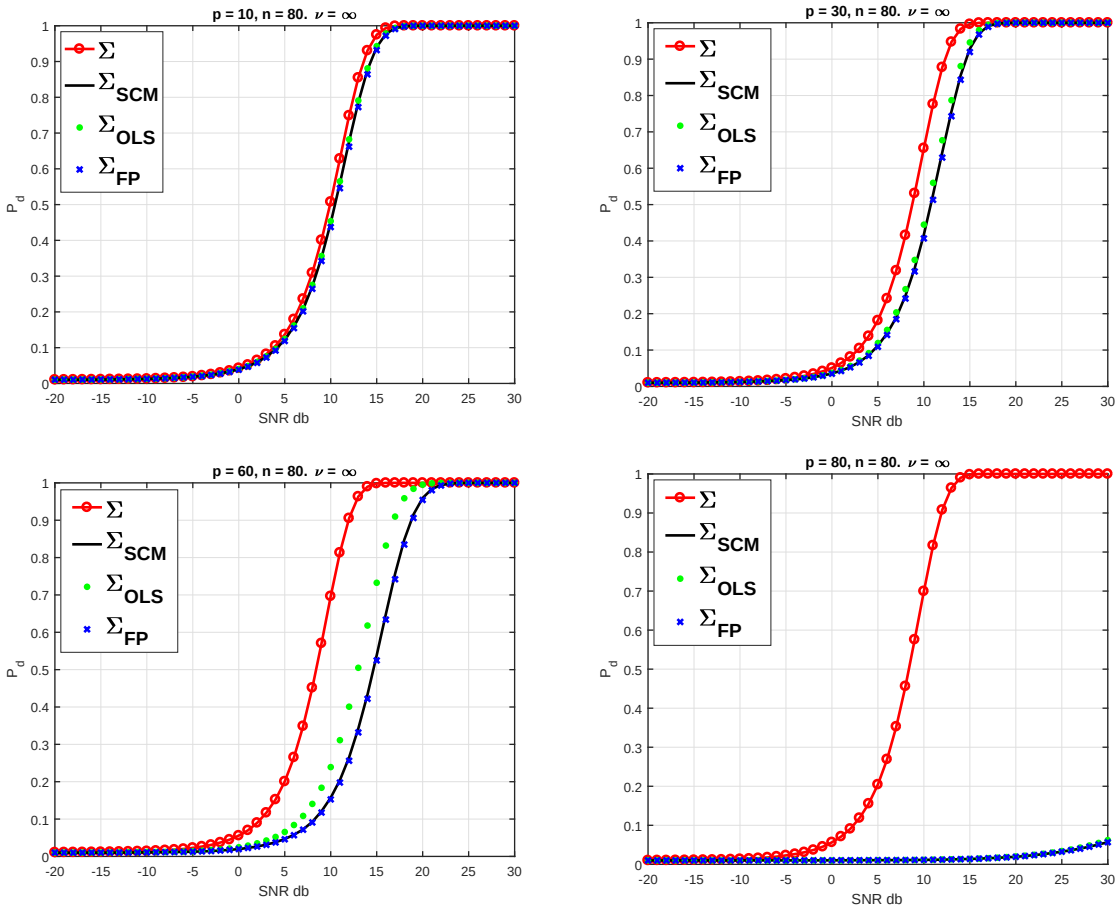


Figure 5.3: Target detection results of the traditional covariance estimators $\{\hat{\mathbf{\Sigma}}_{SCM}, \hat{\mathbf{\Sigma}}_{OLS}, \hat{\mathbf{\Sigma}}_{FP}\}$, using the ANMF detector for $p = \{10, 30, 60, 80\}$, $n = 80$, $\nu = \infty$

to estimate the true zeros as zeros with high probability. These operators (e.g., Soft and SCAD), though simple, do not always guarantee positive definiteness of the thresholded version of $\hat{\mathbf{\Sigma}}_{SCM}$. In this context, Liu et al. [98] have generalized the work in [127] by adding an explicit eigenvalue constraint. and hence, the thresholded estimated covariance matrix simultaneously achieves sparsity and positive definiteness.

In [29], the covariance matrix is constrained to have an eigen decomposition which can be represented as a sparse matrix transform (SMT) that decomposes the eigen-decomposition into a product of very sparse transformations. The resulting estimator, denoted as $\hat{\mathbf{\Sigma}}_{SMT}$ in this dissertation, is always guaranteed to be positive definite.

In addition to the above review, some other works have attempted to enforce sparsity of the covariance matrix via its Cholesky factor $\mathbf{T}$. Hence, yielding sparse covariance estimators that are always guaranteed to be positive definite. For example, in [151], Pourahmadi et al. proposed to smooth the first few subdiagonals of $\hat{\mathbf{T}}_{OLS}$ and set to zero the remaining ones. In [82], Huang *et al.* proposed to directly estimate a sparse version of $\mathbf{T}$ by penalizing the negative normal log-likelihood of the matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n]$ with an l_1-norm penalty function. Hence, allowing the zeros to

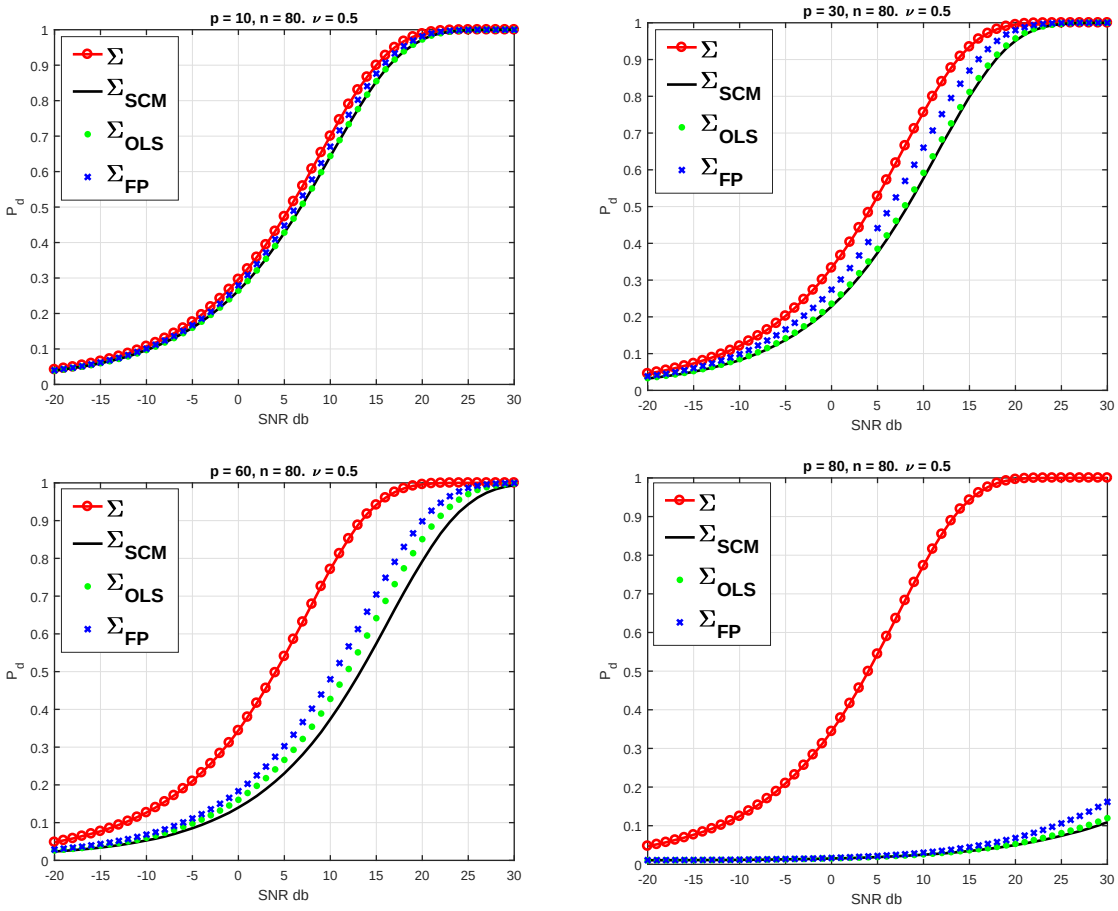


Figure 5.4: Target detection results of the traditional covariance estimators $\{\hat{\boldsymbol{\Sigma}}_{SCM}, \hat{\boldsymbol{\Sigma}}_{OLS}, \hat{\boldsymbol{\Sigma}}_{FP}\}$ using the ANMF detector for $p = \{10, 30, 60, 80\}$, $n = 80$, $\nu = 0.5$

be irregularly placed in the Cholesky factor. This seems to be an advantage over the work in [151].

5.3.1 Banding the Sample Covariance Matrix

Given the $p \times p$ sample covariance matrix $\hat{\boldsymbol{\Sigma}}_{SCM} = [\hat{\sigma}_{g,l}]_{p\times p}$ in section 5.2.1, Bickel et al. [13] have proposed a banded version of $\hat{\boldsymbol{\Sigma}}_{SCM}$, denoted as $B_m\left(\hat{\boldsymbol{\Sigma}}_{SCM}\right)$ in this dissertation:

$$B_m\left(\hat{\boldsymbol{\Sigma}}_{SCM}\right) = \left[\hat{\sigma}_{g,l}\, \mathbb{1}(|g-l| \leq m)\right],$$

where $0 \leq m < p$ is the banding parameter.

The performance of the estimator $B_m(\hat{\boldsymbol{\Sigma}}_{SCM})$ depends critically on the optimal choice of the banding parameter m which is usually selected using a cross-validation method. However, this kind of regularization does not always guarantee positive definiteness.

5.3.2 Thresholding the Sample Covariance matrix

There are several componentwise regularization methods for estimating a sparse (or approximately sparse) covariance matrix. Thresholding is by far the most popular.

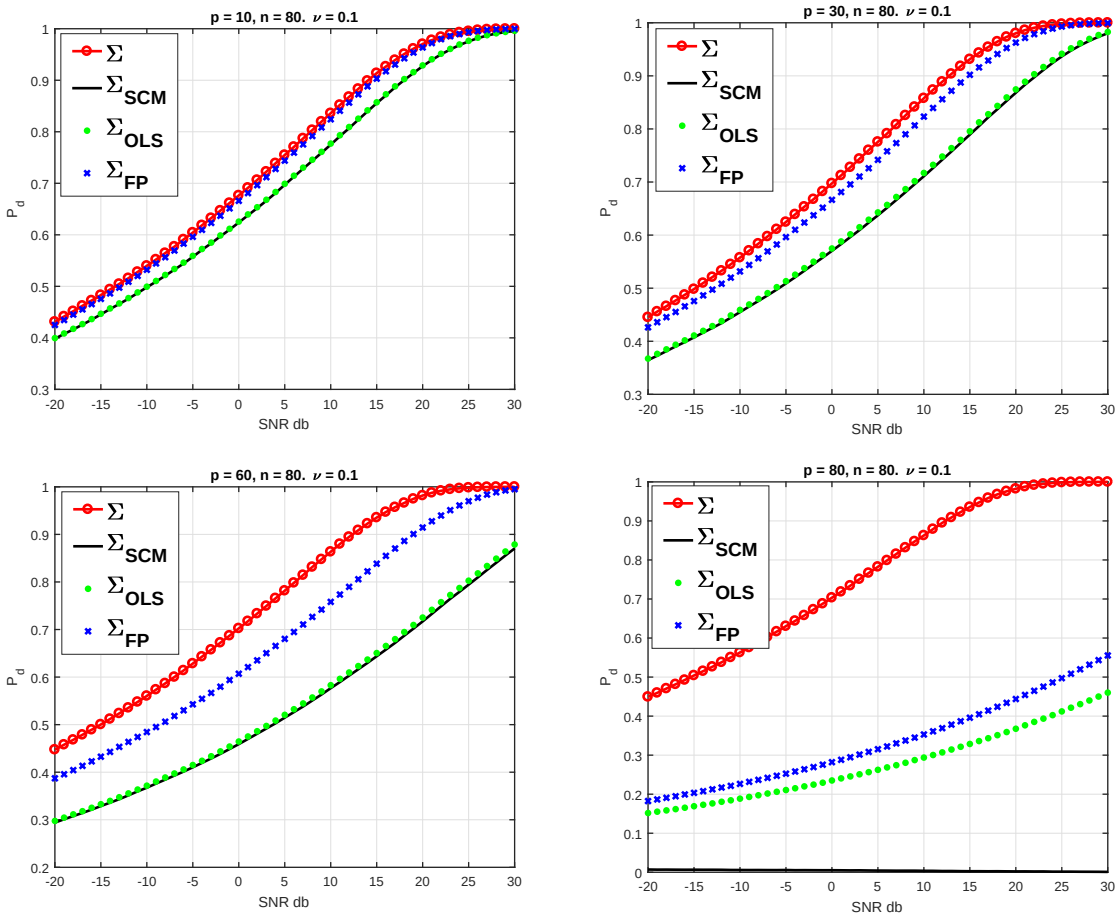


Figure 5.5: Target detection results of the traditional covariance estimators $\{\hat{\mathbf{\Sigma}}_{SCM}, \hat{\mathbf{\Sigma}}_{OLS}, \hat{\mathbf{\Sigma}}_{FP}\}$ using the ANMF detector for $p = \{10, 30, 60, 80\}$, $n = 80$, $\nu = 0.1$

In [127], a class of generalized thresholding operators such as Soft and SCAD applied on the off-diagonal entries of $\hat{\mathbf{\Sigma}}_{SCM}$ have been discussed. These operators combine shrinkage with thresholding and have the advantage to estimate the true zeros as zeros with high probability. For any $\varphi \geq 0$, define a matrix thresholding operator $Th\,(.)$ and denote by $Th\left(\hat{\mathbf{\Sigma}}_{SCM}\right) = [Th\,(\hat{\sigma}_{g,l})]$, with $g \neq l$, the matrix resulting from applying a specific thresholding operator $Th(.)$ to the off-diagonal elements of the matrix $\hat{\mathbf{\Sigma}}_{SCM}$. The thresholding operator $Th\,(.)$ is defined by:

$$Th\left(\hat{\mathbf{\Sigma}}_{SCM}\right) = \underset{\mathbf{\Sigma}}{\operatorname{argmin}} \sum_{g=1}^{p}\sum_{l=1}^{p}\left\{\frac{1}{2}\left(\hat{\sigma}_{g,l} - \sigma_{g,l}\right)^2 + p_{\varphi}\{|\sigma_{g,l}|\}\right\}, \quad g \neq l\,, \tag{5.22}$$

where p_{φ} is a penalty that can include several penalties (i.e. l_1 penalty, and $SCAD$-penalty).

- *For an l_1 penalty*: problem (5.22) is re-written as:

$$\hat{\mathbf{\Sigma}}_{SCM}^{Soft} = \underset{\mathbf{\Sigma}}{\operatorname{argmin}} \sum_{g=1}^{p}\sum_{l=1}^{p}\left\{\frac{1}{2}\left(\hat{\sigma}_{g,l} - \sigma_{g,l}\right)^2 + \varphi\,|\sigma_{g,l}|\right\}, \quad g \neq l\,, \tag{5.23}$$

- *For an SCAD penalty*: problem (5.22) is re-written as:

$$\hat{\mathbf{\Sigma}}_{SCM}^{SCAD} = \underset{\mathbf{\Sigma}}{\operatorname{argmin}} \sum_{g=1}^{p}\sum_{l=1}^{p}\left\{\frac{1}{2}\left(\hat{\sigma}_{g,l}-\sigma_{g,l}\right)^2 + p_{\varphi,\,a}^{SCAD}\{|\sigma_{g,l}|\}\right\}, \quad g \neq l\,, \tag{5.24}$$

  Both problems (5.23) and (5.24) admit a closed-form Soft and SCAD thresholding rules [127].

$Th\,(.)$ is said to have generalized thresholding rule if it satisfies three interesting conditions for all $\hat{\sigma}_{g,l}$:

1. $|Th\,(\hat{\sigma}_{g,l})| \leq |\hat{\sigma}_{g,l}|$, which establishes shrinkage,
2. $Th\,(\hat{\sigma}_{g,l}) = 0$ if $|\hat{\sigma}_{g,l}| \leq \varphi$, which establishes thresholding,
3. $|Th\,(\hat{\sigma}_{g,l}) - \hat{\sigma}_{g,l}| \leq \varphi$, which means that the amount of shrinkage must not exceed the selected thresholding parameter φ.

Obviously, both Soft and SCAD thresholding rules satisfy all the three conditions, and hence, they are a class of generalized thresholding that combine shrinkage with thresholding.

In general, threshold estimators carries almost no computational burden other than the selection of the tuning parameter φ (usually selected using a cross-validation method), but does not necessarily preserve positive-definiteness.

5.3.3 An overview of the work in [82]

Note that $\det(\mathbf{T}) = 1$ and $\mathbf{\Sigma} = \mathbf{T}^{-1}\,\mathbf{D}\,\mathbf{T}^{-T}$. It follows that $\det(\mathbf{\Sigma}) = \det(\mathbf{D}) = \prod_{t=1}^{p}\theta_t^2$. Hence, the negative normal log-likelihood of $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_n] \in \mathbb{R}^{p\times n}$, ignoring an irrelevant constant, satisfies:

$$\begin{aligned}
\Lambda = -2\,\log\left(L\left(\mathbf{\Sigma}, \mathbf{x}_1, \cdots, \mathbf{x}_n\right)\right) &= n\,\log(\det(\mathbf{D})) + \mathbf{X}^T\,\left(\mathbf{T}^T\,\mathbf{D}^{-1}\,\mathbf{T}\right)\,\mathbf{X}\,,\\
&= n\,\log\left(\det(\mathbf{D})\right) + \left(\mathbf{T}\,\mathbf{X}\right)^T\,\mathbf{D}^{-1}\,\left(\mathbf{T}\,\mathbf{X}\right)\,,\\
&= n\sum_{t=1}^{p}\log\theta_t^2 + \sum_{t=1}^{p}\sum_{i=1}^{n}\epsilon_{i,t}^2/\theta_t^2\,.
\end{aligned}$$

By adding a penalty function $\sum_{t=2}^{p}\sum_{j=1}^{t-1} p_\varphi\{|C_{t,j}|\}$ to Λ, where $p_\varphi \in \{p_\varphi^{L_1}, p_{\varphi,a>2}^{SCAD}\}$ with $\varphi \in [0,\infty)$, we have:

$$n\log\theta_1^2 + \sum_{i=1}^{n}\frac{\epsilon_{i,1}^2}{\theta_1^2} + \sum_{t=2}^{p}\left(n\log\,\theta_t^2 + \sum_{i=1}^{n}\frac{\epsilon_{i,t}^2}{\theta_t^2} + \sum_{j=1}^{t-1}p_\varphi\{|C_{t,j}|\}\right) \tag{5.25}$$

Obviously, minimizing (5.25) with respect to θ_1^2 and θ_t^2 gives the solutions:

$$\hat{\theta}_1^2 = \frac{1}{n}\sum_{i=1}^{n}\epsilon_{i,1}^2 = \frac{1}{n}\sum_{i=1}^{n}x_{i,1}^2\,,$$

and

$$\hat{\theta}_t^2 = \frac{1}{n}\sum_{i=1}^{n} \epsilon_{i,t}^2 = \frac{1}{n}\sum_{i=1}^{n} \left(x_{i,t} - \sum_{j=1}^{t-1} C_{t,j}\, x_{i,j} \right)^2 ,$$

respectively.

It remains to estimate the entries of $\mathbf{T}$ by minimizing (5.25) with respect to $C_{t,j}$. From equation (5.14) and (5.16), the minimization problem to solve for each $t \in [2,\, p]$ is:

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_t &= \underset{\boldsymbol{\beta}_t}{\operatorname{argmin}} \sum_{i=1}^{n} \frac{\epsilon_{i,t}^2}{\theta_t^2} + \sum_{j=1}^{t-1} p_\varphi\{|C_{t,j}|\}\,, \\
&= \underset{\boldsymbol{\beta}_t}{\operatorname{argmin}} \frac{1}{\theta_t^2} \sum_{i=1}^{n} \left(x_{i,t} - \sum_{j=1}^{t-1} C_{t,j}\, x_{i,j} \right)^2 + \sum_{j=1}^{t-1} p_\varphi\{|C_{t,j}|\}\,, \\
&= \underset{\boldsymbol{\beta}_t}{\operatorname{argmin}} \frac{1}{\theta_t^2} \left\| \mathbf{y}_t - \mathbf{A}_{n,t}\, \boldsymbol{\beta}_t \right\|_F^2 + \sum_{j=1}^{t-1} p_\varphi\{|C_{t,j}|\}\,. \qquad (5.26)
\end{aligned}$$

In [82], the authors have been interested on the l_1 norm, that is, $p_\varphi\{|C_{t,j}|\} = \varphi\,|C_{t,j}|$. However, the authors have used the local quadratic approximation (LQA) [53] of the l_1-norm in order to get a closed-form solution for $\boldsymbol{\beta}_t$ in equation (5.26).

Summary

An introduction to the linear regression analysis was given in this chapter. More precisely, all the necessary information concerning the least squares method and some of the penalized least squares techniques such as the penalization via an l_2, l_1, and $SCAD$ penalties were briefly outlined. The chapter also provided in detail some of the traditional covariance estimators $\hat{\boldsymbol{\Sigma}}_{SCM}$, $\hat{\boldsymbol{\Sigma}}_{OLS}$ and $\hat{\boldsymbol{\Sigma}}_{FP}$, and a comparative study of them on some Monte-Carlo simulations for hyperspectral target detection. The evaluations show that $\hat{\boldsymbol{\Sigma}}_{OLS}$ always achieve slightly better target detection performances than to those of $\hat{\boldsymbol{\Sigma}}_{SCM}$. Recall that $\hat{\boldsymbol{\Sigma}}_{OLS}$ has been developed mainly to address the positivity definiteness constraint of $\hat{\boldsymbol{\Sigma}}_{SCM}$. The evaluations also demonstrate the robustness of $\hat{\boldsymbol{\Sigma}}_{FP}$. Under a Gaussian assumption, $\hat{\boldsymbol{\Sigma}}_{FP}$ always achieve a slightly lower detection results than to those of $\hat{\boldsymbol{\Sigma}}_{SCM}$ and $\hat{\boldsymbol{\Sigma}}_{OLS}$; whereas in non-Gaussian cases, $\hat{\boldsymbol{\Sigma}}_{FP}$ is robust and achieves higher detection results than to those of $\hat{\boldsymbol{\Sigma}}_{SCM}$ and $\hat{\boldsymbol{\Sigma}}_{OLS}$. However, the three traditional estimators behave poorly when the spectral dimension p is considered large compared to the sample size n. In this regard, a lot of research works have been done to alleviate the high covariance dimensionality challenge by assuming that the covariance matrix is sparse, namely, many entries are zero. This chapter has also provided in detail some of these research works.

Research is creating new knowledge.

— Neil Armstrong

6

Imposing Sparsity on the Covariance Matrix via Its Cholesky Factor

Contents

► *Synopsis* This chapter outlines two methods that are based on pre-existing works to impose sparsity on the covariance matrix $\mathbf{\Sigma}$ via its unit lower triangular matrix (aka Cholesky factor) $\mathbf{T}$. More precisely, four sparse covariance estimators are developed and which are always guaranteed to be positive definite. This chapter is split into four parts.

1. **The first part** "*System Overview of the proposed work*" gives a brief overview of the main contributions as well as the obtained results.

2. **The second part** "*Our main contributions*" outlines in detail the four covariance estimators that are developed under the assumption that the true unknown covariance matrix (of the background surrounding the test pixel) is sparse, namely, many entries are zero.

3. **The third part** "*Experiments and Analysis*" evaluate the proposed sparse covariance estimators on some Monte-Carlo simulations as well as experimental data for hyperspectral

target detection.

4. **The forth part** "*Summary of the whole Part III and some future directions*" outlines a brief summary of the main ideas of Chapter 5 as well as the main contributions of this chapter. This part also provides some directions for future work.

6.1 System Overview of the proposed work

We put forth two simple methods for imposing sparsity on the covariance matrix via its Cholesky factor $\mathbf{T}$. Our first method is very related to the work in [127] (that is, in section 5.3.2 of Chapter 5), but attempts to render $\hat{\mathbf{\Sigma}}_{OLS}$ sparse (instead of $\hat{\mathbf{\Sigma}}_{SCM}$) by thresholding its Cholesky factor $\hat{\mathbf{T}}_{OLS}$ using operators such as Soft and SCAD. The second method aims to generalize the work in [82] in order to be used for various penalty functions (not only the l_1 penalty as in [82]). The two methods allow the zeros to be irregularly placed in the Cholesky factor, and which seems to be an advantage over the work in [151].

Clearly, in real world hyperspectral imagery, the true covariance model is not known, and hence, there is no prior information on its degree of sparsity. Enforcing sparsity on the covariance matrix seems to be a strong assumption, but can be critically important if the true covariance model (of the background surrounding the test pixel) for a given HSI is indeed sparse. That is, taking advantage of the possible sparsity in the estimation can potentially improve the target detection performance, as can be seen from the experimental results later. On the other hand, while the true covariance model may not be sparse (or not highly sparse), there should be no worse detection results than to those of the traditional covariance estimators.

We evaluate our estimators for hyperspectral anomaly detection using the Kelly anomaly detector [91]. More precisely, we first perform a thorough evaluation of our estimators using some Monte-Carlo simulations for three true covariance models of different sparsity levels. From our experiments in section 6.3.2, the detection results show that in trully sparse models, our estimators improve the detection significantly with respect to the traditional ones, and have competitive results with state-of-the-art [13, 29, 127]. When the true model is not sparse, we find that empirically our estimators still have no worse detection results than to those of $\hat{\mathbf{\Sigma}}_{SCM}$, $\hat{\mathbf{\Sigma}}_{OLS}$ and $\hat{\mathbf{\Sigma}}_{FP}$.
Next, in section 6.3.3, our estimators are evaluated on experimental data where a good target detection performances are obtained compared to the traditional estimators and state-of-the-art.

6.2 Our main contributions

6.2.1 Generalized thresholding based Cholesky Factor

For any $0 \leq \omega \leq 1$, we define a matrix thresholding operator $Th(.)$ and denote by $Th(\hat{\mathbf{T}}_{OLS}) = [Th(-\hat{C}_{t,j}^{OLS})]_{p\times p}$ the matrix resulting from applying a specific thresholding operator $Th(.) \in \{$Soft, SCAD$\}$ to each element of the matrix $\hat{\mathbf{T}}^{OLS}$ for $t \in [2,\, p]$ and $j \in [1,\, t-1]$.
Similarly to (5.22), we consider the following minimization problem:

$$\begin{aligned} Th(\hat{\mathbf{T}}_{OLS}) &= \underset{\mathbf{T}}{\operatorname{argmin}} \frac{1}{2} \left\| \mathbf{T} - \hat{\mathbf{T}}_{OLS} \right\|_F^2 + \sum_{t=2}^{p} \sum_{j=1}^{t-1} p_\omega\{|C_{t,j}|\}\,, \\ &= \underset{\mathbf{T}}{\operatorname{argmin}} \sum_{t=2}^{p} \sum_{j=1}^{t-1} \left\{ \frac{1}{2} \left(\hat{C}_{t,j}^{OLS} - C_{t,j} \right)^2 + p_\omega\{|C_{t,j}|\} \right\} \end{aligned} \tag{6.1}$$

where $p_\omega \in \{p_\omega^{l_1}, p_{\omega,a>2}^{SCAD}\}$.

Solving (6.1) with $p_\omega^{l_1}$ and $p_{\omega,a>2}^{SCAD}$, yields a closed-form Soft and SCAD thresholding rules, respectively [127], [53]. The value $a = 3.7$ was recommended by Fan and Li [53]. Despite the application here is different than in [53], for simplicity, we use the same value $a = 3.7$.

We shall designate the two thresholded matrices by $\hat{\mathbf{T}}_{Soft}$ and $\hat{\mathbf{T}}_{SCAD}$, that apply Soft and SCAD on $\hat{\mathbf{T}}_{OLS}$, respectively. We denote our first two estimators as:

$$\boxed{\hat{\mathbf{\Sigma}}_{OLS}^{Soft} = \hat{\mathbf{T}}_{Soft}^{-1}\, \hat{\mathbf{D}}_{OLS}\, \hat{\mathbf{T}}_{Soft}^{-T}}$$

$$\boxed{\hat{\mathbf{\Sigma}}_{OLS}^{SCAD} = \hat{\mathbf{T}}_{SCAD}^{-1}\, \hat{\mathbf{D}}_{OLS} \hat{\mathbf{T}}_{SCAD}^{-T}}$$

Advantages of our estimators $\hat{\mathbf{\Sigma}}_{OLS}^{Soft}$ and $\hat{\mathbf{\Sigma}}_{OLS}^{SCAD}$ over the work in [127] (that is, the work in section 5.3.2 of Chapter 5):
Note that in [127], the authors have demonstrated that for a non sparse true covariance model, the covariance matrix does not suffer any degradation when thresholding is applied to the off-diagonal entries of $\hat{\mathbf{\Sigma}}_{SCM}$. However, this is not the case for the target detection problem where the inverse covariance is used; we found that, and in contrast to our estimators, the scheme in [127] has a deleterious effect on the detection performance when compared to those of $\hat{\mathbf{\Sigma}}_{SCM}$, $\hat{\mathbf{\Sigma}}_{OLS}$ and $\hat{\mathbf{\Sigma}}_{FP}$.

6.2.2 A generalization of the estimator in [82]

We continue the work in section 5.3.3 (the work in [82]), but by modifying the procedure by which the entries of $\mathbf{T}$ have been estimated.
By denoting $l(\boldsymbol{\beta}_t) = \frac{1}{\theta_t^2} \left\| \mathbf{y}_t - \mathbf{A}_{n,t}\boldsymbol{\beta}_t \right\|_F^2$ and $r(\boldsymbol{\beta}_t) = \sum\limits_{j=1}^{t-1} p_\varphi\{|C_{t,j}|\} = \sum\limits_{j=1}^{t-1} r_j(C_{t,j})$, we solve $\boldsymbol{\beta}_t$

iteratively using the General Iterative Shrinkage and Thresholding (GIST) algorithm [70]:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_t^{(k)} &= \underset{\boldsymbol{\beta}_t}{\operatorname{argmin}}\; l\left(\boldsymbol{\beta}_t^{(k-1)}\right) + r(\boldsymbol{\beta}_t) + \left(\nabla l\left(\boldsymbol{\beta}_t^{(k-1)}\right)\right)^T \left(\boldsymbol{\beta}_t - \boldsymbol{\beta}_t^{(k-1)}\right) + \frac{w^{(k-1)}}{2}\left\|\boldsymbol{\beta}_t - \boldsymbol{\beta}_t^{(k-1)}\right\|^2, \\ &= \underset{\boldsymbol{\beta}_t}{\operatorname{argmin}}\; \frac{1}{2}\left\|\boldsymbol{\beta}_t - \mathbf{u}_t^{(k-1)}\right\|^2 + \frac{1}{w^{(k-1)}}\, r(\boldsymbol{\beta}_t)\,. \end{aligned} \tag{6.2}$$

where $\mathbf{u}_t^{(k-1)} = \boldsymbol{\beta}_t^{(k-1)} - \nabla l\left(\boldsymbol{\beta}_t^{(k-1)}\right)/w^{(k-1)}$, and $w^{(k-1)}$ is the step size initialized using the Barzilai-Browein rule [6]. Recall that $\varphi \in [0, \infty)$ (see section 5.3.3 in Chapter 5).

By decomposing (6.2) into independent (t-1) univariate optimization problems, we have for $j = 1, \cdots, t-1$:

$$C_{t,j}^{(k)} = \underset{C_{t,j}}{\operatorname{argmin}}\; \frac{1}{2}\left\|C_{t,j} - u_{t,j}^{(k-1)}\right\|^2 + \frac{1}{w^{(k-1)}}\, r_j(C_{t,j})\,, \tag{6.3}$$

where $\mathbf{u}_t^{(k-1)} = \left[u_{t,1}^{(k-1)}, \cdots, u_{t,t-1}^{(k-1)}\right]^T \in \mathbb{R}^{(t-1)}$.
By solving (6.3) with the l_1-norm penalty, $p_\varphi^{l_1}$, we have the following closed form solution:

$$C_{t,j,(l_1)}^{(k)} = \operatorname{sgn}\left(u_{t,j}^{(k-1)}\right) \max\left(0,\, \left|u_{t,j}^{(k-1)}\right| - \varphi/w^{(k-1)}\right). \tag{6.4}$$

For the SCAD penalty function, $p_{\varphi,a>2}^{SCAD}$, we can observe that it contains three parts for three different conditions (see Subsection 5.1.7 in Chapter 5). In this case, by recasting problem (6.3) into three minimization sub-problems for each condition, and after solving them, one can obtain the following three sub-solutions $h_{t,j}^1$, $h_{t,j}^2$, and $h_{t,j}^3$, where:

$$\begin{aligned} h_{t,j}^1 &= \operatorname{sgn}\left(u_{t,j}^{(k-1)}\right) \min\left(\varphi, \max\left(0, \left|u_{t,j}^{(k-1)}\right| - \varphi/w^{(k-1)}\right)\right), \\ h_{t,j}^2 &= \operatorname{sgn}\left(u_{t,j}^{(k-1)}\right) \min\left(a\,\varphi, \max\left(\varphi, \frac{w^{(k-1)}\left|u_{t,j}^{(k-1)}\right|(a-1) - a\,\varphi}{w^{(k-1)}(a-2)}\right)\right), \\ h_{t,j}^3 &= \operatorname{sgn}\left(u_{t,j}^{(k-1)}\right) \max\left(a\,\varphi, \left|u_{t,j}^{(k-1)}\right|\right). \end{aligned}$$

Hence, we have the following closed form solution:

$$\begin{aligned} C_{t,j,(SCAD)}^{(k)} = \underset{q_{t,j}}{\operatorname{argmin}}\; & \frac{1}{2}\left(q_{t,j} - u_{t,j}^{(k-1)}\right)^2 + \frac{1}{w^{(k-1)}}\, r_j(q_{t,j}) \\ & s.t. \quad q_{t,j} \in \{h_{t,j}^1,\, h_{t,j}^2,\, h_{t,j}^3\} \end{aligned} \tag{6.5}$$

At the end, we denote our last two estimators as:

$$\boxed{\hat{\boldsymbol{\Sigma}}_{l_1} = \hat{\mathbf{T}}_{l_1}^{-1}\, \hat{\mathbf{D}}\, \hat{\mathbf{T}}_{l_1}^{-T}}$$

$$\boxed{\hat{\boldsymbol{\Sigma}}_{SCAD} = \hat{\mathbf{T}}_{SCAD}^{-1}\, \hat{\mathbf{D}}\, \hat{\mathbf{T}}_{SCAD}^{-T}}$$

where $\hat{\mathbf{T}}_{l_1}$ and $\hat{\mathbf{T}}_{SCAD}$ have respectively $-\hat{C}_{t,j,(l_1)}$ and $-\hat{C}_{t,j,(SCAD)}$ in the (t,j)th position for $t \in [2,\, p]$ and $j \in [1,\, t-1]$, whereas $\hat{\mathbf{D}}$ has the entries $(\hat{\theta}_1^2,\, \hat{\theta}_t^2)$ on its diagonal. Note that in [82],

the authors have used the local quadratic approximation (LQA) [53] of the l_1-norm in order to get a closed form solution for $\boldsymbol{\beta}_t$ in equation (5.26) in Chapter 5. Our algorithm is now more general since after exploiting the GIST algorithm to solve (5.26), it can be easily extended to some other penalties such as SCAD [53] , Capped-l_1 penalty [69, 161, 162], Log Sum Penalty[27], and Minimax Concave Penalty [159], and they all have closed-form solutions [70]. In this dissertation, we are only interested to the l_1 and SCAD penalty functions.

6.2.3 Tuning selection of ω (in section 6.2.1) and φ (in section 6.2.2)

In this section, we briefly discuss how the parameters ω and φ can be automatically selected. The standard way to select them is to minimize the following risk:

$$R(\omega \text{ or } \varphi) = \mathbb{E} \left\| \hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma} \right\|_2 ,$$

where $\hat{\boldsymbol{\Sigma}} \in \left\{ \hat{\boldsymbol{\Sigma}}_{OLS}^{Soft}, \hat{\boldsymbol{\Sigma}}_{OLS}^{SCAD}, \hat{\boldsymbol{\Sigma}}_{l_1}, \hat{\boldsymbol{\Sigma}}_{SCAD} \right\}$. The oracle values of ω and φ are chosen to minimize the above risk. Note that the choice of the parameter norm in the above risk is somewhat arbitrary, and we have considered the l_2 norm.

In real hyperspectral applications, the covariance matrix model is completely not known, and thus, has to be estimated. In this case, one needs to find an approximation for the above risk, and hence, to estimate the oracle values of the parameters ω and φ (since these parameters control the amount of sparsity and selecting good values is crucial). By following section 4.1 in [82], the oracle values of ω and φ can be approximated by the following cross-validation procedure:

1. Randomly split the matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n] \in \mathbb{R}^{p \times n}$ into K parts or folds of roughly equal size, $\mathbf{A}_1, \cdots, \mathbf{A}_K$,

2. For each $v \in \{1, \cdots, K\}$:

   Let $\mathbf{A}_v \in \mathbb{R}^{p \times (\frac{n}{K})}$ be its corresponding fold used for validation and $\mathbf{A}_{-v} \in \mathbb{R}^{p \times \left(n-(\frac{n}{K})\right)}$ for training. By $\mathbf{A}_{-v}$, we mean the data matrix $\mathbf{X}$ excluding the fold $\mathbf{A}_v$.

   The parameters ω and φ are continous, but considering all possible values will increase the computational complexity dramatically. That is why, it is more practically to choose to discretize the range of both ω and φ and consider them over some discrete set.
   For each value of ω and φ in the discrete set, compute the error RMS:

$$RMS(v, \omega \text{ or } \varphi) = s_v \log\det\left(\hat{\boldsymbol{\Sigma}}_{-v}\right) + \sum_{i \in I_v} \mathbf{x}_i^T \, \hat{\boldsymbol{\Sigma}}_{-v}^{-1} \, \mathbf{x}_i \,, \tag{6.6}$$

   where I_v is the index set of the data in $\mathbf{A}_v$, s_v denotes the size of $\mathbf{A}_v$, and $\hat{\boldsymbol{\Sigma}}_{-v}$ is computed only on the training set $\mathbf{A}_{-v}$.

**Importantly, Rothman et al. in [127] (that is, in section 5.3.2 in Chapter 5) have selected the thresholding parameter φ for the thresholded $\hat{\boldsymbol{\Sigma}}_{SCM}$ (in equation (5.22)) by the same

cross validation method, but equation (6.6) was replaced by: $\left\|\hat{\mathbf{\Sigma}}_{SCM,\,-v}^{Soft \text{ or } SCAD} - \hat{\mathbf{\Sigma}}_{SCM,\,v}\right\|_F^2$, where $\hat{\mathbf{\Sigma}}_{SCM,\,v}$ represents $\hat{\mathbf{\Sigma}}_{SCM}$ that is computed on the dataset A_v (the validation set) only. For our proposed estimators, we can also think to replace equation (6.6) by $\left\|\hat{\mathbf{\Sigma}}_{-v} - \hat{\mathbf{\Sigma}}_{OLS,\,v}\right\|_F^2$, where $\hat{\mathbf{\Sigma}}_{OLS,\,v}$ represents $\hat{\mathbf{\Sigma}}_{OLS}$ but computed on the dataset A_k. It is interesting to note that the size of the dataset $\mathbf{A}_v$ is usually lower than p, and hence, $\hat{\mathbf{\Sigma}}_{OLS,\,v}$ will be undefined.

3. For each ω or φ value, compute the average error over all folds v, that is:

$$CV(\omega \text{ or } \varphi) = \frac{1}{K}\sum_{v=1}^{K} RMS(v, \omega \text{ or } \varphi)$$

We choose $\omega = \hat{\omega}$ and $\varphi = \hat{\varphi}$ to minimize $CV(\omega)$ and $CV(\varphi)$, respectively.

In order to achieve a good bias-variance trade-off of the average error estimates, the values K=5 and K=10 are preferred in practice. However, it is important to avoid choosing very high values of K (i.e., K=n) and very small values (i.e., K=2). The first (K=n) corresponds to the Leave-one-out cross validation and in which the average error estimates are going to have high variance, whereas the second (K=2) corresponds to the Split-Sample cross validation and in which the average error estimates are going to be biased upwards.

6.3 Experiments and Analysis

Some Monte-Carlo simulations as well as experimental data evaluate our proposed covariance estimators $\left\{\hat{\mathbf{\Sigma}}_{OLS}^{Soft}, \hat{\mathbf{\Sigma}}_{OLS}^{SCAD}, \hat{\mathbf{\Sigma}}_{l_1}, \hat{\mathbf{\Sigma}}_{SCAD}\right\}$ for hyperspectral target detection using the Kelly anomaly detector [91].

Our estimators are compared to the traditional ones $\hat{\mathbf{\Sigma}}_{SCM}$, $\hat{\mathbf{\Sigma}}_{OLS}$, $\hat{\mathbf{\Sigma}}_{FP}$ and state-of-the-art: $\hat{\mathbf{\Sigma}}_{SMT}$ [29], $B_k\left(\hat{\mathbf{\Sigma}}_{SCM}\right)$ [13], and the two estimators that apply Soft and SCAD thresholding on the off-diagonal entries of $\hat{\mathbf{\Sigma}}_{SCM}$ in [127], and which are denoted in the following experiments as $\hat{\mathbf{\Sigma}}_{SCM}^{Soft}$ and $\hat{\mathbf{\Sigma}}_{SCM}^{SCAD}$, respectively (see section 5.3.2 in Chapter 5).

6.3.1 A brief overview of the Kelly anomaly detector

Suppose the following signal model:

$$\begin{cases} H_0 : \mathbf{x} = \mathbf{n}, & \mathbf{x}_i = \mathbf{n}_i, \quad i = 1, \cdots, n \\ H_1 : \mathbf{x} = \delta\,\mathbf{t} + \mathbf{n}, & \mathbf{x}_i = \mathbf{n}_i, \quad i = 1, \cdots, n \end{cases} \tag{6.7}$$

where $\mathbf{n}_1, \cdots, \mathbf{n}_n$ are n i.i.d p-vectors having a multivariate normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$. $\mathbf{t}$ is an unknown steering vector which denotes the presence of an anomalous signal with unknown amplitude $\delta > 0$. After some calculation (refer to [91] and both Subsection II. B and Remark II. 1 in [58] for details), the Kelly anomaly detector is described as follows:

$$D_{KellyAD\hat{\mathbf{\Sigma}}}(\mathbf{x}) = \mathbf{x}^T\,\hat{\mathbf{\Sigma}}_{SCM}^{-1}\,\mathbf{x} \underset{H_0}{\overset{H_1}{\gtrless}} \eta\,, \tag{6.8}$$

where η is a prescribed threshold value.

In the following two subsections, both parameters ω and φ has been selected using the cross-validation method described in section 6.2.3.

The detection performances of the estimators, when are plugged in $D_{KellyAD,\hat{\mathbf{\Sigma}}}$ are evaluated by the Receiver Operating Characteristics (ROC) curves (which describe the probability of detection P_d against the probability of false alarms P_{fa}) and their corresponding Area Under Curves (AUC) values.

6.3.2 Monte-Carlo simulations

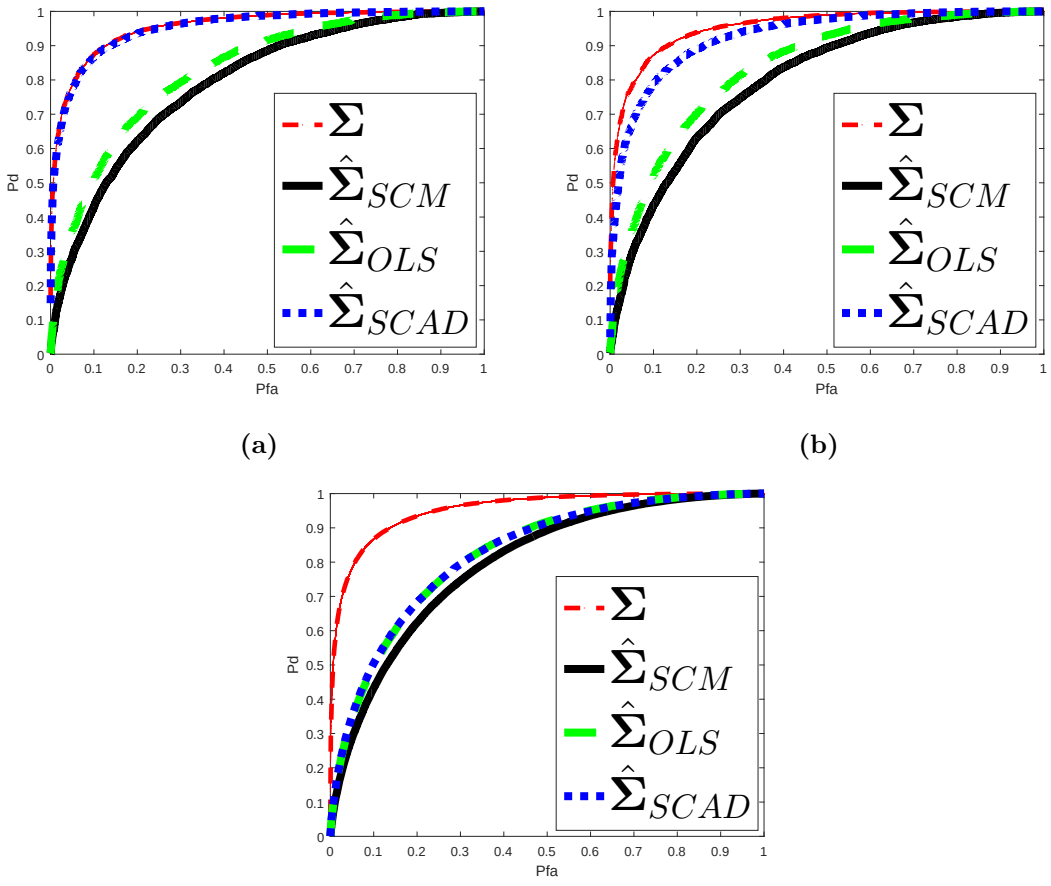
The experiments are conducted on three covariance models:

- Model 1: $\mathbf{\Sigma} = \mathbf{I}$, the identity matrix,
- Model 2: the autoregressive model order 1, AR(1), $\mathbf{\Sigma} = [\sigma_{gl}]_{p\times p}$, where $\sigma_{gl} = c^{|g-l|}$, for $c = 0.3$,
- Model 3: $\Sigma = [\sigma_{gl}]_{p\times p}$, where $\sigma_{gl} = (1 - ((|g-l|)/r))_+$, for $r = p/2$: the triangular matrix.

Model 1 is very sparse and Model 2 is approximately sparse. Model 3 with $r = p/2$ is considered the least sparse [127] among the three models.

The computations have been made through 10^5 Monte-Carlo trials and the ROC curves (P_d - P_{fa}) are drawn for a signal to noise ratio equal to 15dB. We choose $n = 80$ for covariance estimation under a GAUSSIAN assumption (that is, $\nu = \infty$), and set $p = 60$. The artificial anomaly we consider is a vector containing normally distributed pseudorandom numbers (to have fair results, the same vector is used for the three covariance models). The ROC curves for Model 1, 2 and 3 are shown in Figure. 6.1**(a)**, 6.1**(b)** and 6.1**(c)**, respectively, and their corresponding Area Under Curves (AUC) values are presented in Table 1. Note that in Figure 6.1**(a)**, 6.1**(b)** and 6.1**(c)**, we only exhibit the ROC curves for $\mathbf{\Sigma}$, $\hat{\mathbf{\Sigma}}_{SCM}$, $\hat{\mathbf{\Sigma}}_{OLS}$, and $\hat{\mathbf{\Sigma}}_{SCAD}$. For the other estimators, their AUC values are presented in Table 1.

Our findings: For both Model 1 and 2, our estimators significantly improve the detection performances comparing to those of the traditional estimators ($\hat{\mathbf{\Sigma}}_{SCM}$, $\hat{\mathbf{\Sigma}}_{OLS}$, $\hat{\mathbf{\Sigma}}_{FP}$), and have competitive detection results with state-of-the-art. An important finding is that even for a non sparse covariance model (that is, Model 3) , our estimators do not show a harm on the detection when compared to those of $\hat{\mathbf{\Sigma}}_{SCM}$, $\hat{\mathbf{\Sigma}}_{OLS}$ and $\hat{\mathbf{\Sigma}}_{FP}$. Although that $\hat{\mathbf{\Sigma}}_{OLS}^{Soft}$, $\hat{\mathbf{\Sigma}}_{OLS}^{SCAD}$ and $\hat{\mathbf{\Sigma}}_{l_1}$ have slightly lower AUC values than $\hat{\mathbf{\Sigma}}_{OLS}$, this is still a negligible degradation on the detection. Thus, considering that $\hat{\mathbf{\Sigma}}_{OLS}^{Soft}$, $\hat{\mathbf{\Sigma}}_{OLS}^{SCAD}$ and $\hat{\mathbf{\Sigma}}_{l_1}$ have no worse detection results than $\hat{\mathbf{\Sigma}}_{OLS}$ is still acceptable.



(a) **(b)**

(c)

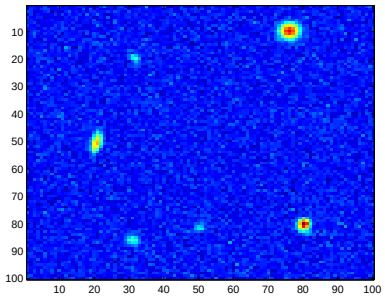
Figure 6.1: ROC curves for the three Models. **(a)**: Model 1. **(b)**: Model 2. **(c)**: Model 3.

6.3.3 Real experiments

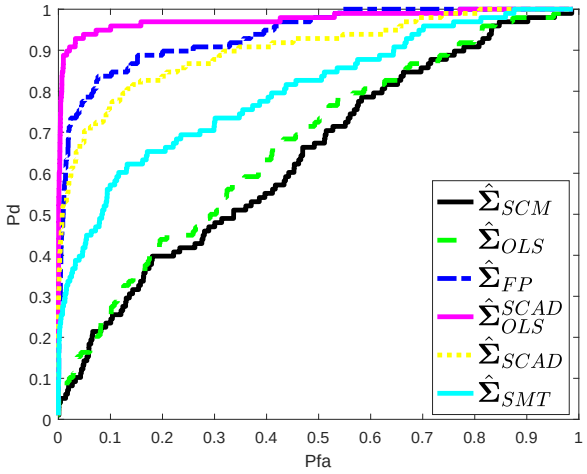
Our estimators are now evaluated for galaxy detection on the Multi Unit Spectroscopic Explorer (MUSE) data cube (see [111]). It is a 100 × 100 image and consists of 3600 bands in wavelengths ranging from 465-930 nm. We used one band of each 60, so that 60 bands in total. Figure 6.4**(a)** exhibits the mean power in dB over the 60 bands. The covariance matrix is estimated using a sliding window of size 9×9, having $n = 80$ secondary data (after excluding only the test pixel). The mean has been removed from the given HSI. Figure 6.2**(b)** exhibits the ROC curves [163] of our estimators when compared to some others, and their AUC values are shown in Table 1.
The estimators $\hat{\mathbf{\Sigma}}_{OLS}^{Soft}$, $\hat{\mathbf{\Sigma}}_{OLS}^{SCAD}$ achieve higher detection results than to those of all the others (especially to that of the Fixed-Point $\hat{\mathbf{\Sigma}}_{FP}$), whereas both $\hat{\mathbf{\Sigma}}_{l_1}$ and $\hat{\mathbf{\Sigma}}_{SCAD}$ achieve only a lower AUC values than to that of $B_k\left(\hat{\mathbf{\Sigma}}_{SCM}\right)$ and $\hat{\mathbf{\Sigma}}_{FP}$.

Covariances	Model 1	Model 2	Model 3	MUSE
$\boldsymbol{\Sigma}$	0.9541	0.9540	0.9541	Not known
$\hat{\boldsymbol{\Sigma}}_{SCM}$	0.7976	0.7977	0.7978	0.6277
$\hat{\boldsymbol{\Sigma}}_{OLS}$	0.8331	0.8361	0.8259	0.6575
$\hat{\boldsymbol{\Sigma}}_{FP}$	0.7941	0.7942	0.7876	0.9269
$\hat{\boldsymbol{\Sigma}}_{OLS}^{Soft}$	0.9480	0.9124	0.8169	0.9620
$\hat{\boldsymbol{\Sigma}}_{OLS}^{SCAD}$	0.9480	0.9124	0.8257	0.9643
$\hat{\boldsymbol{\Sigma}}_{l_1}$	0.9509	0.9264	0.8236	0.8844
$\hat{\boldsymbol{\Sigma}}_{SCAD}$	0.9509	0.9264	0.8261	0.8844
$\hat{\boldsymbol{\Sigma}}_{SMT}$	0.9503	0.9184	0.7798	0.7879
$B_k(\hat{\boldsymbol{\Sigma}}_{SCM})$	0.9509	0.9478	0.5321	0.9277
$\hat{\boldsymbol{\Sigma}}_{SCM}^{Soft}$	0.9509	0.9274	0.5969	0.7180
$\hat{\boldsymbol{\Sigma}}_{SCM}^{SCAD}$	0.9509	0.9270	0.5781	0.7180

Table 1. A List of Area Under Curve (AUC) values of our estimators $\hat{\boldsymbol{\Sigma}}_{OLS}^{Soft}$, $\hat{\boldsymbol{\Sigma}}_{OLS}^{SCAD}$, $\hat{\boldsymbol{\Sigma}}_{l_1}$, $\hat{\boldsymbol{\Sigma}}_{SCAD}$ when compared to some others.



(a)



(b)

Figure 6.2: **(a)** MUSE HSI (average). **(b)** ROC curves for MUSE

6.4 Summary of the whole Part III and some future directions

6.4.1 Summary

In hyperspectral, the covariance matrix $\boldsymbol{\Sigma}$ (of the background surrounding the test pixel) is completely not known, and thus, has to be carefully estimated especially in large dimensions. In Chapter 5, an interesting comparative study has been done on three traditional covariance estimators: $\hat{\boldsymbol{\Sigma}}_{SCM}$, $\hat{\boldsymbol{\Sigma}}_{OLS}$ and $\hat{\boldsymbol{\Sigma}}_{FP}$. It has been shown that $\hat{\boldsymbol{\Sigma}}_{OLS}$ always achieve slightly higher target detection results than $\hat{\boldsymbol{\Sigma}}_{SCM}$. For a Gaussian model assumption, $\hat{\boldsymbol{\Sigma}}_{FP}$ always achieve slightly lower target detection results than $\hat{\boldsymbol{\Sigma}}_{SCM}$; whereas for a non-Gaussian model assumption, $\hat{\boldsymbol{\Sigma}}_{FP}$ is robust, in contrast to both $\hat{\boldsymbol{\Sigma}}_{SCM}$ and $\hat{\boldsymbol{\Sigma}}_{OLS}$. When the spectral dimension

p is considered large compared to the sample size n, $\hat{\mathbf{\Sigma}}_{SCM}$, $\hat{\mathbf{\Sigma}}_{OLS}$ and $\hat{\mathbf{\Sigma}}_{FP}$ behave very poorly. This is due to the fact that these traditional covariance estimators are ideal only when n tends to ∞.

Usually, compounding the large dimensionality problem can be alleviated by leveraging on the assumption that the true unknown covariance matrix is sparse, namely, many entries are zero. In this regard, Chapter 6 has outlined two methods to impose sparsity on the covariance matrix via its unit lower triangular matrix (aka Cholesky factor) $\mathbf{T}$. The first method serves to estimate the entries of $\mathbf{T}$ using the Ordinary Least Squares (OLS), then imposes sparsity by exploiting some generalized thresholding techniques such as Soft and Smoothly Clipped Absolute Deviation (SCAD). The second method directly estimates a sparse version of $\mathbf{T}$ by penalizing the negative normal log-likelihood of the matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n]$ with l_1 and SCAD penalty functions. More precisely, we aimed to generalize the work in [82] in order to be used for various penalty functions. Our proposed sparse covariance estimators are always guaranteed to be positive definite.

Our findings show that in trully sparse covariance models, our proposed estimators potentially improve the target detection performances compared to those of the traditional estimators, and have competitive detection results with state-of-the-art. Note that when the true covariance is not sparse (or not highly sparse), our proposed covariance estimators do not achieve worse detection results than to those obtained with the traditional covariance estimators.

6.4.2 Some directions for future work

As for future enhancements:;

- We will evaluate our proposed covariance estimators on more experimental datasets,
- Extend the proposed methods to the case of "curse of dimensionality", that is, when $p \gg n$,
- Our Monte-Carlo simulations have been done only on the Gaussian case (for $\nu = \infty$). The question now is "*Are our findings the same in a non-Gaussian case (that is, for small values of ν)?*". In the future, the evaluation of the robustness of our proposed estimators will be of our interest.

Part IV

Concluding remarks

"I may not have gone where I intended to go, but I think I have ended up where I needed to be.

— Douglas Adams

7 Conclusion

Contents

► *Synopsis* Finally, the dissertation concludes with this chapter, which summarizes the main contributions and raises several additional research questions. It also provides some of our additional works that still need to be tested in the future.

7.1 Looking Back

In hyperspectral, the main challenges lie in large spectral dimensionality, and data variation modelling due to material spectral variability, atmospheric effects, and sensor noise. This dissertation has concentrated on alleviating the aforementioned challenges by proposing new techniques and detectors. The proposed works have been grouped into two different directions:

7.1.1 First thesis direction ($PartII$)

We have exploited the well-known Robust Principal Component Analysis (RPCA) for target detection in hyperspectral imagery. By taking similar assumptions to those used in RPCA, a given hyperspectral image (HSI) has been decomposed into the sum of a low-rank background HSI (denoted as $\mathbf{L}$) consisting only of the background without the targets, and a sparse target HSI (denoted as $\mathbf{E}$) that only contains the targets (with the background is suppressed). After

evaluating the RPCA on several real hyperspectral images, we proved that the direct use of RPCA is inadequate to distinghuising the true targets from the surrounding background. In particular, we found that RPCA searches for small heterogeneous and high contrast regions that may not necessarily be the true targets of interest.

In order to potentially alleviate this problem, we have incorporated into the RPCA imaging, the prior target information that can often be provided to the user. In hyperspectral, this prior information is usually related to the target spectral signature which can be already measured by some laboratories and present in some online spectral libraries. In this regard, we have constructed a pre-learned target dictionary $\mathbf{A}_t$, and the given HSI is being decomposed as the sum of a low-rank background HSI $\mathbf{L}$ and a sparse target HSI denoted as $(\mathbf{A}_t\mathbf{C})^T$, where $\mathbf{C}$ is a sparse activation matrix. Following our proposed modification of RPCA, two detection strategies were available to us to realize the target detection:

First detection strategy: the sparse component $(\mathbf{A}_t\mathbf{C})^T$ is the object of interest

We use the sparse image $(\mathbf{A}_t\mathbf{C})^T$ directly as a detector. Note that for this scheme to work, we require as few false alarms as possible to be deposited in the target image, but we do not need the target fraction to be entirely removed from the background. As long as enough of the target fractions are moved to the target image such that non-zero support is detected at the corresponding pixel location, it will be adequate for our detection scheme.

Second detection strategy: the low-rank component ($\mathbf{L}$) is the object of interest

We use the background HSI $\mathbf{L}$ for a more accurate construction of the background dictionary $\mathbf{A}_b$. For each test pixel in the original HSI, we create a concentric window of size $m \times m$ on the background HSI $\mathbf{L}$, and all the pixels within the window (except the center pixel) will each contribute to one column in $\mathbf{A}_b$. Note that this concentric window amounts to an OWR of size $m \times m$ with IWR of size 1×1. Next, we make use of the SRBBH detector, but with the background dictionary $\mathbf{A}_b$ constructed in the preceding manner. Note that for this scheme to work, we do not need a clean separation (by clean separation, we mean that all the targets are present in the sparse image with no false alarms); specifically, we require the entire target fraction to be separated from the background and deposited in the target image, but some of the background objects can also be deposited in the target image. As long as enough signatures of these background objects remain in the background HSI $\mathbf{L}$, the $\mathbf{A}_b$ constructed will be adequately representative of the background.

Evaluation performances of the two detection strategies

Both detection strategies are evaluated independently to each other on both synthetic as well as real experiments, and the results of which demonstrate their effectiveness for hyperspectral target detection. In particular, they can deal with targets of any shapes or targets that occur in close

proximity, and are resilient to most values of fill-fractions unless they are too small.

7.1.2 Second thesis direction ($PartIII$)

When the spectral dimension p is considered large compared to the sample size n, the traditional covariance estimators behave poorly. In many studies, the researchers assume that compounding the large dimensionality problem can be alleviated by leveraging on the assumption that the true unknown covariance matrix (of the background surrounding the test pixel) $\mathbf{\Sigma}$ is sparse, namely, many entries are zero. In this thesis direction, the covariance matrix $\mathbf{\Sigma}$ is first regarded as being made up of $\mathbf{\Sigma} = \mathbf{T}^{-1}\,\mathbf{D}\,\mathbf{T}^{-T}$, where $\mathbf{T}$ is a unit lower triangular matrix (aka Cholesky factor) and $\mathbf{D}$ is a diagonal matrix with positive entries. Then, two kinds of sparsity are imposed for large covariance matrices through their Cholesky factor $\mathbf{T}$. The first method aims to estimate the entries of $\mathbf{T}$ using such a method as the ordinary least squares (OLS), and then imposes sparsity by exploiting some of the generalized thresholding techniques such as Soft and Smoothly Clipped Absolute Deviation (SCAD). The second aims to directly estimate a sparse version of $\mathbf{T}$ by proposing a generalized version of the estimator already developed in [82]. The resulting estimated sparse covariance matrices are always guaranteed to be positive definite.

Evaluation performances: some important findings

Enforcing sparsity on the covariance matrix seems to be a strong assumption, but can be critically important if the true covariance model for a given HSI is indeed sparse or approximately sparse. We have demonstrated that taking advantage of the possible sparsity in the estimation can potentially improve the target detection performance. On the other hand, if the true covariance model is not sparse (or not highly sparse), our estimators do not have worse detection results than to those of the traditional ones.

7.2 Looking Ahead

Throughout this PhD thesis, new opportunities of investigation presented themselves as areas of future research. These opportunities include:

1. **Changing the $l_{2,1}$ norm in problem** (2.11): we have already mentioned in part 4.2 of Chapter 4, that there is a need to change the $l_{2,1}$ norm in (2.11) by another one closer to the ideal $l_{2,0}$ norm in order to probably alleviate the manual selection problem for the parameters.

2. **Extension of the work in [163] by taking into consideration the spatial correlation in hyperspectral imagery (see [33])**: the fact is that in hyperspectral imagery, neighboring pixels have a similar spectral characteristics. This property has not been exploited in the work of Zhang *et al.* in [163]. Hence, it will be important to exlploit this fact (more precisely, the work of Chen *et al.* in [33]) into the work of [163].

7.2.1 Changing the $l_{2,1}$ norm in problem (2.11)

Recall that our proposed modified version of RPCA in Chapter 2 was (see problem (2.10)):

$$\min_{\mathbf{L},\mathbf{C}} \left\{ \tau \operatorname{rank}(\mathbf{L}) + \lambda \left\| \mathbf{C} \right\|_{2,0} + \left\| \mathbf{D} - \mathbf{L} - (\mathbf{A}_t \, \mathbf{C})^T \right\|_F^2 \right\}.$$

Since problem (2.10) is NP-HARD to solve due to the rank term and $l_{2,0}$ norm, hence, we used the nuclear norm $||\mathbf{L}||_*$ as a surrogate for the $rank(\mathbf{L})$ term, and the $l_{2,1}$ norm for the $l_{2,0}$ norm, as can be seen in problem (2.11)
The l_1-norm regularizer, a continuous and convex surrogate, has been studied extensively in the literature [50, 139] and has been applied successfully to many applications including signal/image processing, biomedical informatics and computer vision [8, 129, 148, 150, 158]. Although the l_1 norm based sparse learning formulations have achieved great success, they have been shown to be suboptimal in many cases [28, 161, 162], since the l_1 is still too far away from the ideal l_0 norm. To address this issue, many non-convex regularizers, interpolated between the l_0 norm and l_1 norm, have been proposed to better approximate the l_0 norm. They include l_q norm $(0 < q < 1)$ [56], Smoothly Clipped Absolute Deviation [53], Log-Sum Penalty [27], Minimax Concave Penalty [159], Geman Penalty [64, 145], and Capped-l_1 penalty [69, 161, 162].

In this regard, in our problem (2.11), it will be interesting to use other proxies than the $l_{2,1}$ norm, closer to $l_{2,0}$, in order to probably alleviate the $l_{2,1}$ artifact and also the manual selection problem of both τ and λ. But although the non-convex regularizers (penalties) are appealing in sparse learning, it remains a big challenge to solve the corresponding non-convex optimization problems.

7.2.2 Extension of the work in [163] by taking into consideration the spatial correlation in hyperspectral imagery (see [33])

The SRBBH detector [163], that has been developed very recently, does not take into account any joint sparsity model that can incorporate the interpixel correlation within the HSI by assuming that neighboring pixels usually consist of similar materials and thus have the same spectral characteristics (this joint sparsity approach has been already developed in [33], but we just aim to exploit it into the work in [163]).

Hence, as our future enhancements, we attempt to develop the Simultaneous SRBBH (S-$SRBBH$) target detector that is similar to the $SRBBH$ in [163], but further considers a simultaneous joint sparsity model [33] that incorporates the spatial correlation that exists within neighboring pixels in HSI. More precisely, for each test pixel, all its neighbors within a small neighborhood can be simultaneously represented by a linear combination of a few common atom signals but weighted with a different set of coefficients for each pixel.

The Simultaneous SRBBH (S-$SRBBH$)

We aim to extend the sparsity model of Zhang *et al.* in [163] (which is briefly overviewed in section 3.1.1).

We define the matrix $\mathbf{X}_d = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_d] \in \mathbb{R}^{p\times d}$, where d is the total number of pixels in the neighborhood. Note that if $d = 1$, we return back to the SRBBH model. Hence, the S-SRBBH can be considered as a generalization of SRBBH when $d > 1$. Therefore if $\mathbf{x} \in H_0$, we have:

$$\begin{aligned}
\mathbf{x}_1 &= c_{1,1}\mathbf{a}_1^b + c_{1,2}\mathbf{a}_2^b + \cdots + c_{1,N_b}\mathbf{a}_{N_b}^b\,,\\
&\vdots\\
\mathbf{x}_d &= c_{d,1}\mathbf{a}_1^b + c_{d,2}\mathbf{a}_2^b + \cdots + c_{d,N_b}\mathbf{a}_{N_b}^b\,.
\end{aligned}$$

This implies:

$$\mathbf{X}_d = \left[\mathbf{a}_1^b\, \mathbf{a}_2^b \cdots \mathbf{a}_{N_b}^b\right]\, [\mathbf{c}_1\, \mathbf{c}_2 \cdots \mathbf{c}_d] = \mathbf{A}_b\mathbf{C}_b\,, \tag{7.1}$$

where $\mathbf{a}_1^b, \mathbf{a}_2^b \cdots, \mathbf{a}_{N_b}^b$ are the background samples, N_b is the total number of background samples, $\mathbf{A}_b \in \mathbb{R}^{p\times N_b}$, and $\mathbf{C}_b \in \mathbb{R}^{N_b\times d}$.

If $\mathbf{x} \in H_1$, we have:

$$\begin{aligned}
\mathbf{x}_1 = c'_{1,1}\mathbf{a}_1^b + c'_{1,2}\mathbf{a}_2^b + \cdots + c'_{1,N_b}\mathbf{a}_{N_b}^b\\
+z_{1,1}\mathbf{a}_1^t + z_{1,2}\mathbf{a}_2^t + \cdots + z_{1,N_t}\mathbf{a}_{N_t}^t\,,\\
\vdots\\
\mathbf{x}_d = c'_{d,1}\mathbf{a}_1^b + c'_{d,2}\mathbf{a}_2^b + \cdots + c'_{d,N_b}\mathbf{a}_{N_b}^b\\
+z_{d,1}\mathbf{a}_1^t + z_{d,2}\mathbf{a}_2^t + \cdots + z_{d,N_t}\mathbf{a}_{N_t}^t\,.
\end{aligned}$$

This implies:

$$\begin{aligned}\mathbf{X}_d &= \left[\mathbf{a}_1^b\, \mathbf{a}_2^b\, \cdots \mathbf{a}_{N_b}^b\right]\, [\mathbf{c'}_1\, \mathbf{c'}_2\, \cdots\, \mathbf{c'}_d] \\ &+ \left[\mathbf{a}_1^t\, \mathbf{a}_2^t\, \cdots \mathbf{a}_{N_t}^t\right]\, [\mathbf{z}_1\, \mathbf{z}_2\, \cdots\, \mathbf{z}_d] \\ &= [\mathbf{A}_b\, \mathbf{A}_t] \begin{pmatrix} \mathbf{C}'_b \\ \mathbf{Z}_t \end{pmatrix} = \mathbf{A}\, \mathbf{S}\,.\end{aligned} \tag{7.2}$$

where $\mathbf{a}_1^t, \mathbf{a}_2^t \cdots, \mathbf{a}_{N_t}^t$ are the target samples, N_t denotes the total number of target samples, $\mathbf{A}_t \in \Re^{p\times N_t}$, $\mathbf{A} \in \mathbb{R}^{p\times(N_b+N_t)}$ is the union of $\mathbf{A}_b$ and $\mathbf{A}_t$, $\mathbf{C}'_b \in \mathbb{R}^{N_b\times d}$, $\mathbf{Z}_t \in \mathbb{R}^{N_t\times d}$, and $\mathbf{S} \in \mathbb{R}^{(N_b+N_t)\times d}$.

Both $\mathbf{C}_b$ and $\mathbf{S}$ stand to be sparse, and therefore, the two (non-convex and NP-HARD) minimization subproblems to solve are:

$$\hat{\mathbf{C}}_b = \underset{\mathbf{C}_b}{\operatorname{argmin}} \left\|\mathbf{A}_b\mathbf{C}_b - \mathbf{X}_d\right\|_F \quad \text{s.t.} \quad \left\|\mathbf{C}_b\right\|_{0,2} \leq k_0\,, \tag{7.3a}$$

$$\hat{\mathbf{S}} = \underset{\mathbf{S}}{\operatorname{argmin}} \left\|\mathbf{A}\mathbf{S} - \mathbf{X}_d\right\|_F \quad \text{s.t.} \quad \left\|\mathbf{S}\right\|_{0,2} \leq k'_0\,, \tag{7.3b}$$

where $\left\|\mathbf{C}_b\right\|_{0,2}$ and $\left\|\mathbf{S}\right\|_{0,2}$ denote the total number of nonzero rows of $\mathbf{C}_b$ and $\mathbf{S}$, respectively. We can solve each subproblem using the Simultaneous Orthogonal Matching Pursuit (SOMP) [143] greedy algorithm. After that $\mathbf{C}_b$ and $\mathbf{S}$ are being estimated (to contain a few nonzero rows), the detection performance can be evaluated as:

$$D_{S-SRBBH}\left(\mathbf{x}\right) = \left\|\mathbf{X}_d - \mathbf{A}_b\hat{\mathbf{C}}_b\right\|_F - \left\|\mathbf{X}_d - \mathbf{A}\hat{\mathbf{S}}\right\|_F\,, \tag{7.4}$$

where $\hat{\mathbf{S}} = \begin{pmatrix} \hat{\mathbf{C}}'_b \\ \hat{\mathbf{Z}}_t \end{pmatrix}$. If $D_{S-SRBBH}\left(\mathbf{x}\right) > \eta$ with η is being a prescribed threshold value, then $\mathbf{x}$ is declared as target; otherwise, $\mathbf{x}$ will be labeled as background.

The first kind of intellectual and artistic personality belongs to the hedgehogs, the second to the foxes ...

— Sir Isaiah Berlin [10]

Bibliography

[1] C. Aicher. "A variational Bayes approach to robust principal component analysis". In: REU. 2013.

[2] Anestis Antoniadis and Jianqing Fan. "Regularization of Wavelet Approximations". In: *Journal of the American Statistical Association* 96.455 (2001), pp. 939–955. ISSN: 01621459. URL: http://www.jstor.org/stable/2670237.

[3] N. V. Aravind, K. Abhinandan, V. V. Acharya, and S. Sumam David. "Comparison of OMP and SOMP in the reconstruction of compressively sensed hyperspectral images". In: *2011 International Conference on Communications and Signal Processing*. Feb. 2011, pp. 188–192. DOI: 10.1109/ICCSP.2011.5739298.

[4] S. D. Babacan, M. Luessi, R. Molina, and A. K. Katsaggelos. "Sparse Bayesian Methods for Low-Rank Matrix Estimation". In: *IEEE Transactions on Signal Processing* 60.8 (Aug. 2012), pp. 3964–3977. ISSN: 1053-587X. DOI: 10.1109/TSP.2012.2197748.

[5] A.M. Baldridge, S.J. Hook, C.I. Grove, and G. Rivera. "The ASTER Spectral Library Version 2.0". In: *Remote Sensing of Environment* 113 (2009), pp. 711–715.

[6] J. Barzilai and J. M. Borwein. "Two-Point Step Size Gradient Methods". In: *IMA Journal of Numerical Analysis* 8.1 (1988), pp. 141–148. DOI: 10.1093/imanum/8.1.141. eprint: http://imajna.oxfordjournals.org/content/8/1/141.full.pdf+html.

[7] R. Basri and D. W. Jacobs. "Lambertian reflectance and linear subspaces". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25.2 (Feb. 2003), pp. 218–233. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2003.1177153.

[8] Amir Beck and Marc Teboulle. "A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems". In: *SIAM J. Img. Sci.* 2.1 (Mar. 2009), pp. 183–202. ISSN: 1936-4954. DOI: 10.1137/080716542. URL: http://dx.doi.org/10.1137/080716542.

[9] A. Berk, L. Bernstein, and D. Robertson. *MORTRAN: A moderate resolution model for LOWTRAN 7*. Tech. Rep. GL-TR-90-0122. G, 1989.

[10] Isaiah Berlin. *The Hedgehog and the Fox: An Essay on Tolstoy's View of History*. English. Ed. by Henry Hardy. 2nd. Princeton University Press, June 2013.

[11] O. Besson, S. Bidon, and J. Y. Tourneret. "Covariance Matrix Estimation With Heterogeneous Samples". In: *IEEE Transactions on Signal Processing* 56.3 (Mar. 2008), pp. 909–920. ISSN: 1053-587X. DOI: 10.1109/TSP.2007.908995.

[12] O. Besson, J. Y. Tourneret, and S. Bidon. "Knowledge-Aided Bayesian Detection in Heterogeneous Environments". In: *IEEE Signal Processing Letters* 14.5 (May 2007), pp. 355–358. ISSN: 1070-9908. DOI: 10.1109/LSP.2006.888088.

[13] P. J. Bickel and E. Levina. "Regularized Estimation of Large Covariance Matrices". In: *The Annals of Statistics* 36.1 (2008), pp. 199–227.

[14] S. Bidon, O. Besson, and J. Y. Tourneret. "A Bayesian Approach to Adaptive Detection in Nonhomogeneous Environments". In: *IEEE Transactions on Signal Processing* 56.1 (Jan. 2008), pp. 205–217. ISSN: 1053-587X. DOI: 10.1109/TSP.2007.901664.

[15] S. Bidon, O. Besson, and J. Y. Tourneret. "The Adaptive Coherence Estimator is the Generalized Likelihood Ratio Test for a Class of Heterogeneous Environments". In: *IEEE Signal Processing Letters* 15 (2008), pp. 281–284. ISSN: 1070-9908. DOI: 10.1109/LSP.2007.916044.

[16] A. W. Bitar, L.-F. Cheong, and J.-P. Ovarlez. "Simultaneous sparsity-based binary hypothesis model for real hyperspectral target detection". In: *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-17*. New Orleans, USA, Mar. 2017.

[17] A. W. Bitar, L.-F. Cheong, and J.-P. Ovarlez. "Sparse and Low-Rank Decomposition for Automatic Target Detection in Hyperspectral Imagery". In: (Nov. 2017). URL: https://arxiv.org/pdf/1711.08970.pdf.

[18] A. W. Bitar, J.-P. Ovarlez, and L.-F. Cheong. "Sparsity-Based Cholesky Factorization and Its Application to Hyperspectral Anomaly Detection". In: *IEEE Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP-17)*. Curaçao, Dutch Antilles, Dec. 2017. URL: https://arxiv.org/pdf/1711.08240.pdf.

[19] Thierry Bouwmans and El Hadi Zahzah. "Robust PCA via Principal Component Pursuit: A review for a comparative evaluation in video surveillance". In: *Computer Vision and Image Understanding* 122 (2014), pp. 22–34. ISSN: 1077-3142. DOI: https://doi.org/10.1016/j.cviu.2013.11.009. URL: http://www.sciencedirect.com/science/article/pii/S1077314213002294.

[20] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers". In: *Found. Trends Mach. Learn.* 3.1 (Jan. 2011), pp. 1–122. ISSN: 1935-8237. DOI: 10.1561/2200000016. URL: http://dx.doi.org/10.1561/2200000016.

[21] S. Charles Brubaker. "Robust PCA and Clustering in Noisy Mixtures". In: *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '09. New York, New York: Society for Industrial and Applied Mathematics, 2009, pp. 1078–1087. URL: http://dl.acm.org/citation.cfm?id=1496770.1496887.

[22] Alfred M. Bruckstein, David L. Donoho, and Michael Elad. "From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images". In: *SIAM Rev.* 51.1 (Feb. 2009), pp. 34–81. ISSN: 0036-1445. DOI: 10.1137/060657704. URL: http://dx.doi.org/10.1137/060657704.

[23] Jian-Feng Cai, Emmanuel J. Candès, and Zuowei Shen. "A Singular Value Thresholding Algorithm for Matrix Completion". In: *SIAM J. on Optimization* 20.4 (Mar. 2010), pp. 1956–1982. ISSN: 1052-6234. DOI: 10.1137/080738970. URL: http://dx.doi.org/10.1137/080738970.

[24] T. T. Cai and L. Wang. "Orthogonal Matching Pursuit for Sparse Signal Recovery With Noise". In: *IEEE Transactions on Information Theory* 57.7 (July 2011), pp. 4680–4688. ISSN: 0018-9448. DOI: 10.1109/TIT.2011.2146090.

[25] E. J. Candès, X. Li, Y. Ma, and J. Wright. "Robust Principal Component Analysis?" In: *J. ACM* 58.3 (June 2011), 11:1–11:37. ISSN: 0004-5411. DOI: 10.1145/1970392.1970395. URL: http://doi.acm.org/10.1145/1970392.1970395.

[26] E. J. Candes and B. Recht. "Exact low-rank matrix completion via convex optimization". In: *2008 46th Annual Allerton Conference on Communication, Control, and Computing*. Sept. 2008, pp. 806–812. DOI: 10.1109/ALLERTON.2008.4797640.

[27] E. Candès, M. B. Wakin, and S. P. Boyd. "Enhancing Sparsity by Reweighted l1 Minimization". In: *Journal of Fourier Analysis and Applications* 14.5 (2008), pp. 877–905.

[28] Emmanuel J. Candès, Michael B. Wakin, and Stephen P. Boyd. "Enhancing Sparsity by Reweighted l_1 Minimization". In: *Journal of Fourier Analysis and Applications* 14.5 (Dec. 2008), pp. 877–905. ISSN: 1531-5851. DOI: 10.1007/s00041-008-9045-x. URL: https://doi.org/10.1007/s00041-008-9045-x.

[29] G. Cao and C. Bouman. "Covariance Estimation for High Dimensional Data Vectors Using the Sparse Matrix Transform". In: *Advances in Neural Information Processing Systems 21*. Ed. by D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou. Curran Associates, Inc., 2009, pp. 225–232. URL: http://papers.nips.cc/paper/3409-covariance-estimation-for-high-dimensional-data-vectors-using-the-sparse-matrix-transform.pdf.

[30] Venkat Chandrasekaran, Sujay Sanghavi, Pablo A. Parrilo, and Alan S. Willsky. "Rank-Sparsity Incoherence for Matrix Decomposition". In: *SIAM Journal on Optimization* 21.2 (2011), pp. 572–596. DOI: 10.1137/090761793. eprint: https://doi.org/10.1137/090761793. URL: https://doi.org/10.1137/090761793.

[31] Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders. "Atomic Decomposition by Basis Pursuit". In: *SIAM J. Sci. Comput.* 20.1 (Dec. 1998), pp. 33–61. ISSN: 1064-8275. DOI: 10.1137/S1064827596304010. URL: http://dx.doi.org/10.1137/S1064827596304010.

[32] Shih-Yu Chen, Shiming Yang, Konstantinos Kalpakis, and Chein-I Chang. *Low-rank decomposition-based anomaly detection*. 2013. DOI: 10.1117/12.2015652. URL: http://dx.doi.org/10.1117/12.2015652.

[33] Y. Chen, N. M. Nasrabadi, and T. D. Tran. "Simultaneous Joint Sparsity Model for Target Detection in Hyperspectral Imagery". In: *IEEE Geoscience and Remote Sensing Letters* 8.4 (July 2011), pp. 676–680. ISSN: 1545-598X. DOI: 10.1109/LGRS.2010.2099640.

[34] Y. Chen, A. Wiesel, Y. C. Eldar, and A. O. Hero. "Shrinkage Algorithms for MMSE Covariance Estimation". In: *IEEE Transactions on Signal Processing* 58.10 (Oct. 2010), pp. 5016–5029. ISSN: 1053-587X. DOI: 10.1109/TSP.2010.2053029.

[35] Y. Chen, A. Wiesel, and A. O. Hero. "Robust shrinkage estimation of high-dimensional covariance matrices". In: *2010 IEEE Sensor Array and Multichannel Signal Processing Workshop*. Oct. 2010, pp. 189–192. DOI: 10.1109/SAM.2010.5606730.

[36] Z. Chen, S. D. Babacan, R. Molina, and A. K. Katsaggelos. "Variational Bayesian Methods For Multimedia Problems". In: *IEEE Transactions on Multimedia* 16.4 (June 2014), pp. 1000–1017. ISSN: 1520-9210. DOI: 10.1109/TMM.2014.2307692.

[37] Z. Chen and D. P. W. Ellis. "Speech enhancement by sparse, low-rank, and dictionary spectrogram decomposition". In: *2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. Oct. 2013, pp. 1–4. DOI: 10.1109/WASPAA.2013.6701883.

[38] Roger N. Clark, Gregg A. Swayze, Andrea J. Gallagher, Trude V.V. King, and Wendy M. Calvin. *The U. S. Geological Survey, Digital Spectral Library: Version 1: 0.2 to 3.0 micros*. Open file report. U.S. Geological Survey, 1993.

[39] A. Combernoux, F. Pascal, G. Ginolhac, and M. Lesturgie. "Random matrix theory applied to low rank stap detection". In: *21st European Signal Processing Conference (EUSIPCO 2013)*. Sept. 2013, pp. 1–5.

[40] A. Combernoux, F. Pascal, M. Lesturgie, and G. Ginolhac. "Performances of low rank detectors based on random matrix theory with application to STAP". In: *2014 International Radar Conference*. Oct. 2014, pp. 1–4. DOI: 10.1109/RADAR.2014.7060457.

[41] 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". In: *IEEE Transactions on Signal Processing* 50.8 (Aug. 2002), pp. 1908–1915. ISSN: 1053-587X. DOI: 10.1109/TSP.2002.800412.

[42] Noel Cressie. "STATISTICS FOR SPATIAL DATA". In: *Terra Nova* 4.5 (), pp. 613–617. DOI: 10.1111/j.1365-3121.1992.tb00605.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-3121.1992.tb00605.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-3121.1992.tb00605.x.

[43] B. Datt, T. R. McVicar, T. G. van Niel, D. L. B. Jupp, and J. S. Pearlman. "Preprocessing eo-1 hyperion hyperspectral data to support the application of agricultural indexes". In: *IEEE Transactions on Geoscience and Remote Sensing* 41 (June 2003), pp. 1246–1259. DOI: 10.1109/TGRS.2003.813206.

[44] G. Davis, S. Mallat, and M. Avellaneda. "Adaptive greedy approximations". In: *Constructive Approximation* 13.1 (Mar. 1997), pp. 57–98. ISSN: 1432-0940. DOI: 10.1007/BF02678430. URL: https://doi.org/10.1007/BF02678430.

[45] Fernando De La Torre and Michael J. Black. "A Framework for Robust Subspace Learning". In: *Int. J. Comput. Vision* 54.1-3 (Aug. 2003), pp. 117–142. ISSN: 0920-5691. DOI: 10.1023/A:1023709501986. URL: http://dx.doi.org/10.1023/A:1023709501986.

[46] Y. Deng, Q. Dai, R. Liu, Z. Zhang, and S. Hu. "Low-Rank Structure Learning via Nonconvex Heuristic Recovery". In: *IEEE Transactions on Neural Networks and Learning Systems* 24.3 (Mar. 2013), pp. 383–396. ISSN: 2162-237X. DOI: 10.1109/TNNLS.2012.2235082.

[47] X. Ding, L. He, and L. Carin. "Bayesian Robust Principal Component Analysis". In: *IEEE Transactions on Image Processing* 20.12 (Dec. 2011), pp. 3419–3430. ISSN: 1057-7149. DOI: 10.1109/TIP.2011.2156801.

[48] DAVID L. DONOHO and JAIN M. JOHNSTONE. "Ideal spatial adaptation by wavelet shrinkage". In: *Biometrika* 81.3 (1994), pp. 425–455. DOI: 10.1093/biomet/81.3.425. eprint: /oup/backfile/content_public/journal/biomet/81/3/10.1093/biomet/81.3.425/2/81-3-425.pdf. URL: +%20http://dx.doi.org/10.1093/biomet/81.3.425.

[49] Carl Eckart and Gale Young. "The approximation of one matrix by another of lower rank". In: *Psychometrika* 1.3 (1936), pp. 211–218. URL: http://EconPapers.repec.org/RePEc:spr:psycho:v:1:y:1936:i:3:p:211-218.

[50] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. "Least Angle Regression". In: *The Annals of Statistics* 32.2 (2004), pp. 407–451. ISSN: 00905364. URL: http://www.jstor.org/stable/3448465.

[51] M. T. Eismann, A. D. Stocker, and N. M. Nasrabadi. "Automated Hyperspectral Cueing for Civilian Search and Rescue". In: *Proceedings of the IEEE* 97.6 (2009), pp. 1031–1055. ISSN: 0018-9219. DOI: 10.1109/JPROC.2009.2013561.

[52] Michael Elad. *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. 1st. Springer Publishing Company, Incorporated, 2010. ISBN: 144197010X, 9781441970107.

[53] J. Fan and R. Li. "Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties". In: *Journal of the American Statistical Association* 96.456 (2001), pp. 1348–1360. DOI: 10.1198/016214501753382273. eprint: http://dx.doi.org/10.1198/016214501753382273. URL: http://dx.doi.org/10.1198/016214501753382273.

[54] Jiashi Feng, Huan Xu, and Shuicheng Yan. "Online Robust PCA via Stochastic Optimization". In: *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 1*. NIPS'13. Lake Tahoe, Nevada: Curran Associates Inc., 2013, pp. 404–412. URL: http://dl.acm.org/citation.cfm?id=2999611.2999657.

[55] Martin A. Fischler and Robert C. Bolles. "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography". In: *Commun. ACM* 24.6 (June 1981), pp. 381–395. ISSN: 0001-0782. DOI: 10.1145/358669.358692. URL: http://doi.acm.org/10.1145/358669.358692.

[56] S. Foucart and M.J Lai. "Sparsest solutions of underdetermined linear systems via l_q-minimization for $0 < q < 1$". In: *Applied and Computational Harmonic Analysis* 26.3 (2009), pp. 395–407. URL: http://alpha.math.uga.edu/~mjlai/papers/FL08.pdf.

[57] J. Frontera-Pons, F. Pascal, and J. P. 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*. Dec. 2013, pp. 161–164. DOI: 10.1109/CAMSAP.2013.6714032.

[58] J. Frontera-Pons, M. A. Veganzones, F. Pascal, and J. P. Ovarlez. "Hyperspectral Anomaly Detectors Using Robust Estimators". In: *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of* 9.2 (Feb. 2016), pp. 720–731.

[59] J. Frontera-Pons, M. A. Veganzones, S. Velasco-Forero, F. Pascal, J. P. Ovarlez, and J. Chanussot. "Robust anomaly detection in Hyperspectral Imaging". In: *2014 IEEE Geoscience and Remote Sensing Symposium*. July 2014, pp. 4604–4607. DOI: 10.1109/IGARSS.2014.6947518.

[60] "Frontmatter". In: *Variance Components*. Wiley-Blackwell, 2008, pp. i–23. ISBN: 9780470316856. DOI: 10.1002/9780470316856.fmatter. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470316856.fmatter. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470316856.fmatter.

[61] Box. G, Jenkins. G, and Reinsel. G. *Times Series Analysis*. 3rd ed. Prentice Hall, 1994.

[62] Z. Gao, L. F. Cheong, and Y. X. Wang. "Block-Sparse RPCA for Salient Motion Detection". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36.10 (Oct. 2014), pp. 1975–1987. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2014.2314663.

[63] Zhi Gao, Loong-Fah Cheong, and Mo Shan. "Block-Sparse RPCA for Consistent Foreground Detection". In: *Computer Vision – ECCV 2012*. Ed. by Andrew Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato, and Cordelia Schmid. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 690–703. ISBN: 978-3-642-33715-4.

[64] D. Geman and Chengda Yang. "Nonlinear image recovery with half-quadratic regularization". In: *IEEE Transactions on Image Processing* 4.7 (July 1995), pp. 932–946. ISSN: 1057-7149. DOI: 10.1109/83.392335.

[65] F. Gini. "Sub-optimum coherent radar detection in a mixture of K-distributed and Gaussian clutter". In: *IEE Proceedings - Radar, Sonar and Navigation* 144.1 (Feb. 1997), pp. 39–48. ISSN: 1350-2395. DOI: 10.1049/ip-rsn:19970967.

[66] F. Gini and M. V Greco. "Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter". In: *Signal Processing, special section on Signal Processing with Heavy Tailed Distributions,* 82 (Dec. 2002), pp. 1847–1859.

[67] R. Gnanadesikan and J. R. Kettenring. "Robust Estimates, Residuals, and Outlier Detection with Multiresponse Data". In: *Biometrics* 28.1 (1972), pp. 81–124. ISSN: 0006341X, 15410420. URL: http://www.jstor.org/stable/2528963.

[68] J. Goes, T. Zhang, R. Arora, and G. Lerman. "Robust stochastic principal component analysis". In: AISTATS. 2014.

[69] P. Gong, J. Ye, and C. Zhang. "Multi-Stage Multi-Task Feature Learning". In: *Advances in Neural Information Processing Systems 25*. Ed. by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger. Curran Associates, Inc., 2012, pp. 1988–1996. URL: http://papers.nips.cc/paper/4729-multi-stage-multi-task-feature-learning.pdf.

[70] P. Gong, C. Zhang, Z. Lu, J. Huang, and J. Ye. *GIST: General Iterative Shrinkage and Thresholding for Non-convex Sparse Learning*. 2013. (Visited on).

[71] I. F. Gorodnitsky and B. D. Rao. "Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm". In: *IEEE Transactions on Signal Processing* 45.3 (Mar. 1997), pp. 600–616. ISSN: 1053-587X. DOI: 10.1109/78.558475.

[72] Z. Guo, T. Wittman, and S. Osher. *L1 unmixing and its application to hyperspectral image enhancement*. 2009. DOI: 10.1117/12.818245. URL: https://doi.org/10.1117/12.818245.

[73] C. Guyon, T. Bouwmans, and E. H. Zahzah. "Foreground detection via robust low rank matrix factorization including spatial constraint with Iterative reweighted regression". In: *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*. Nov. 2012, pp. 2805–2808.

[74] Charles Guyon, Thierry Bouwmans, and El-Hadi Zahzah. "Foreground Detection via Robust Low Rank Matrix Decomposition Including Spatio-Temporal Constraint". In: *Computer Vision - ACCV 2012 Workshops*. Ed. by Jong-Il Park and Junmo Kim. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 315–320. ISBN: 978-3-642-37410-4.

[75] Charles Guyon, Thierry Bouwmans, and El-Hadi Zahzah. "Moving Object Detection via Robust Low Rank Matrix Decomposition with IRLS Scheme". In: *Advances in Visual Computing*. Ed. by George Bebis, Richard Boyle, Bahram Parvin, Darko Koracin, Charless Fowlkes, Sen Wang, Min-Hyung Choi, Stephan Mantler, Jürgen Schulze, Daniel Acevedo, Klaus Mueller, and Michael Papka. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 665–674. ISBN: 978-3-642-33179-4.

[76] X. Hang and F.-X Wu. "Sparse Representation for Classification of Tumors Using Gene Expression Data". In: *Journal of Biomedicine and Biotechnology*. 2009, p. 6.

[77] G. Healey and D. Slater. "Models and methods for automated material identification in hyperspectral imagery acquired under unknown illumination and atmospheric conditions". In: *IEEE Transactions on Geoscience and Remote Sensing* 37.6 (1999), pp. 2706–2717. ISSN: 0196-2892. DOI: 10.1109/36.803418.

[78] Arthur E. Hoerl and Robert W. Kennard. "Ridge Regression: Biased Estimation for Nonorthogonal Problems". In: *Technometrics* 12.1 (1970), pp. 55–67. DOI: 10.1080/00401706.1970.10488634. eprint: https://www.tandfonline.com/doi/pdf/10.1080/00401706.1970.10488634. URL: https://www.tandfonline.com/doi/abs/10.1080/00401706.1970.10488634.

[79] B. Hörig, F. Kühn, F. Oschütz, and F. Lehmann. "HyMap hyperspectral remote sensing to detect hydrocarbons". In: *International Journal of Remote Sensing* 22 (May 2001), pp. 1413–1422. DOI: 10.1080/01431160010013450.

[80] H. Hotelling. "Analysis of a complex of statistical variables into principal components". In: *Journal of Educational Psychology* 24 (1933), pp. 417–441.

[81] D. Hsu, S. M. Kakade, and T. Zhang. "Robust Matrix Decomposition With Sparse Corruptions". In: *IEEE Transactions on Information Theory* 57.11 (Nov. 2011), pp. 7221–7234. ISSN: 0018-9448. DOI: 10.1109/TIT.2011.2158250.

[82] J. Z. Huang, N. Liu, M. Pourahmadi, and L. Liu. "Covariance matrix selection and estimation via penalised normal likelihood". In: *Biometrika* 93.1 (2006), pp. 85–98. DOI: 10.1093/biomet/93.1.85. eprint: http://biomet.oxfordjournals.org/content/93/1/85.full.pdf+html. URL: http://biomet.oxfordjournals.org/content/93/1/85.abstract.

[83] P. Huber. *Robust Statistics*. Wiley and Sons, 1981.

[84] W. James and Charles Stein. "Estimation with Quadratic Loss". In: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. Berkeley, Calif.: University of California Press, 1961, pp. 361–379. URL: https://projecteuclid.org/euclid.bsmsp/1200512173.

[85] Sajid Javed, Seon Ho Oh, JunHyeok Heo, and Soon Ki Jung. "Robust Background Subtraction via Online Robust PCA Using Image Decomposition". In: *Proceedings of the 2014 Conference on Research in Adaptive and Convergent Systems*. RACS '14. Towson, Maryland: ACM, 2014, pp. 105–110. ISBN: 978-1-4503-3060-2. DOI: 10.1145/2663761.2664195. URL: http://doi.acm.org/10.1145/2663761.2664195.

[86] Wei Jiang, Jing Liu, Heng Qi, and Qionghai Dai. "Robust subspace segmentation via nonconvex low rank representation". In: *Information Sciences* 340-341 (2016), pp. 144–158. ISSN: 0020-0255. DOI: https://doi.org/10.1016/j.ins.2015.12.038. URL: http://www.sciencedirect.com/science/article/pii/S0020025516000219.

[87] I. Jolliffe. *Principal Component Analysis*. Springer-Verlag, 1986.

[88] J Nathan K., G. Jacob, B. Steven, F. Xing, and P. Seth. "USING DYNAMIC MODE DECOMPOSITION FOR REAL-TIME BACKGROUND/FOREGROUND SEPARATION IN VIDEO". In: 20160050343 (2016). URL: http://www.freepatentsonline.com/y2016/0050343.html.

[89] Qifa Ke and Takeo Kanade. "Robust L" Norm Factorization in the Presence of Outliers and Missing Data by Alternative Convex Programming". In: *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01*. CVPR '05. Washington, DC, USA: IEEE Computer Society, 2005, pp. 739–746. ISBN: 0-7695-2372-2. DOI: 10.1109/CVPR.2005.309. URL: http://dx.doi.org/10.1109/CVPR.2005.309.

[90] Douglas Kelker. "Distribution Theory of Spherical Distributions and a Location-Scale Parameter Generalization". In: *Sankhy?: The Indian Journal of Statistics, Series A (1961-2002)* 32.4 (1970), pp. 419–430. ISSN: 0581572X. URL: http://www.jstor.org/stable/25049690.

[91] E. J. Kelly. "An Adaptive Detection Algorithm". In: *Aerospace and Electronic Systems, IEEE Transactions on* 23.1 (Nov. 1986), pp. 115–127.

[92] E. J. Kelly. "An Adaptive Detection Algorithm". In: *IEEE Transactions on Aerospace and Electronic Systems* AES-22.2 (Mar. 1986), pp. 115–127. ISSN: 0018-9251. DOI: 10.1109/TAES.1986.310745.

[93] S. Kraut and L.L. Scharf. "The CFAR adaptive subspace detector is a scale-invariant GLRT". In: *Signal Processing, IEEE Transactions on* 47.9 (1999), pp. 2538–2541.

[94] Olivier Ledoit and Michael Wolf. "A well-conditioned estimator for large-dimensional covariance matrices". In: *Journal of Multivariate Analysis* 88.2 (2004), pp. 365–411. ISSN: 0047-259X. DOI: http://dx.doi.org/10.1016/S0047-259X(03)00096-4. URL: http://www.sciencedirect.com/science/article/pii/S0047259X03000964.

[95] Olivier Ledoit and Michael Wolf. "Honey, I Shrunk the Sample Covariance Matrix". In: *UPF Economics and Business Working Paper* 691 (2003). DOI: http://dx.doi.org/10.2139/ssrn.433840.

[96] A.S. Lewis. "The mathematics of eigenvalue optimization". In: *Mathematical Programming* 97.1 (July 2003), pp. 155–176. ISSN: 1436-4646. DOI: 10.1007/s10107-003-0441-3. URL: https://doi.org/10.1007/s10107-003-0441-3.

[97] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma. "Robust Recovery of Subspace Structures by Low-Rank Representation". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35.1 (Jan. 2013), pp. 171–184. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2012.88.

[98] H. Liu, L. Wang, and T. Zhao. "Sparse Covariance Matrix Estimation With Eigenvalue Constraints". In: *J Comput Graph Stat* (23).2 (Apr. 2014), pp. 439–459. ISSN: 1350-2395. DOI: 10.1080/10618600.2013.782818.

[99] S. Mallat. *A Wavelet Tour of Signal Processing*. 2st. Academic-Press, Second-Edition, 2009. ISBN: 144197010X, 9781441970107.

[100] S. G. Mallat and Zhifeng Zhang. "Matching pursuits with time-frequency dictionaries". In: *IEEE Transactions on Signal Processing* 41.12 (Dec. 1993), pp. 3397–3415. ISSN: 1053-587X. DOI: 10.1109/78.258082.

[101] D. Manolakis, R. Lockwood, T. Cooley, and J. Jacobson. "Is there a best hyperspectral detection algorithm?" In: *Proc. SPIE 7334* (2009), p. 733402.

[102] D. Manolakis and D. Marden. "Non Gaussian models for hyperspectral algorithm design and assessment". In: *IEEE International Geoscience and Remote Sensing Symposium*. Vol. 3. June 2002, 1664–1666 vol.3. DOI: 10.1109/IGARSS.2002.1026214.

[103] D. Manolakis, D. Marden, and G. Shaw. "Hyperspectral image processing for automatic target detection applications". In: *Lincoln Laboratory Journal* 14.1 (2003), pp. 79–116.

[104] D. Manolakis and G. Shaw. "Detection algorithms for hyperspectral imaging applications". In: *Signal Processing Magazine, IEEE* 19.1 (2002), pp. 29–43.

[105] D. Manolakis, G. Shaw, and N. Keshava. "Comparative analysis of hyperspectral adaptive matched filter detectors". In: *Proc. SPIE 4049, Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI* 2 (2000).

[106] D. Manolakis, E. Truslow, M. Pieper, T. Cooley, and M. Brueggeman. "Detection Algorithms in Hyperspectral Imaging Systems: An Overview of Practical Algorithms". In: *IEEE Signal Processing Magazine* 31.1 (Jan. 2014), pp. 24–33. ISSN: 1053-5888. DOI: 10.1109/MSP.2013.2278915.

[107] Dimitris G. Manolakis, Ronald B. Lockwood, and Thomas W. Cooley. *Hyperspectral Imaging Remote Sensing: Physics, Sensors, and Algorithms*. Cambridge University Press, 2016. DOI: 10.1017/CBO9781316017876.

[108] G. Mateos and G. B. Giannakis. "Robust PCA as Bilinear Decomposition With Outlier-Sparsity Regularization". In: *IEEE Transactions on Signal Processing* 60.10 (Oct. 2012), pp. 5176–5190. ISSN: 1053-587X. DOI: 10.1109/TSP.2012.2204986.

[109] G. Mateos and G. B. Giannakis. "Sparsity control for robust principal component analysis". In: *2010 Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers*. Nov. 2010, pp. 1925–1929. DOI: 10.1109/ACSSC.2010.5757875.

[110] S. Matteoli, M. Diani, and G. Corsini. "A tutorial overview of anomaly detection in hyperspectral images". In: *Aerospace and Electronic Systems Magazine, IEEE* 25.7 (2010), pp. 5–28.

[111] Official Website of the MUSE Project. "http://muse.univ-lyon.fr/". In: ().

[112] S. Nakajima, M. Sugiyama, and S.D. Babacan. "Sparse Additive Matrix Factorization for Robust PCA and Its Generalization". In: *Proceedings of the Asian Conference on Machine Learning*. Ed. by Steven C. H. Hoi and Wray Buntine. Vol. 25. Proceedings of Machine Learning Research. Singapore Management University, Singapore: PMLR, Apr. 2012, pp. 301–316. URL: http://proceedings.mlr.press/v25/nakajima12.html.

[113] Shinichi Nakajima, Masashi Sugiyama, and S. Derin Babacan. "Variational Bayesian sparse additive matrix factorization". In: *Machine Learning* 92.2 (Sept. 2013), pp. 319–347. ISSN: 1573-0565. DOI: 10.1007/s10994-013-5347-6. URL: https://doi.org/10.1007/s10994-013-5347-6.

[114] N. M. Nasrabadi. "Regularized Spectral Matched Filter for Target Recognition in Hyperspectral Imagery". In: *IEEE Signal Processing Letters* 15 (2008), pp. 317–320. ISSN: 1070-9908. DOI: 10.1109/LSP.2008.917805.

[115] B. K. Natarajan. "Sparse Approximate Solutions to Linear Systems". In: *SIAM J. Comput.* 24.2 (Apr. 1995), pp. 227–234. ISSN: 0097-5397. DOI: 10.1137/S0097539792240406. URL: http://dx.doi.org/10.1137/S0097539792240406.

[116] J. P. Ovarlez, S. K. Pang, F. Pascal, V. Achard, and T. K. Ng. "Robust detection using the SIRV background modelling for hyperspectral imaging". In: *2011 IEEE International Geoscience and Remote Sensing Symposium*. July 2011, pp. 4316–4319. DOI: 10.1109/IGARSS.2011.6050186.

[117] F. Pascal and Y. Chitour. "Shrinkage covariance matrix estimator applied to STAP detection". In: *2014 IEEE Workshop on Statistical Signal Processing (SSP)*. June 2014, pp. 324–327. DOI: 10.1109/SSP.2014.6884641.

[118] F. Pascal, Y. Chitour, and Y. Quek. "Generalized Robust Shrinkage Estimator and Its Application to STAP Detection Problem". In: *IEEE Transactions on Signal Processing* 62.21 (Nov. 2014), pp. 5640–5651. ISSN: 1053-587X. DOI: 10.1109/TSP.2014.2355779.

[119] Frédéric Pascal. "Détection et estimation en environnement non gaussien". 2006PA100124. PhD thesis. 2006, 1 vol. (153 p.) URL: http://www.theses.fr/2006PA100124.

[120] N. K. Patel, C. Patnaik, S. Dutta, A. M. Shekh, and A. J. Dave. "Study of crop growth parameters using Airborne Imaging Spectrometer data". In: *International Journal of Remote Sensing* 22.12 (2001), pp. 2401–2411. DOI: 10.1080/01431160117383. eprint: http://www.tandfonline.com/doi/pdf/10.1080/01431160117383. URL: http://www.tandfonline.com/doi/abs/10.1080/01431160117383.

[121] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad. "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition". In: *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*. Nov. 1993, 40–44 vol.1. DOI: 10.1109/ACSSC.1993.342465.

[122] J. K. Pillai, V. M. Patel, and R. Chellappa. "Sparsity inspired selection and recognition of iris images". In: *2009 IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems*. Sept. 2009, pp. 1–6. DOI: 10.1109/BTAS.2009.5339067.

[123] Joana Maria Frontera Pons. "Robust target detection for Hyperspectral Imaging". PhD thesis. CentraleSupelec, 2014. URL: https://tel.archives-ouvertes.fr/tel-01165283/document.

[124] M. Pourahmadi. "Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation". In: *Biometrika* 86.3 (1999), pp. 677–690. DOI: 10.1093/biomet/86.3.677. eprint: http://biomet.oxfordjournals.org/content/86/3/677.full.pdf+html.

[125] I. S. Reed and X. Yu. "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution". In: *IEEE Transactions on Acoustics Speech and Signal Processing* 38 (Oct. 1990), pp. 1760–1770.

[126] John A. Richards and Xiuping Jia. *Remote Sensing Digital Image Analysis: An Introduction*. Ed. by D. E. Ricken and W. Gessner. 3rd. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 1999. ISBN: 3540648607.

[127] A. J. Rothman, E. Levina, and J. Zhu. "Generalized Thresholding of Large Covariance Matrices". In: *Journal of the American Statistical Association* 104.485 (2009), pp. 177–186. ISSN: 01621459. URL: http://www.jstor.org/stable/40591909.

[128] G. Shaw and D. Manolakis. "Signal processing for hyperspectral image exploitation". In: *IEEE Signal Processing Magazine* 19.1 (Jan. 2002), pp. 12–16. ISSN: 1053-5888. DOI: 10.1109/79.974715.

[129] S. K. Shevade and S. S. Keerthi. "A simple and efficient algorithm for gene selection using sparse logistic regression". In: *Bioinformatics* 19.17 (2003), pp. 2246–2253. DOI: 10.1093/bioinformatics/btg308. eprint: /oup/backfile/content_public/journal/bioinformatics/19/17/10.1093/bioinformatics/btg308/2/btg308.pdf. URL: +%20http://dx.doi.org/10.1093/bioinformatics/btg308.

[130] Charles Stein. "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution". In: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. Berkeley, Calif.: University of California Press, 1956, pp. 197–206. URL: http://projecteuclid.org/euclid.bsmsp/1200501656.

[131] D. W. J. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker. "Anomaly detection from hyperspectral imagery". In: *IEEE Signal Processing Magazine* 19.1 (Jan. 2002), pp. 58–69. ISSN: 1053-5888. DOI: 10.1109/79.974730.

[132] P. Sun and J. Qin. "Low-Rank and Sparsity Analysis Applied to Speech Enhancement Via Online Estimated Dictionary". In: *IEEE Signal Processing Letters* 23.12 (Dec. 2016), pp. 1862–1866. ISSN: 1070-9908. DOI: 10.1109/LSP.2016.2627029.

[133] Gregg A. Swayze, Roger N. Clark, Alexander F. H. Goetz, Thomas G. Chrien, and Noel S. Gorelick. "Effects of spectrometer band pass, sampling, and signal-to-noise ratio on spectral identification using the Tetracorder algorithm". In: *Journal of Geophysical Research: Planets* 108.E9 (2003). 5105. ISSN: 2156-2202. DOI: 10.1029/2002JE001975. URL: http://dx.doi.org/10.1029/2002JE001975.

[134] Gregg A. Swayze, Roger N. Clark, Alexander F.H. Goetz, K. Eric Livo, George N. Breit, Fred A. Kruse, Stephen J. Sutley, Lawrence W. Snee, Heather A. Lowers, James L. Post, Roger E. Stoffregen, and Roger P. Ashley. "Mapping Advanced Argillic Alteration at Cuprite, Nevada, Using Imaging Spectroscopy". In: *Economic Geology* 109.5 (2014), p. 1179. DOI: 10.2113/econgeo.109.5.1179. eprint: /gsw/content_public/journal/economicgeology/109/5/10.2113_econgeo.109.5.1179/4/1179-1221.pdf. URL: +%20http://dx.doi.org/10.2113/econgeo.109.5.1179.

[135] E. Terreaux, J. P. Ovarlez, and F. Pascal. "Anomaly detection and estimation in hyperspectral imaging using Random Matrix Theory tools". In: *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2015 IEEE 6th International Workshop on*. Dec. 2015, pp. 169–172. DOI: 10.1109/CAMSAP.2015.7383763.

[136] E. Terreaux, J. P. Ovarlez, and F. Pascal. "New model order selection in large dimension regime for complex elliptically symmetric noise". In: *2017 25th European Signal Processing Conference (EUSIPCO)*. Aug. 2017, pp. 1090–1094. DOI: 10.23919/EUSIPCO.2017.8081376.

[137] E. Terreaux, J. P. Ovarlez, and F. Pascal. "Robust Model Order Selection in Large Dimensional Elliptically Symmetric Noise". In: (Oct. 2017). URL: https://arxiv.org/pdf/1710.06735.pdf.

[138] B. Thai and G. Healey. "Invariant subpixel material detection in hyperspectral imagery". In: *IEEE Transactions on Geoscience and Remote Sensing* 40.3 (2002), pp. 599–608. ISSN: 0196-2892. DOI: 10.1109/TGRS.2002.1000320.

[139] R. Tibshirani. "Regression Shrinkage and Selection via the Lasso." In: *Journal of the Royal Statistical Society. Series B (Methodological)* 58.1 (1996), pp. 267–288.

[140] Robert Tibshirani. "Regression Shrinkage and Selection via the Lasso". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 58.1 (1996), pp. 267–288. ISSN: 00359246. URL: http://www.jstor.org/stable/2346178.

[141] F. De la Torre and M. J. Black. "Robust principal component analysis for computer vision". In: *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*. Vol. 1. 2001, 362–369 vol.1. DOI: 10.1109/ICCV.2001.937541.

[142] J. A. Tropp and A. C. Gilbert. "Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit". In: *IEEE Transactions on Information Theory* 53.12 (Dec. 2007), pp. 4655–4666. ISSN: 0018-9448. DOI: 10.1109/TIT.2007.909108.

[143] J. A. Tropp, A. C. Gilbert, and M. J. Strauss Martin. "Algorithms for Simultaneous Sparse Approximation: Part I: Greedy Pursuit". In: *Signal Process.* 86.3 (Mar. 2006), pp. 572–588. ISSN: 0165-1684. DOI: 10.1016/j.sigpro.2005.05.030. URL: http://dx.doi.org/10.1016/j.sigpro.2005.05.030.

[144] J. A. Tropp and S. J. Wright. "Computational Methods for Sparse Solution of Linear Inverse Problems". In: *Proceedings of the IEEE* 98.6 (June 2010), pp. 948–958. ISSN: 0018-9219. DOI: 10.1109/JPROC.2010.2044010.

[145] J. Trzasko and A. Manduca. "Relaxed Conditions for Sparse Signal Recovery With General Concave Priors". In: *IEEE Transactions on Signal Processing* 57.11 (Nov. 2009), pp. 4347–4354. ISSN: 1053-587X. DOI: 10.1109/TSP.2009.2025979.

[146] airborne Visible / Infrared Imaging Spectrometer. "https://aviris.jpl.nasa.gov". In: ().

[147] G.A. Watson. "Characterization of the subdifferential of some matrix norms". In: *Linear Algebra and its Applications* 170 (1992), pp. 33–45. ISSN: 0024-3795. DOI: https://doi.org/10.1016/0024-3795(92)90407-2. URL: http://www.sciencedirect.com/science/article/pii/0024379592904072.

[148] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. "Robust Face Recognition via Sparse Representation". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31.2 (Feb. 2009), pp. 210–227. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2008.79.

[149] John Wright, Arvind Ganesh, Shankar Rao, Yigang Peng, and Yi Ma. "Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices via Convex Optimization". In: *Advances in Neural Information Processing Systems 22*. Ed. by Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta. Curran Associates, Inc., 2009, pp. 2080–2088. URL: http://papers.nips.cc/paper/3704-robust-principal-component-analysis-exact-recovery-of-corrupted-low-rank-matrices-via-convex-optimization.pdf.

[150] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo. "Sparse Reconstruction by Separable Approximation". In: *IEEE Transactions on Signal Processing* 57.7 (July 2009), pp. 2479–2493. ISSN: 1053-587X. DOI: 10.1109/TSP.2009.2016892.

[151] Wei Biao Wu and Mohsen Pourahmadi. "Nonparametric estimation of large covariance matrices of longitudinal data". In: *Biometrika* 90.4 (2003), p. 831. DOI: 10.1093/biomet/90.4.831. eprint: /oup/backfile/content_public/journal/biomet/90/4/10.1093/biomet/90.4.831/2/900831.pdf. URL: +%20http://dx.doi.org/10.1093/biomet/90.4.831.

[152] H. Xu, C. Caramanis, and S. Sanghavi. "Robust PCA via Outlier Pursuit". In: *IEEE Transactions on Information Theory* 58.5 (May 2012), pp. 3047–3064. ISSN: 0018-9448. DOI: 10.1109/TIT.2011.2173156.

[153] Y. Xu, Z. Wu, J. Chanussot, M. Dalla Mura, A. L. Bertozzi, and Z. Wei. "Low-Rank Decomposition and Total Variation Regularization of Hyperspectral Video Sequences". In: *IEEE Transactions on Geoscience and Remote Sensing* 56.3 (Mar. 2018), pp. 1680–1694. ISSN: 0196-2892. DOI: 10.1109/TGRS.2017.2766094.

[154] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei. "Joint Reconstruction and Anomaly Detection From Compressive Hyperspectral Images Using Mahalanobis Distance-Regularized Tensor RPCA". In: *IEEE Transactions on Geoscience and Remote Sensing* 56.5 (May 2018), pp. 2919–2930. ISSN: 0196-2892. DOI: 10.1109/TGRS.2017.2786718.

[155] Y. Xu, Z. Wu, J. Li, A. Plaza, and Z. Wei. "Anomaly Detection in Hyperspectral Images Based on Low-Rank and Sparse Representation". In: *IEEE Transactions on Geoscience and Remote Sensing* 54.4 (Apr. 2016), pp. 1990–2000. ISSN: 0196-2892. DOI: 10.1109/TGRS.2015.2493201.

[156] Junfeng Yang, Wotao Yin, Yin Zhang, and Yilun Wang. "A Fast Algorithm for Edge-Preserving Variational Multichannel Image Restoration". In: *SIAM Journal on Imaging Sciences* 2.2 (2009), pp. 569–592. DOI: 10.1137/080730421. eprint: https://doi.org/10.1137/080730421. URL: https://doi.org/10.1137/080730421.

[157] K. Yao, M. K. Simon, and E. Bigileri. "Unified theory on wireless communication fading statistics based on SIRP". In: *IEEE 5th Workshop on Signal Processing Advances in Wireless Communications, 2004.* July 2004, pp. 135–139. DOI: 10.1109/SPAWC.2004.1439219.

[158] Jieping Ye and Jun Liu. "Sparse Methods for Biomedical Data". In: *SIGKDD Explor. Newsl.* 14.1 (Dec. 2012), pp. 4–15. ISSN: 1931-0145. DOI: 10.1145/2408736.2408739. URL: http://doi.acm.org/10.1145/2408736.2408739.

[159] Cun-Hui Zhang. "Nearly unbiased variable selection under minimax concave penalty". In: *Ann. Statist.* 38.2 (Apr. 2010), pp. 894–942. DOI: 10.1214/09-AOS729. URL: http://dx.doi.org/10.1214/09-AOS729.

[160] Huishuai Zhang, Yi Zhou, and Yingbin Liang. "Analysis of Robust PCA via Local Incoherence". In: *Advances in Neural Information Processing Systems 28*. Ed. by C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett. Curran Associates, Inc., 2015, pp. 1819–1827. URL: http://papers.nips.cc/paper/6018-analysis-of-robust-pca-via-local-incoherence.pdf.

[161] Tong Zhang. "Analysis of Multi-stage Convex Relaxation for Sparse Regularization". In: *J. Mach. Learn. Res.* 11 (Mar. 2010), pp. 1081–1107. ISSN: 1532-4435. URL: http://dl.acm.org/citation.cfm?id=1756006.1756041.

[162] Tong Zhang. "Multi-stage convex relaxation for feature selection". In: *Bernoulli* 19.5B (Nov. 2013), pp. 2277–2293. DOI: 10.3150/12-BEJ452. URL: http://dx.doi.org/10.3150/12-BEJ452.

[163] Y. Zhang, B. Du, and L. Zhang. "A Sparse Representation-Based Binary Hypothesis Model for Target Detection in Hyperspectral Images". In: *IEEE Transactions on Geoscience and Remote Sensing* 53.3 (Mar. 2015), pp. 1346–1354. ISSN: 0196-2892. DOI: 10.1109/TGRS.2014.2337883.

[164] Y. Zhang, B. Du, L. Zhang, and S. Wang. "A Low-Rank and Sparse Matrix Decomposition-Based Mahalanobis Distance Method for Hyperspectral Anomaly Detection". In: *IEEE Transactions on Geoscience and Remote Sensing* 54.3 (Mar. 2016), pp. 1376–1389. ISSN: 0196-2892. DOI: 10.1109/TGRS.2015.2479299.

[165] Qian Zhao, Deyu Meng, Zongben Xu, Wangmeng Zuo, and Lei Zhang. "Robust Principal Component Analysis with Complex Noise". In: *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32*. ICML'14. Beijing, China: JMLR.org, 2014, pp. II-55–II-63. URL: http://dl.acm.org/citation.cfm?id=3044805.3044899.

[166] T. Zhou and D. Tao. "Manifold elastic net for sparse learning". In: *2009 IEEE International Conference on Systems, Man and Cybernetics*. Oct. 2009, pp. 3699–3704. DOI: 10.1109/ICSMC.2009.5346879.

[167] Tianyi Zhou and Dacheng Tao. "Godec: Randomized low-rank and sparse matrix decomposition in noisy case". In: *in International Conference on Machine Learning*. 2011.

[168] Zihan Zhou, Xiaodong Li, John Wright, Emmanuel J. Candès, and Yi Ma. "Stable Principal Component Pursuit". In: *CoRR* abs/1001.2363 (2010). URL: http://arxiv.org/abs/1001.2363.

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

Titre : Exploitation de la parcimonie pour la détection de cibles dans les images hyperspectrales

Mots clés : Parcimonie, image hyperspectrale, détection de cibles

Résumé : Le titre de cette thèse de doctorat est formé de trois mots clés: parcimonie, image hyperspectrale, et détection de cibles. La parcimonie signifie généralement « petit en nombre ou quantité, souvent répartie sur une grande zone ». Une image hyperspectrale est constituée d'une série d'images de la même scène spatiale, mais prises dans plusieurs dizaines de longueurs d'onde contiguës et très étroites, qui correspondent à autant de "couleurs". Lorsque la dimension spectrale est très grande, la détection de cibles devient délicate et caractérise une des applications les plus importantes pour les images hyperspectrales. Le but principal de cette thèse de doctorat est de répondre à la question « Comment et Pourquoi la parcimonie peut-elle être exploitée pour détecter de cibles dans les images hyperspectrales? ». La réponse à cette question nous a permis de développer des méthodes de détection de cibles prenant en compte l'hétérogénéité de l'environnement, le fait que les objets d'intérêt sont situés dans des parties relativement réduites de l'image observée et enfin que l'estimation de la matrice de covariance d'un pixel d'une image hyperspectrale peut être compliquée car cette matrice appartient à un espace de grande dimension. Les méthodes proposées sont évaluées sur des données synthétiques ainsi que réelles, dont les résultats démontrent leur efficacité pour la détection de cibles dans les images hyperspectrales.

Title : Exploitation of Sparsity for Hyperspectral Target Detection

Keywords : Sparsity, hyperspectral image, target detection

Abstract : The title of this PhD thesis is formed by three keywords: sparsity, hyperspectral image, and target detection. Sparsity is a word that is used everywhere and in everyday life. It generally means « small in number or amount, often spread over a large area ». A hyperspectral image is a three dimensional data cube consisting of a series of images of the same spatial scene in a contiguous and multiple narrow spectral wavelength (color) bands. According to the high spectral dimensionality, target detection is not surprisingly one of the most important applications in hyperspectral imagery. The main objective of this PhD thesis is to answer the question « How and Why can sparsity be exploited for hyperspectral target detection? ». Answering this question has allowed us to develop different target detection methods that mainly take into consideration the heterogeneity of the environment, the fact that the total image area of all the targets is very small relative to the whole image, and the estimation challenge of the covariance matrix (surrounding the test pixel) in large dimensions. The proposed mehods are evaluated on both synthetic and real experiments, the results of which demonstrate their effectiveness for hyperspectral target detection.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France