UNIVERSIDADE TÉCNICA DE LISBOA

INSTITUTO SUPERIOR TÉCNICO

Unsupervised Hyperspectral Unmixing

José Manuel Peixoto Nascimento

(Mestre)

Dissertação para obtenção do Grau de Doutor em

Engenharia Electrotécnica e de Computadores

Orientador: Doutor José Manuel Bioucas Dias

Júri

Presidente: Reitor da Universidade Técnica de Lisboa

Vogais: Doutor Luís Henrique Martins Borges de Almeida

Doutor António Miguel Plaza

Doutor Mário Alexandre Teles de Figueiredo

Doutor Arnaldo Joaquim de Castro Abrantes

Doutor José Manuel Bioucas Dias

Dezembro de 2006

Abstract

This thesis addresses hyperspectral unmixing, which is the decomposition of pixel spectra acquired by spectral sensors into a collection of constituent spectra, or endmember spectral signatures, and their corresponding abundance fractions. The need for hyperspectral unmixing has been fostered by the development of extremely powerful hyperspectral sensors, with hundreds of narrow (on the order of $10nm$) contiguous bands, spanning the visible, the near-infrared, and the mid-infrared wavelengths (between $0.3\mu m$ and $2.5\mu m$) of the electromagnetic spectrum. These sensors have improved the recognition of Earth substances based on their signatures, *i.e.*, based on the scattered electromagnetic energy on each wavelength. However, since the spatial resolution of any sensor is finite, those substances are very often spatially mixed. In given circumstances, these substances can be unmixed and thus resolved without increasing the spatial resolution of the sensor. This thesis introduces three new methodologies playing relevant roles in the hyperspectral unmixing processing chain.

The first method, termed *hyperspectral signal identification by minimum error* (HySime), estimates the signal subspace to reduce the dimensionality of the data to be processed. Since the number of components of hyperspectral vector is large (typically more than 200 bands) and the dimension of the signal subspace is often much smaller, dimensionality reduction leads to reductions in memory requirements and gains in the signal-to-noise ratio.

The second method, termed *vertex component analysis* (VCA), extracts the endmembers exploiting the fact that the spectra of pure pixels are in vertices of a simplex. VCA searches for these vertices based on the assumption that at least one pure pixel per endmember is present in the data set. VCA achieves state-of-the-art performance with a computational complexity between one and two orders of magnitude lower than the best algorithms.

The third method, termed *dependent component analysis* (DECA), formalizes spectral unmixing as a maximum likelihood (ML) problem. DECA uses the expectation maximization (EM) framework to infer the unmixing matrix. The abundance fraction densities are modeled by a mixture of Dirichlet densities, thus modeling the statistical dependence normally found in hyperspectral data. Compared with VCA, DECA does not need to assume the presence of pure pixels in the observations.

Independent component analysis (ICA) and *independent factor analysis* (IFA) have been used by many authors to unmix hyperspectral data. Due to physical constraints ever present in hyperspectral data, the ICA and IFA central assumption of source statistical independence does not hold. The impact of this source statistical dependence is addressed. It is concluded that ICA and IFA algorithms do not correctly unmix hyperspectral data. We give evidence that the unmixing matrix minimizing the mutual information might be far from the true one.

A set of tests with simulated and real hyperspectral data evaluates the performance and illustrates the effectiveness of the proposed methods.

Keywords:

Hyperspectral Imagery; Linear Unmixing; Unsupervised Endmember Extraction; Signal Subspace Estimation; Dependent Component Analysis; Spectral Mixing Model.

Resumo

A tese aborda o problema da separação de materiais em imagens hiper-espectrais. Este problema consiste na decomposição de cada vector espectral (pixel) de uma dada imagem num conjunto de vectores espectrais (designadas por assinaturas) e nas respectivas concentrações (designadas por abundâncias). A separação de dados hiper-espectrais tem ganho relevância com o desenvolvimento de sensores de detecção remota de alta resolução espectral, denominados por sensores hiper-espectrais, com centenas de bandas contíguas de alta resolução espectral, tipicamente $10nm$, no espectro óptico e infra-vermelho ($0.3\mu m$ a $2.5\mu m$), permitindo assim distinguir elementos na superfície terrestre com base nas suas assinaturas espectrais. A informação gerada por estes sensores explora o facto de todas as substâncias dispersarem energia electromagnética, em diferentes comprimentos de onda, consoante a sua composição molecular. Em muitos casos a resolução espacial destes sensores é de 10 a 20 metros. Em consequência, o vector de observação associado a cada pixel é uma mistura de vários elementos com concentrações distintas. Em determinadas circunstâncias estas substâncias podem ser separadas, sem que para isso se necessite de aumentar a resolução espacial do sensor. A tese propõe três novas metodologias de separação de materiais na cadeia de processamento da separação de dados hiper-espectrais.

O primeiro método, designado por *hyperspectral signal identification by minimum error* (HySime), estima o sub-espaço de sinal com vista à redução da dimensão dos dados a serem processados. Dado que o número de bandas (tipicamente maior que 200) é muito superior à ordem do sub-espaço de sinal, a redução da dimensão dos dados permite ganhos na relação sinal-ruído e elevadas reduções

na memória utilizada no armazenamento de dados.

O segundo método, designado por *vertex component analysis* (VCA), determina assinaturas com base nas propriedades geométricas dos dados hiper-espectrais. O método VCA determina os vértices do simplex que contém os dados baseando-se na hipótese de que existe pelo menos um elemento puro de cada material presente no conjunto de dados. Para um desempenho ao nível do estado-da-arte, o VCA tem uma complexidade computacional entre uma a duas ordens de magnitude mais baixa do que os melhores métodos.

O terceiro método, designado por *dependent component analysis* (DECA), formaliza a separação de dados espectrais como um problema de máxima verosimilhança (ML). O algoritmo *expectation-maximization* (EM) é usado para inferir a matriz de separação. Este método modela as densidades das abundâncias como misturas de Dirichlet, impondo assim as restrições habitualmente presentes em dados hiper-espectrais, ou seja, soma constante e não negatividade. Este algoritmo apresenta a vantagem de não necessitar da presença de pixeis puros nos dados.

A tese estuda ainda o impacto que a dependência estatística exibida pelos dados hiper-espectrais tem no desempenho dos algoritmos de análise de componentes independentes (ICA) e na análise de factores independentes (IFA), quando aplicados ao problema da separação de dados hiper-espectrais. Conclui-se que estes algoritmos não separam correctamente os dados hiper-espectrais e que a matriz de separação pode estar muito distante da solução correcta.

Na tese apresenta-se um conjunto de testes com dados hiper-espectrais simulados e reais para avaliar o desempenho das metodologias propostas e ilustrar a sua efectividade.

Palavras-Chave: Imagens Hiper-espectrais; Separação Linear; Extracção Não Supervisionada; Estimação do Sub-espaço de Sinal; Análise de Componentes Dependentes; Modelo de Mistura Espectral.

Agradecimentos

Em primeiro lugar, quero agradecer ao Prof. José Bioucas Dias, meu orientador, pela sua paciência, disponibilidade permanente e envolvimento constante na orientação desta tese. A sua competência, sensibilidade, rigor científico e espírito crítico foram muito gratificantes durante este trabalho.

Quero agradecer ao Instituto de Telecomunicações pelos apoios logístico e financeiro que me foram proporcionados e ao Departamento de Engenharia de Electrónica e Telecomunicações e de Computadores do Instituto Superior de Engenharia de Lisboa pelo apoio financeiro, para a participação em conferências e publicação de artigos, e pelas facilidades concedidas.

Ao Prof. Mário Figueiredo por todos os bons conselhos e incentivos na fase inicial da tese e pela sua amizade. Aos colegas e amigos que comigo partilharam o espaço, conversas e alguns bons momentos no Instituto de Telecomunicações. Em particular ao Paulo Marques pelo seu bom senso e incentivo que me deu ao longo destes anos e ao Martijn Kuipers pela leitura atenta da tese e pelas suas sugestões.

A todos os colegas do Instituto Superior de Engenharia de Lisboa que me apoiaram e incentivaram a percorrer este trajecto. Em particular aos colegas da Secção de Comunicações e Processamento de Sinal e em especial ao Arnaldo, Meneses e Amaral pela sua determinação para a criação de condições que tornaram possível este trabalho.

Aos amigos do CEDET quero agradecer todo o estímulo e incentivo que me deram durante a execução deste trabalho.

Um agradecimento muito especial aos meus pais e irmãos, pela forma como souberam apoiar-me, são os primeiros responsáveis pelo meu trajecto, sobretudo ao Jacinto pelos momentos partilhados, cheios de discussões e desafios gratificantes.

Deixo para último, o eterno agradecimento à Guida, ao Jonas, ao Ian e à Iara por toda a sua compreensão e incentivos inexcedíveis que me deram ao longo desta fase.

To Guida.

Table of Contents

List of Tables

List of Figures

List of Algorithms

List of Acronyms

AFAE	Abundance Fraction Angle Error
AIC	Akaike Information Criterion
AMEE	Automatic Morphological Endmember Extraction
ATREM	Atmospheric Removal
AVIRIS	Airborne Visible/Infrared Imaging Spectrometer
CCA	Convex Cone Analysis
DECA	Dependent Component Analysis
ELM	Empirical Line Method
EM	Expectation Maximization
FastICA	Fast Independent Component Analysis
FFC	Flat Field Correction
FLAASH	Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes
GEM	Generalized Expectation Maximization
HFC	Harsanyi, Farrand, and Chang algorithm
Hydice	Hyperspectral Digital Imagery Collection Experiment
HyMap	Hyperspectral Mapping
HySime	Hyperspectral Signal Identification by Minimum Error
IARR	Internal Average Relative Reflectance
ICA	Independent Component Analysis
ID	Intrinsic Dimensionality
IEA	Iterative Error Analysis
IFA	Independent Factor Analysis

IFOV	Instantaneous Field of View
i.i.d.	Independently and Identically Distributed
JADE	Joint Approximate Diagonalization of Eigen-matrices
MAP	Maximum a Posteriori
MDL	Minimum Description Length
MESMA	Multiple Endmember Spectral Mixture Analysis
MEST	Manual Endmember Selection Tool
ML	Maximum Likelihood
MM	Majorization Maximization
MNF	Maximum Noise Fraction
MOD	Mixtures of Dirichlets
MODTRAN	Moderate Resolution Transmittance
MOG	Mixture of Gaussians
MSE	Mean Squared Error
MVT	Minimum Volume Transform
NAPC	Noise Adjusted Principal Components
N-FINDR	N-Finder Algorithm
NND	Nearest Neighbor Difference
NWHFC	Noise-Whitened HFC
ORASIS	Optical Real-time Adaptive Signature Identification System
OSP	Orthogonal Subspace Projection
PCA	Principal Component Analysis
PPI	Pixel Purity Index
PSF	Point Spread Function
RMS	Root Mean Square
SAE	Signature Angle Error
SAM	Spectral Angle Mapper

SID	Spectral Information Divergence
SNR	Signal-to-Noise Ratio
SVD	Singular Value Decomposition
USGS	U.S. Geological Survey
VCA	Vertex Component Analysis
VD	Virtual Dimensionality

List of Symbols

General Notation

a, A	lowercase and uppercase lightface letters denote scalars
$\mathbf{a}$	lowercase boldface letters denote column vectors
$\mathbf{A}$	uppercase boldface letters denote matrices
$\equiv$	equivalent to, by definition
$\simeq$	approximated to
$a \sim F$	a has distribution F
$\sum, \sum_{j=1}^{N}$	summation
$\prod, \prod_{j=1}^{N}$	product
log	natural logarithm
$\log_{10}$	logarithm to the base 10
$\lvert a \rvert$	absolute value of a
$\max(\mathbf{a})$	maximum of value of a vector $\mathbf{a}$
$\lVert \mathbf{a} \rVert$	Euclidean norm of a vector $\mathbf{a}$
$\mathbf{a} \succeq 0$	components of vector $\mathbf{a}$ are nonnegative
$\mathbb{T}[\mathbf{a}], \bar{\mathbf{a}}$	sample average of $\mathbf{a}$, *i.e.*, $1/N \sum_{i=1}^{N} a_i$, where a_i are components of $\mathbf{a}$

Sets and Spaces

$\{x : P\}$	set of values x with property P
{i,. . . ,j}	set of integers between i and j
$\mathbb{R}$	Real numbers
$\mathbb{R}^L$	L-dimensional Euclidean space
$\langle \mathbf{A} \rangle$	subspace spanned by column vectors of matrix $\mathbf{A}$
$\langle \mathbf{A} \rangle^{\perp}$	orthogonal complement of subspace $\langle \mathbf{A} \rangle$

General Matrices and Vectors

$[a_1, a_2, \ldots, a_N]$	$(1 \times N)$ vector with components a_i, $i = 1, \ldots, N$
$[\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N]$	matrix with columns $\mathbf{a_i}$, $i = 1, \ldots, N$
$[\mathbf{A}]_{i,j}$	element in the ith row and jth column of a matrix $\mathbf{A}$
$[\mathbf{A}]_{:,j}$	jth column of a matrix $\mathbf{A}$
$[\mathbf{A}]_{:,i:j}$	submatrix of a matrix $\mathbf{A}$ consisting of columns between ith and jth
$[\mathbf{A}]_{j,:}$	jth row of a matrix $\mathbf{A}$
$[\mathbf{A}]_{i:j,:}$	submatrix of a matrix $\mathbf{A}$ consisting of rows between ith and jth
$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$	partitioned matrix, consisting of submatrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$

Special Matrices and Vectors

$\mathbf{I}_N$	$(N \times N)$ identity matrix
$\mathbf{1}_N$	$(N \times 1)$ vector of ones
$\mathbf{0}_N$	$(N \times 1)$ vector of zeros

Matrix Transformations and Functions

$(\cdot)^T$	transpose operator
$\mathbf{A}^{-1}$	inverse of a matrix $\mathbf{A}$
$\mathbf{A}^{-T}$	transpose-inverse of a matrix $\mathbf{A}$
$\mathbf{A}^{\#}$	pseudo-inverse of a matrix $\mathbf{A}$, *i.e.*, $\mathbf{A}^{\#} \equiv (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$
$\det(\mathbf{A})$	determinant of a matrix $\mathbf{A}$
$\mathrm{tr}(\mathbf{A})$	trace of a matrix $\mathbf{A}$
$\mathrm{diag}(\mathbf{A})$	vector with the diagonal values of a matrix $\mathbf{A}$
$\mathrm{diag}(a_1, a_2, \ldots, a_N)$	$(N \times N)$ diagonal matrix with value a_i in the ith column and ith row
$\|\mathbf{A}\|_F$	Frobenius norm of a matrix $\mathbf{A}$, *i.e.*,$\|\mathbf{A}\|_F = \sqrt{\mathrm{tr}(\mathbf{A}\mathbf{A}^T)}$

Special Functions

$\Gamma(\cdot)$	Gamma function
$\mathrm{psi}(\cdot)$	psi function (derivative of the logarithm of the Gamma function)
$\mathrm{E}[\cdot]$	expectation operator
$\mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$	Normal distribution with mean $\boldsymbol{\mu}$ and covariance $\mathbf{C}$
$\widehat{\mathbf{A}}$, $\widehat{\mathbf{a}}$, $\widehat{a}$	estimative of a matrix $\mathbf{A}$, a vector $\mathbf{a}$, and a scalar a, respectively.

Overview

Scope

Terrestrial remote sensing imagery involves the acquisition of information from the Earth's surface without physical contact with the area under study. Among the remote sensing modalities, hyperspectral imaging has recently emerged as a powerful passive technology. Hyperspectral sensors sample the scattered or emitted electromagnetic radiation in the portion spectrum extending from the visible region through the near-infrared and mid-infrared (wavelengths between $0.3\mu m$ and $2.5\mu m$), in hundreds of narrow (on the order of $10nm$) contiguous bands [106, 168]. This capability enables the acquisition of images where each pixel is a vector with high spectral resolution.

This unprecedent spectral resolution has opened the door to a series of civilian and military applications among which we refer to: land use, agriculture assessment, ecological and environmental monitoring, ground-cover classification, mineral exploitation, change detection, man-made materials identification and detection, target detection, target activities, and surveillance [152, 153]. Underlying all these applications is the fact that all substances scatter electromagnetic energy, at specific wavelengths, in distinctive patterns related to their molecular composition [66].

The majority of image processing and analysis algorithms dealing with hyperspectral imagery can be classified as follows [56, 157]: detect known or unknown objects or materials in a given scenario; classify/segment the image into regions where a material is predominant; estimate the materials and the respective area fractions that they occupy within a pixel, the so-called hyperspectral unmixing.

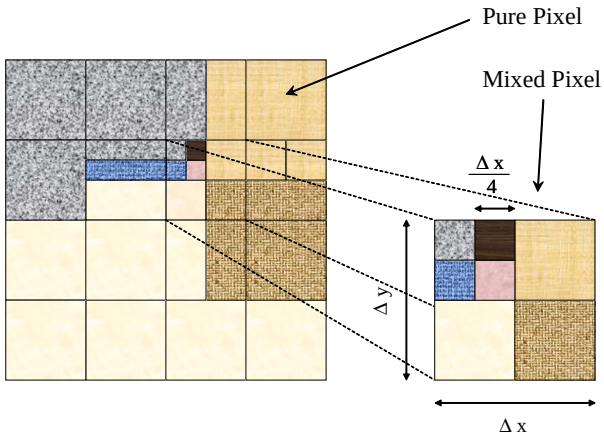


Figure 1: Spectral resolution versus spatial resolution.

Spectral Resolution Versus Spatial Resolution

This thesis addresses the hyperspectral unmixing problem, which is the decomposition of the pixel spectrum into a collection of constituent spectra, or endmember spectral signatures, and their corresponding abundance fractions [86, 142]. Although important in it's own right, unmixing also permits the change of spectral resolution with spatial resolution.

To illustrate this, let's assume that a given spectral vector $\mathbf{r} \in \mathbb{R}^L$ (L is the number of bands) is a linear combination of $\mathbf{m}_j$ signatures, weighted by the respective abundance fractions α_j, for $j = 1, \ldots, p$. Thus,

$$\mathbf{r} = \mathbf{M}\boldsymbol{\alpha},$$

where $\mathbf{M} = [\mathbf{m}_1, \ldots, \mathbf{m}_p]$ is the mixing matrix and $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_p]^T$ is the abundance fractions vector.

Fig. 1 illustrates the concept of pure and mixed pixels. Suppose we want to resolve the material in brown. The spatial resolution of the sensor shall be increased from (Δx $\times$ Δy) to (Δx/4 $\times$ Δx/4), *i.e.*, 16 times. On the other hand, if the mixing matrix $\mathbf{M}$ can be inferred and the columns of $\mathbf{M}$ are linearly independent, then $\boldsymbol{\alpha}$ can be computed and thus the materials can be resolved within pixels without increasing the spectral resolution; *i.e.*, $\boldsymbol{\alpha} = \mathbf{M}^{\#}\mathbf{r}$ ($\mathbf{M}^{\#}$ stands for the pseudo-inverse

of $\mathbf{M}$). Of course, a necessary condition for the columns of $\mathbf{M}$ to be independent is that $L \geq p$. As a conclusion, hyperspectral sensors offer the possibility of changing spatial resolution with spectral resolution.

The primary objective of this thesis is to recover the matrix $\mathbf{M}$ and the vector $\boldsymbol{\alpha}$ for each pixel of the hyperspectral image. This is a blind source separation problem to which many methods and frameworks have been developed [6, 41, 74, 117]. However, most of them do not apply directly to hyperspectral data, since the abundance fractions have particular constraints due to the physics of the acquisition process, namely: i) $\mathbf{m}_j \succeq 0$; ii) $\alpha_j \in \Delta_p$, where

$$\Delta_p \equiv \{\boldsymbol{\alpha} \in \mathbb{R}^p : \alpha_j \geq 0, \sum_{j=1}^{p} \alpha_j = 1\}.$$

Thesis Contributions

This section summarizes the topics addressed in the thesis and its main contributions. These topics are schematized in Fig. 2.

Linear Mixture Model

Chapter 1 describes a linear model for the observed spectral vectors. The model accounts for the degradation mechanisms normally found in hyperspectral data, namely, signature variability, topographic modulation, and sensor noise. This model is part of the material published in [119, 122, 124, 128].

Hysime, an Hyperspectral Signal Subspace Estimator

Chapter 2 proposes a new method termed *hyperspectral signal identification by minimum error* (HySime). The method is based on a mean squared error (MSE) approach to determine the

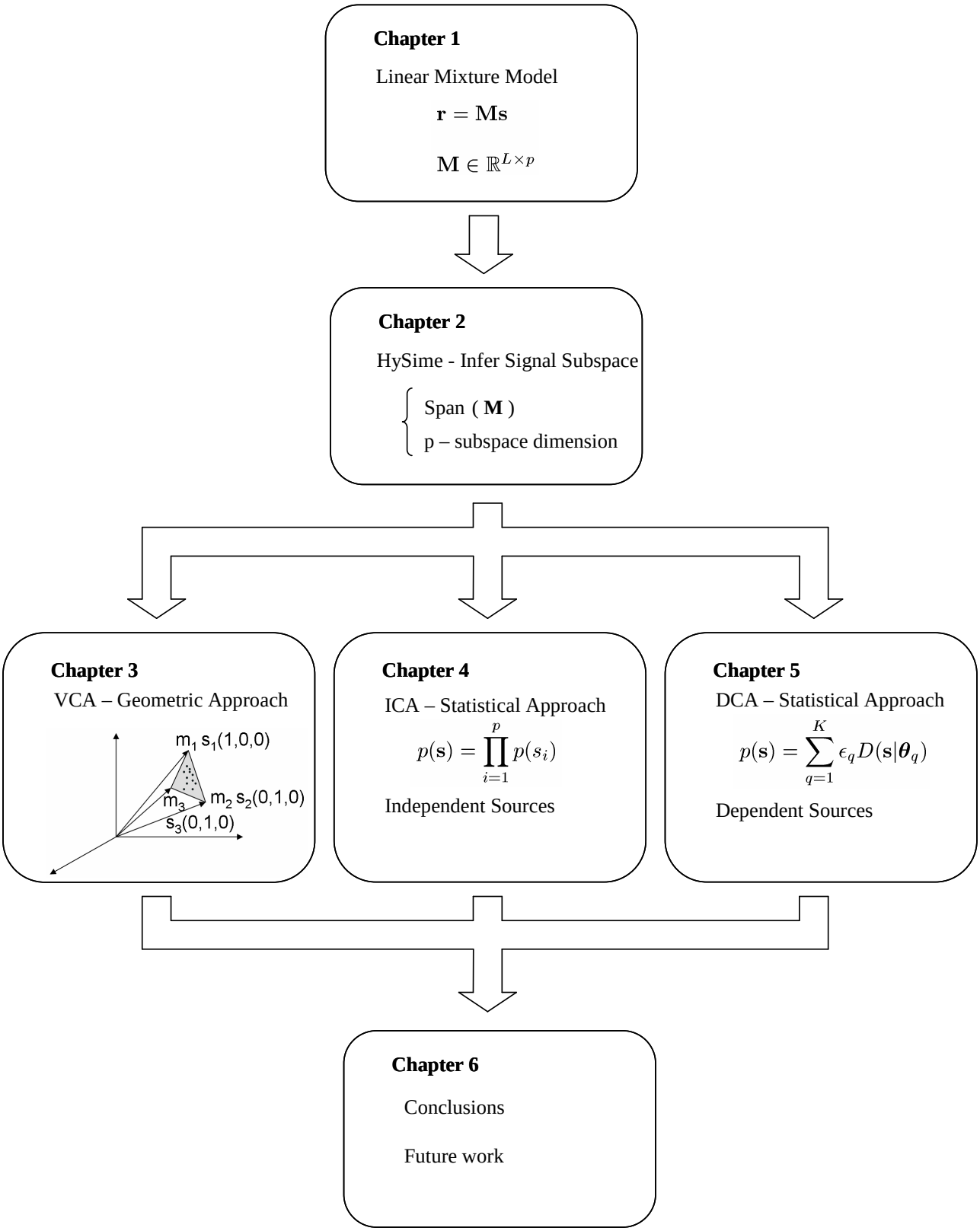


Figure 2: Thesis organization.

signal subspace in hyperspectral imagery. HySime is an eigendecomposition based method; it first estimates the signal and noise correlations matrices, then it selects the subset of eigenvalues that best represents the signal subspace in the least square sense. A variation of this method, exploiting the fact that hyperspectral mixtures are nonnegative, is also proposed.

The noise estimation, a necessary step in the HySime algorithm, is based on the multiple regression theory and exploits the high correlation existing between contiguous hyperspectral bands. This method has been optimized to reduce the computational complexity. Parts of HySime were published in [16, 18, 125, 126].

VCA, A Fast Algorithm to Unmix Hyperspectral Data

Chapter 3 proposes a new unsupervised method for linear unmixing termed *vertex component analysis* (VCA). VCA is a fast method based on the following hyperspectral geometric features: i) the endmembers are the vertices of a simplex; ii) the affine transformation of a simplex is also a simplex. Briefly, VCA algorithm iteratively projects the spectral vectors onto a direction orthogonal to the subspace spanned by the endmembers already determined. The new endmember signature corresponds to the extreme of the projection. This algorithm competes with the current state-of-the-art methods, yet it has much lower computational complexity. Parts of this work were published in [18, 120, 121, 123, 127].

The Role of ICA on Hyperspectral Data Unmixing

VCA, as other geometrical approaches, assumes the presence of pure pixels on the data set. This is a strong requirement that may not hold in some hyperspectral data sets. Statistical methods, such as independent component analysis (ICA) or independent factor analysis (IFA), do not require such assumption. However, hyperspectral mixtures have statistical dependence among the endmembers' abundance fractions. Chapter 4 addresses the impact of the source (*i.e.*, abundance

fractions) statistical dependence on unmixing hyperspectral data using ICA and IFA algorithms. It is shown that these algorithms do not correctly unmix hyperspectral data. We give evidence that the unmixing matrix minimizing the mutual information might be very far from the true one.

A method based on the source entropy to sort the output of ICA or IFA algorithms according to the likelihood of being correctly separated is also proposed. Parts of the study were published in [119, 122, 124, 128].

Dependent Component Analysis

Chapter 5 introduces a new direction to blindly unmix hyperspectral data, termed *dependent component analysis* (DECA). This method ensures abundance fractions constraints by using mixtures of Dirichlet (MOD) densities and it uses an *expectation maximization* (EM) algorithm to estimate the unmixing matrix and the Dirichlet parameters. Compared with the geometric based approaches, the proposed direction does not need to assume the presence of pure pixels in the observations. Parts of this work were published in [17, 128].

Thesis Organization

The thesis is organized as follows: Chapter 1 describes the hyperspectral unmixing problem and the hyperspectral linear mixture model; Chapter 2 presents the *hyperspectral signal identification by minimum error* (HySime) method to determine the signal subspace of the hyperspectral image. Chapter 3 presents the *vertex component analysis* (VCA), a new geometrical method to unmix hyperspectral data. Chapter 4 addresses the impact of the source statistical dependence on unmixing hyperspectral data using ICA and IFA frameworks. Chapter 5 presents the *dependent component analysis* (DECA) method to blindly unmix hyperspectral data. Chapter 6 concludes the thesis summarizing the main contributions and presenting future research directions.

Chapter 1

Introduction

1.1 Hyperspectral Data Acquisition

The science of remote sensing has advanced over the recent past by using increasingly capable sensors. The development of an extremely powerful class of Earth remote sensing instruments has improved the capability of ground-based data collection in the fields of agriculture, geography, geology, mineral identification, detection [33, 108], and classification [13, 97, 109, 115, 153].

Hyperspectral imaging sensors collect two dimensional spatial images over many contiguous bands of high spectral resolution covering the visible, near-infrared, and shortwave infrared spectral bands [106, 168]. Table 1.1 presents five examples of aircraft and satellite hyperspectral sensors and their principal characteristics.

Table 1.1: Hyperspectral imaging sensors.

	AVIRIS [62]	HyDICE [143]	Hymap [134]	Probe-1 [104]	Hyperion [167]
Introduction Year	1987	1995	1996	1997	2000
Platform	airborne	airborne	airborne	airborne	spaceborn
Nominal Altitude (km)	20	6	5	2.5	705
Spatial Resolution (m)	20	3	10	5	30
Spectral Resolution (nm)	10	10	17	10	10
Spectral Coverage (μm)	0.4-2.5	0.4-2.5	0.4-2.5	0.4-2.5	0.4-2.5
Number of Channels	224	210	128	128	220
Swath Width (km)	12	0.9	6	3	7.7

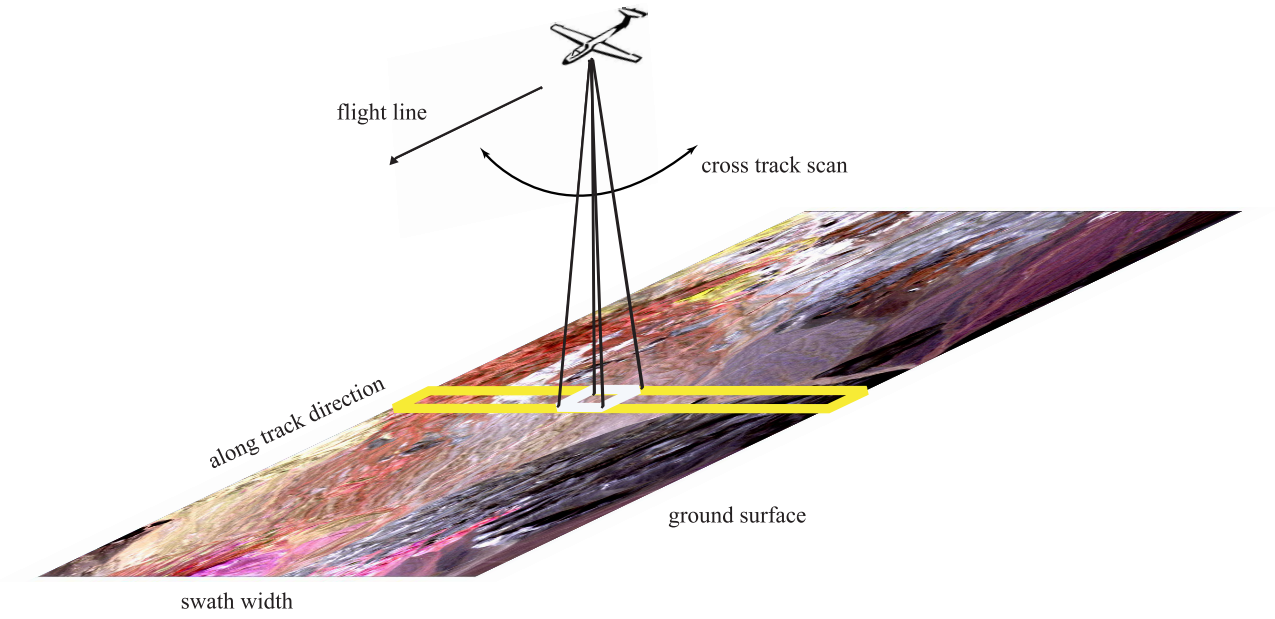


Figure 1.1: Principle of hyperspectral data acquisition.

The AVIRIS [62, 168], Probe-1 [37, 104], and Hymap [134] instruments perform the collection of data in a whisk-broom mode to the cross-track direction by mechanical scanning and in the along-track direction by movement of the platform. Hyperion [135, 136, 167] and HyDICE [143] instruments use a push-broom imaging sensor, which acquires data in a cross-track line without any mechanical scanning. Usually an optical dispersing element such as a grating or a prism splits the received ground surface radiance from the spectrometer into discrete channels registered by several line array detectors. The digital data is produced by an analog to digital converter, which samples the radiance measured in each spectral channel with a given radiometric resolution.

Fig. 1.1 illustrates the principles involved in the hyperspectral data acquisition. The spatially and spectrally sample information of the ground surface can be described by a three dimensional structure, referred to as a data cube. Fig. 1.2 shows an example of such a data cube and the radiance of a pixel vector as a function of the wavelength. The ground pixel size varies from meters to tens of meters depending on the altitude of the platform and on the instantaneous field of view (IFOV) of the sensor.

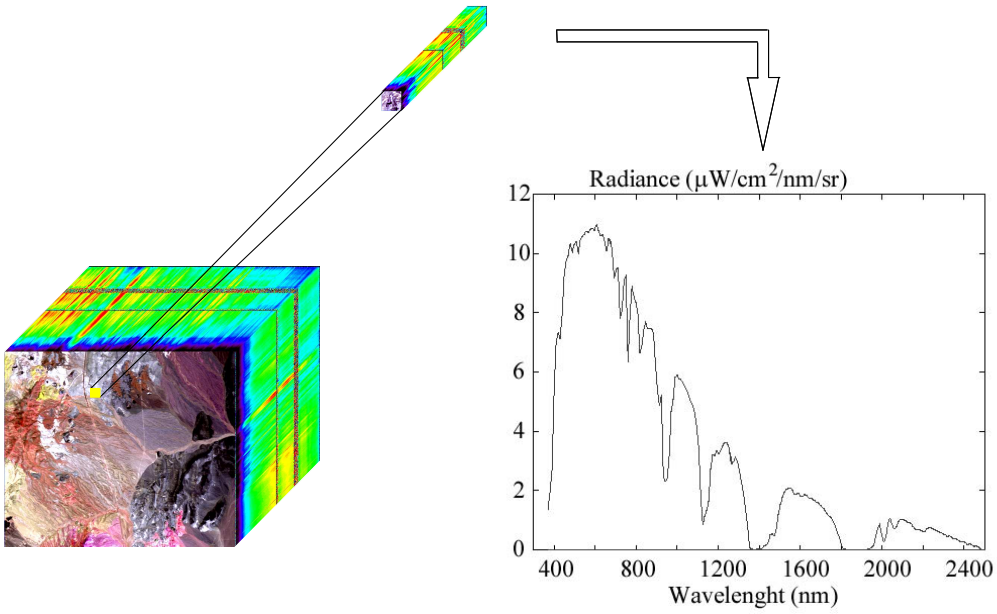


Figure 1.2: Hyperspectral data cube and pixel radiance example.

The information content of hyperspectral images with thousands of pixels and hundreds of channels allows us to remotely identify ground materials, based on their spectral signature, and to perform land characterization based on the abundance of each material.

Very often, the spatial portion occupied by a substance is smaller than the ground pixel size (tens of meters). As a result, the signal read by the sensor from a given spatial element of resolution and at a given spectral band is a mixing of components originated from the constituent substances, termed endmembers, located at that element of resolution [106]. In this situation, the scattered energy is a mixing of the endmember spectrum [71, 137, 151]. Hyperspectral unmixing is the decomposition of the pixel spectrum into a collection of constituent spectra, or spectral signatures, and their corresponding abundance fractions, occasionally termed sources, indicating the proportion of each endmember present in the pixel [86, 142].

Depending on the mixing scales at each pixel, the observed mixture is either linear or nonlinear [87, 105]. The linear mixing model holds approximately when the mixing scale is macroscopic [156] and there is negligible interaction among distinct endmembers [39, 65]. Fig. 1.3(a) illustrates a

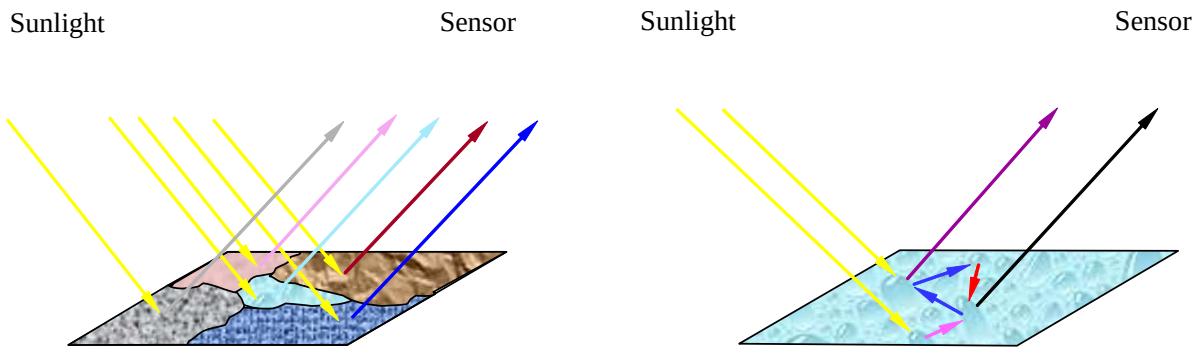


Figure 1.3: Illustration of a pixel: (a) linear mixture (macroscopic mixing scale); (b): nonlinear mixture (intimate mixture).

linear mixing scale. The nonlinear model holds when the mixing scale is microscopic (*i.e.*, intimate mixtures) [129, 155]. Fig. 1.3(b) illustrates an intimate mixture, yielding a nonlinear scenario. The linear model assumes negligible interaction among distinct endmembers [39, 65]. The nonlinear model assumes that incident solar radiation is scattered by the scene through multiple bounces involving several endmembers [23]. Very often, the effects of multiple scattering are assumed to be negligible and thus the linear model is adopted [110].

Fig. 1.4 shows the most common steps to unmix hyperspectral data. Due to illumination and atmospheric effects, the radiance acquired by hyperspectral sensors cannot be directly compared with a digital spectral library or even with other radiance data sets. This comparison is made possible by the atmospheric correction, which transforms the radiance spectra into reflectance. This operation accounts for solar spectrum, sensor and sun directions, path radiance, secondary illumination, and shadowing. The second operation, data reduction, is a consequence of the fact that the number of endmembers present in the scene is usually much smaller than the number of bands of an hyperspectral data set. This operation has a great impact since it reduces the amount of data, implying computational savings in the unmixing step, and it improves the signal-to-noise ratio (SNR). The third operation, spectral unmixing, usually embodies two steps: endmember determination and inversion. The first step estimates the signatures of the distinct endmembers

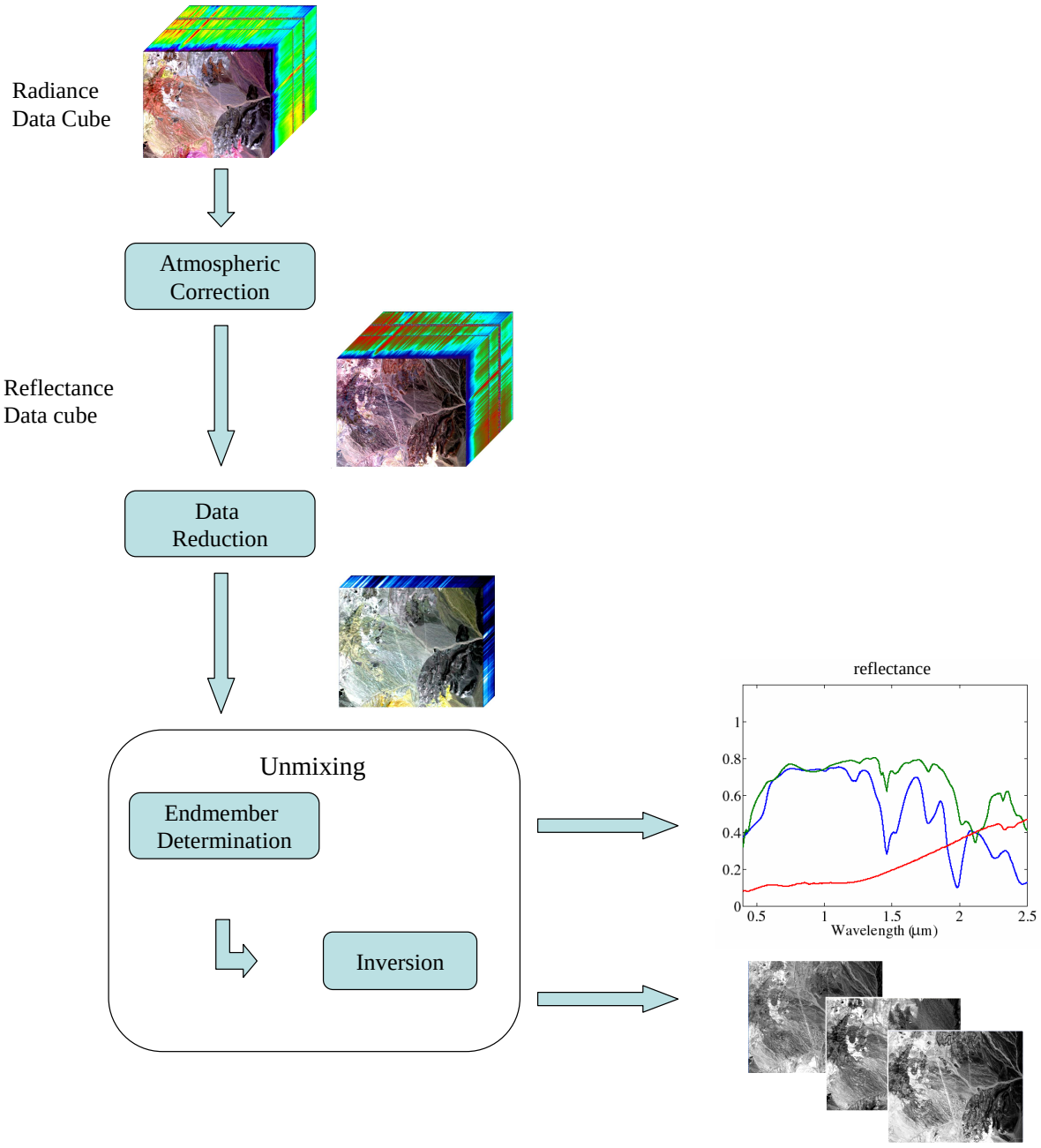


Figure 1.4: Schematic diagram of hyperspectral unmixing.

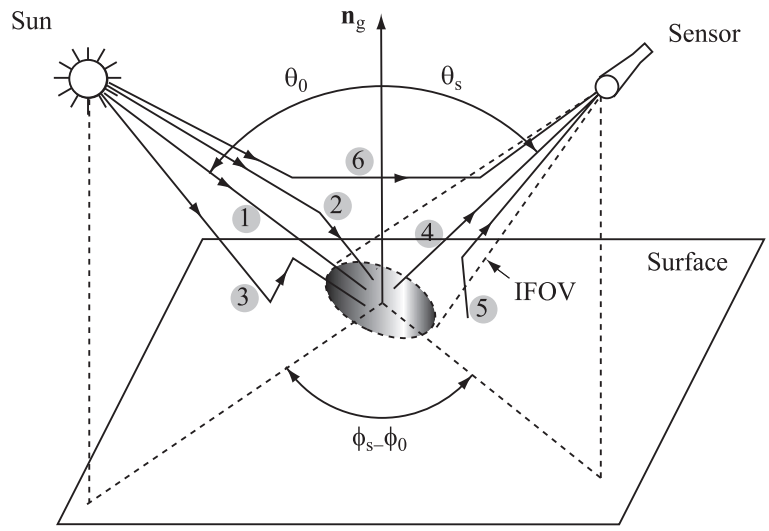


Figure 1.5: Schematic diagram of the main contributions to the radiance read by the sensor in the solar spectrum.

present in the scene. The second step estimates the abundance fractions of each endmember.

To conduct the hyperspectral unmixing operation, a mixture model must be adopted to describe how the constituent endmembers and how the atmosphere scatters the sun light at a given pixel. This process is addressed in the next section.

1.2 Spectral Radiance Model

This section describes a typical passive remote sensing scenario (see Fig. 1.5). The sun illuminates a random media formed by the earth surface and by the atmosphere; a sensor (airborne or spaceborn) reads, within its instantaneous field of view, the scattered radiance in the solar-reflectance region extending from 0.4 to 2.5 μm, encompassing the visible, near-infrared, and shortwave infrared bands. Angles θ and ϕ, with respect to the normal $\mathbf{n}_g$ on the ground, are the colatitude and the longitude, respectively. The solar and sensor directions are (θ_0, ϕ_0) and (θ_s, ϕ_s), respectively.

The total radiance at the surface level is the sum of three components, as schematized in Fig. 1.5: the sunlight (ray 1), the skylight (ray 2), and the light due to the *adjacency effect* (ray 3), *i.e.*,

due to the successive reflections and scattering between the surface and the atmosphere. Following [163, 169], the spectral radiances of these components are, at a given wavelength λ, respectively, given by

$\mathcal{L}_1 = \mu_0 \mathcal{E}_0 T_\downarrow$, where $\mathcal{E}_0$ is the solar flux at the top of the atmosphere, $\mu_0 = \cos(\theta_0)$, and $T_\downarrow = T_\downarrow(\theta_0)$ is the downward transmittance.

$\mathcal{L}_2 = \mu_0 \mathcal{E}_0 t_\downarrow$, where $t_\downarrow = t_\downarrow(\theta_0)$ is the downward diffuse transmittance factor.

$\mathcal{L}_3 = \mu_0 \mathcal{E}_0 T'_\downarrow [\rho_t S + (\rho_t S)^2 + (\rho_t S)^3 \dots]$, where $T'_\downarrow = [T_\downarrow + t_\downarrow]$, ρ_t is the mean reflectance of the surroundings with respect to the atmospheric point spread function (PSF), and S is the spherical albedo of the atmosphere.

The total radiance incident upon the sensor location is the sum of three components: the light scattered by the surface (ray 4), the light scattered by the surface and by the atmosphere (ray 5), and light scattered by the atmosphere (ray 6), the so-called *path radiance*. Assuming a Lambertian surface, and again following [163, 169], these radiances at the top of the atmosphere are, at wavelength λ, respectively, given by

$\mathcal{L}_4 = \dfrac{\mu_0 \mathcal{E}_0}{\pi} \dfrac{T'_\downarrow T_\uparrow}{1 - \rho_t S} \rho$, where ρ is the surface reflectance and $T_\uparrow = T_\uparrow(\theta_0)$ is the upward transmittance.

$\mathcal{L}_5 = \dfrac{\mu_0 \mathcal{E}_0}{\pi} \dfrac{T'_\downarrow t_\uparrow}{1 - \rho_t S} \rho_t$, where $t_\uparrow = t_\uparrow(\theta_s)$ is the upward diffuse transmittance factor.

$\mathcal{L}_6 = \dfrac{\mu_0 \mathcal{E}_0}{\pi} \rho_a$, where $\rho_a(\theta_0, \theta_s, \phi_s - \phi_0)$ is the atmosphere reflectance.

Thus, the total radiance, $\mathcal{L}_s$, incident upon the sensor location is

$$\mathcal{L}_s = a\rho + b,$$

where

$$a = \frac{\mu_0 \mathcal{E}_0}{\pi} \frac{T'_{\downarrow} T_{\uparrow}}{1 - \rho_t S} \tag{1.1}$$

$$b = \frac{\mu_0 \mathcal{E}_0}{\pi} \left(\frac{T'_{\downarrow} t_{\uparrow}}{1 - \rho_t S} \rho_t + \rho_a \right). \tag{1.2}$$

Let us assume that the sensor has L channels (wavebands). Assuming linear receivers and narrow wavebands, the signal at the output of the ith channel (waveband centered at wavenumber λ_i) is given by

$$r_i = c_i \rho + d_i + n_i,$$

where c_i and d_i are proportional to $a(\lambda_i)$ and $b(\lambda_i)$, respectively, and n_i denotes the receiver electronic noise at channel i plus the Poisson (photonic) signal noise (see, *e.g.*, [79]).

Terms a and b in expressions (1.1) and (1.2) depend in a complex way on the sun and sensor directions, on the atmosphere composition, on the topography, and on the scene materials and configurations [107, 163, 169]. The compensation for these terms, the so-called atmospheric correction, is a necessary step in many quantitative algorithms aiming at extracting information from multispectral or hyperspectral imagery [14, 146, 169].

In the thesis, linear unmixing of abundance fractions at the pixel level is addressed. The term linear indicates that the observed entities are linear combinations of the endmember spectral signatures weighted by the correspondent abundance fractions. Therefore, it is assumed that atmospheric correction has been applied to ensure a linear relation between the radiance $\mathcal{L}_s$ and the reflectance ρ; *i.e.*, for each channel, the relation between the radiance and the reflectivity is linear with coefficients independent of the pixel.

A brief description of the most used atmospheric correction methods is presented in Appendix A. Note, however, that no correction may be necessary. This is the case when the scene is a surface of approximately constant altitude, the atmosphere is horizontally homogeneous, and ρ_t, the mean reflectance of the surroundings, exhibits negligible variation.

1.2.1 Linear Spectral Mixture Model

In spectral mixture modeling, the basic assumption is that the surface is made of a few number of endmembers with relatively constant spectral signatures, or, at least, constant spectral shapes. If the multiple scattering among distinct endmembers is negligible and the surface is partitioned according to the abundance fractions, then the spectral radiance upon the sensor location is well approximated by a linear mixture of endmember radiances weighted by the correspondent abundance fractions [2, 66, 86, 87, 105].

Under the linear mixing model and assuming that the sensor radiation pattern is ideal (*i.e.*, constant in the IFOV and zero outside), the output of channel i from a given pixel is

$$r_i = c_i \sum_{j=1}^{p} \rho_{ij}\alpha_j + d_i + n_i, \tag{1.3}$$

where ρ_{ij} denotes the reflectance of endmember j at wavenumber λ_i, α_j denotes the abundance fraction of endmember j at the considered pixel, and p is the number of endmembers.

Abundance fractions $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_p]^T$ as the name indicates is a vector of fractions belonging to the simplex Δ_p given by

$$\Delta_p \equiv \{\boldsymbol{\alpha} \in \mathbb{R}^p : \alpha_j \geq 0, \sum_{j=1}^{p} \alpha_j = 1\}, \tag{1.4}$$

where $\sum_{j=1}^{p} \alpha_j = 1$ and $\alpha_j \geq 0$, for $j = 1, \ldots, p$ are the so-called full additivity and positivity constraints, respectively. For a real sensor, the output of channel i is still formally given by

expression (1.3), but α_j depends on the sensor PSF $h_{x,y}(u,v)$ according to

$$\alpha_j = \frac{\int_{A_j} h_{x,y}(-u,-v)\,dudv}{\int_{\mathbb{R}^2} h_{x,y}(-u,-v)\,dudv},$$

where A_j denotes the set of points on the surface belonging to the jth endmember. The PSF $h_{x,y}(u,v)$ may be spatially variant (*i.e.*, it depends on surface coordinates x and y) and it includes the effect of finite aperture and receiver impulse response. Note that the meaning of α_j depends not only on the true abundance fractions, but also on the endmember distribution inside the IFOV and on the surface point (x,y). An image-restoration-based approach to mitigate the errors introduced by the non-ideal nature of the PSF is proposed in [175].

Herein it is assumed that the signal at the output of channel i is given by expression (1.3). A usual procedure is to remove low-SNR channels from the spectral vectors. Thus, since only high-SNR channels are considered, Poisson noise is neglected. Concerning the atmospheric correction, it is assumed that the images have been processed to remove the path radiance (ray 5 in Fig. 1.5) and the light scattered by interaction between surface and the atmosphere (ray 6 in Fig. 1.5). Note, however, that the corrected image is still in radiance units.

Let $\mathbf{r}$ be an $L\times 1$ vector, where L is the total number of bands and $\mathbf{m}_j^o \equiv [c_1\rho_{1j}, c_2\rho_{2j}, \ldots, c_L\rho_{Lj}]^T$ is the so-called signature of the jth endmember. In this case, expression (1.3) can be written as

$$\mathbf{r} = \mathbf{M}\boldsymbol{\alpha} + \mathbf{n}', \tag{1.5}$$

where $\mathbf{M} \equiv [\mathbf{m}_1^o, \mathbf{m}_2^o, \ldots, \mathbf{m}_p^o]$ is a matrix with the signatures of the endmembers present in the covered area, $\boldsymbol{\alpha} \equiv [\alpha_1, \alpha_2, \ldots, \alpha_p]^T$, and $\mathbf{n}'$ models additive receiver electronic noise.

The model (1.5) is an oversimplification of reality, because it does not take into account signature variability (from pixel to pixel) due to changes in the configuration and in the composition

of substances, surface contaminants, variation in the substances such as age-induced color fading due to oxidation or bleaching, uncompensated atmospheric and environmental effects, and uncompensated errors in the sensor. Signature variability has been studied and accounted for in a few unmixing algorithms (see, *e.g.*, [10, 22, 92]).

Signature variability is primarily characterized by spectral shape invariance [152]; *i.e.*, while the spectral shapes of the endmembers are fairly consistent, their amplitude varies considerably over the scene. Based on this rationale, the spectral variability of the jth endmember at a given pixel is modeled as

$$\mathbf{m}_j = \psi_j \, \mathbf{m}_j^o + \boldsymbol{\eta}_j, \tag{1.6}$$

where $\psi_j \geq 0$ is a scale factor and $\boldsymbol{\eta}_j$ is a zero-mean random vector. Noise $\boldsymbol{\eta}_j$ accounts for signature variability not modeled by ψ_j. Introducing expression (1.6) into expression (1.5), for a given pixel, leads to

$$\mathbf{r} = \mathbf{M}\boldsymbol{\psi}\boldsymbol{\alpha} + \sum_{j=1}^{p} \alpha_j \boldsymbol{\eta}_j + \mathbf{n}', \tag{1.7}$$

where $\boldsymbol{\psi} \equiv \text{diag}\,(\psi_1, \ldots, \psi_p)$ is a $p \times p$ diagonal matrix.

Having in mind that illumination variability due to surface topography affects equally all bands and that additive noise is assumed, the spatial vector associated to a given pixel can be written as

$$\begin{aligned} \mathbf{r} &= \mathbf{M} \underbrace{\gamma \boldsymbol{\psi}\boldsymbol{\alpha}}_{\mathbf{s}} + \underbrace{\gamma \sum_{j=1}^{p} \alpha_j \boldsymbol{\eta}_j + \mathbf{n}'}_{\mathbf{n}} \\ &= \mathbf{M}\mathbf{s} + \mathbf{n}, \end{aligned} \tag{1.8}$$

where γ models the surface topography effect.

Model (1.8) is adopted in the experiments conducted in the following chapters of the thesis. It is still linear and accounts for illumination fluctuations, signature variability, and sensor noise.

Chapter 2

HySime: An Hyperspectral Signal Subspace Estimator

2.1 Introduction

Hyperspectral sensors provide more accurate and more detailed data than their multispectral ancestors. From each pixel, they acquire spectral vectors with hundreds of components, yielding large amounts of data. For example, AVIRIS collects a 512 (along track) $\times$ 614 (across track) $\times$ 224 (bands) $\times$ 12 (bits) data cube in 43 seconds, corresponding to more than 700 Mbits, and Hyperion collects 4 Mbits in 3 seconds, corresponding to 366Kbytes/Km^2 [85]. Such huge data volumes put stringent requirements in what concerns communications, storage, and processing.

Each pixel of an hyperspectral image can be represented as a vector in the Euclidian space $\mathbb{R}^L$, where L is the number of bands and each channel is assigned to one axis of space. Under the linear mixing scenario, the spectral vectors are a linear combination of the so-called endmember signatures. The number of endmembers present in a given scene is, very often, much less than the number of bands L. Therefore, hyperspectral vectors lie in a low dimensional linear subspace. The identification of this subspace enables the representation spectral vectors in a low dimensional subspace, thus yielding gains in computational time and complexity and in data storage. The computation of statistical estimates is a relevant example of the advantages of dimensionality reduction, since the number of samples required to obtain accurate estimates increases drastically

with the dimensionality of the data [97].

2.1.1 Hyperspectral Dimensionality Reduction Methods

Dimensionality reduction has been approached in many ways. Band selection or band extraction, as the name suggests, exploits the high correlation existing between adjacent bands to select a few spectral components among those with higher SNR [7, 72, 78, 154, 160]. The signal subspace methods can be classified either as global or local [80]. The global methods estimate the signal subspace using the complete data set. The local methods use local information contained in pixel neighborhoods, thus avoiding projections of data onto a lower-dimensional subspace. Projection techniques, which are usually used in the global approaches, seek for the best subspace to project data by minimizing an objective function. For example, principal component analysis (PCA) [82], computes the *Karhunen-Loéve* transform, which seeks for the projection that best represents data in the least square sense; singular value decomposition (SVD) [149] provides the projection that best represents data in the maximum power sense; maximum noise fraction (MNF)[61] and noise adjusted principal components (NAPC)[99] seek for the projection that optimizes the ratio of noise power to signal power. This is in contrast with PCA and SVD, where no noise model is used. Orthogonal subspace projection (OSP) [68] is a technique that projects the data set onto a subspace orthogonal to the undesired signatures. This operation is an optimal interference suppression process in the least squares sense.

Topological methods are local approaches that infer the manifold, usually of low dimension, where data set live [28]. For example, curvilinear component analysis [45], curvilinear distance analysis [101], and manifold learning [57] are non-linear projections based on the preservation of the local topology. Independent component analysis [102], projection pursuit [77, 81], and wavelet decomposition [83, 132] have also been considered.

U. S. Naval Research Laboratory developed ORASIS [25, 27] aiming at real-time implementations. This framework consists of several algorithms, where the exemplar selector module uses a non-statistical technique, based on a squared error optimization, to reduce the dimension of hyperspectral data.

The identification of the signal subspace is a model order inference problem to which information theoretic criteria like the *minimum description length* (MDL) [144, 150] or the *Akaike information criterion* (AIC) [4] comes to mind. These criteria have in fact been used in hyperspectral applications [32] adopting the approach introduced by Wax and Kailath in [172].

Harsanyi, Farrand, and Chang [67] developed a Neyman-Pearson detection theory-based thresholding method (HFC) to determine the number of spectral endmembers in hyperspectral data (referred to in [32] as virtual dimensionality - VD). The HFC method uses the eigenvalues to measure signal energies in the detection model. A modified version of this method, termed noise-whitened HFC (NWHFC), includes a noise-whitening process as a preprocessing step to remove the correlation [31]. This method requires a noise estimation step.

2.1.2 Proposed Approach

This chapter proposes two new minimum mean squared error based approaches to determine the signal subspace in hyperspectral imagery. The first method, termed *hyperspectral signal identification by minimum error* (HySime), starts by estimating the signal and the noise correlation matrices using multiple regression. The eigenvectors of the signal correlation matrix are then used to build a sequence of nested subspaces. The signal subspace is inferred by minimizing the sum of the projection error power with the noise power, which are, respectively, decreasing and increasing functions of the subspace dimension. Therefore, if the subspace dimension is overestimated the noise power term is dominant, whereas if the subspace dimension is underestimated the projection error power term is the dominant. The overall scheme is adaptive in the sense that it does not depend on any

tuning parameters. The second method is a mean-value based approach. It exploits the structure of hyperspectral mixtures, namely, the fact that spectral vectors are nonnegative. The effectiveness of the proposed methods is illustrated using simulated and real hyperspectral images.

In spite of very good nonlinear existing methods, it makes sense to research the linear approach because, besides being light from the computational point of view, the linear mixing model is a good approximation in many real scenarios. Even in the nonlinear mixing case, the data is, very often, in a subspace of dimension much lower than the number of bands, *i.e.*, $p \ll L$.

The chapter is structured as follows. Section 2.2 formulates the signal subspace identification problem and reviews the SVD and MNF methods. Section 2.3 describes the fundamentals of the proposed method. Section 2.4 evaluates the proposed algorithm using simulated and real data. Section 2.6 ends the chapter by presenting some concluding remarks. Parts of the proposed approach herein presented were published in [16, 18, 125, 126].

2.2 Problem Formulation and Classical Dimensionality Reduction Methods

Assume that the observed spectral vectors, $\mathbf{r} \in \mathbb{R}^L$, are given by (see Section 1.2.1)

$$\mathbf{r} = \mathbf{x} + \mathbf{n}, \tag{2.1}$$

where $\mathbf{x}$ and $\mathbf{n}$ are L-dimensional vectors standing for signal and additive noise, respectively. Furthermore, assume that signal vectors are in an unknown p-dimensional subspace, *i.e.*,

$$\mathbf{x} = \mathbf{M}\mathbf{s},$$

with $\mathbf{M}$ being a full-rank $L \times p$ matrix. Under the linear mixing scenario, the columns of $\mathbf{M} \equiv [\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_p]$ are the endmember signatures and $\mathbf{s}$ is the abundance fraction vector. To be physically meaningful [108], abundance fractions are subject nonnegativity and full additivity constraints, *i.e.*, $\mathbf{s} \in \Delta_p$ [see expression (1.4)].

Herein, we do not assume any special structure for the scattering mechanism; *i.e.*, our approach works both under the linear and nonlinear scenarios. Even in the nonlinear mixing scenario, it often happens that signal subspace dimension, although larger than the number of endmembers, is much smaller than the number of bands L. In these cases, it is still worthy, then, to estimate the signal subspace and represent the data on it. Note that this procedure does not preclude the application of future nonlinear projection techniques; on the contrary, it is an advantage, since the data is represented by vectors of smaller dimension, thus lightening the computational complexity of any posterior processing scheme.

Let us assume, for a while, that noise $\mathbf{n}$ is zero-mean Gaussian i.i.d., (*i.e.*, the components of $\mathbf{n}$ are independent and identical distributed) with variance σ_n^2 per band. Under these circumstances, the *maximum likelihood* (ML) estimate of the signal subspace is spanned by the p-dominant eigenvectors of the sample correlation matrix of $\mathbf{r}$ [149, Ch. 6]; *i.e.*, $\langle \mathbf{M} \rangle = \langle [\mathbf{e_1}, \ldots, \mathbf{e_p}] \rangle$, (the notation $\langle \mathbf{M} \rangle$ represents the subspace spanned by the columns of $\mathbf{M}$) where $\mathbf{e}_i$, for $i = 1, \ldots, p$, are the p-dominant eigenvectors of the sample correlation matrix $\widehat{\mathbf{K}}_r$.

2.2.1 Eigenanalysis of the Sample Correlation Matrix

The ML estimator of the subspace signal just presented assumes that the dimension of the subspace is known beforehand. However, this dimension is often *a priori* unknown. Nevertheless, a similar approach has been extensively used as a dimensionality reduction tool in hyperspectral image processing [87, 153]. It consists in assuming that the noise is zero-mean i.i.d.. Thus, the correlation matrix of the observed vectors may be written as $\mathbf{K}_r = \mathbf{E}(\mathbf{\Sigma} + \sigma_n^2 \mathbf{I}_L)\mathbf{E}^T$, where $\mathbf{E}$ and $\mathbf{\Sigma}$ are the

eigenvector and eigenvalue matrices of $\mathbf{K}_x$, respectively. Assuming that $\mathbf{K}_x$ has just p positive eigenvalues and that they are ordered along the diagonal of $\mathbf{\Sigma}$ by decreasing magnitude, we have then $\langle \mathbf{M} \rangle = \langle [\mathbf{e}_1 \ldots \mathbf{e}_p] \rangle$; *i.e.*, the estimate of the signal subspace is the span of the eigenvectors of $\mathbf{K}_r$ whose respective eigenvalues values are larger than σ_n^2 [149].

This is, basically, the idea behind SVD-based dimensionality reduction. Two limitations of this approach are the following:

i) the noise present in most hyperspectral data sets is not i.i.d. and, thus, the signal subspace is no longer given by the span of the first p singular vectors nor by any other set of eigenvalues.

ii) even if the noise was i.i.d., the procedure described above to infer the subspace dimension is prone to errors owing to random perturbations always present in the estimates of σ_n^2, $\mathbf{E}$, and $\mathbf{\Sigma}$.

We illustrate these limitations with an experiment built on a simulated hyperspectral image composed of 10^5 pixels and generated according to the linear mixing scattering mechanism. Each pixel is a mixture of five endmembers signatures ($p = 5$) selected from the USGS digital spectral library [40]. Abundance fractions are generated according to a Dirichlet distribution given by

$$D(s_1, s_2, \ldots, s_p | \theta_1, \theta_2, \ldots, \theta_p) = \frac{\Gamma(\sum_{j=1}^{p} \theta_j)}{\prod_{j=1}^{p} \Gamma(\theta_j)} \prod_{j=1}^{p} s_j^{\theta_j - 1}, \tag{2.2}$$

where $\{s_1, \ldots, s_p\} \in \Delta_p$ and the mean value of the jth endmember fraction, s_j, is $\mathrm{E}[s_j] = \theta_j / \sum_{l=1}^{p} \theta_l$ [54]. This density, besides enforcing positivity and full additivity constraints, displays a wide range of shapes, depending on the parameters of the distribution. On the other hand, as noted in [116], the Dirichlet density is suited to model fractions.

Consider that the noise correlation matrix is $\mathbf{K}_n = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_L^2)$ and that the diagonal

elements follow a Gaussian shape centered at the band $L/2$, *i.e.*,

$$\sigma_i^2 = \sigma^2 \frac{e^{-\frac{(i-L/2)^2}{(2\eta^2)}}}{\sum_{j=1}^{L} e^{-\frac{(j-L/2)^2}{(2\eta^2)}}}, \tag{2.3}$$

for $i = 1, \ldots, L$. Parameter η plays the role of variance in the Gaussian shape ($\eta \to \infty$ corresponds to white noise; $\eta \to 0$ corresponds to *one-band* noise). Parameter σ^2 controls the total noise power. We set $\sigma^2 = 8.1 \times 10^{-3}$ leading to SNR $= 17\,\text{dB}$, where

$$\text{SNR} \equiv 10 \log_{10} \frac{\mathbb{E}\big[\mathbf{x}^T\mathbf{x}\big]}{\mathbb{E}\big[\mathbf{n}^T\mathbf{n}\big]}. \tag{2.4}$$

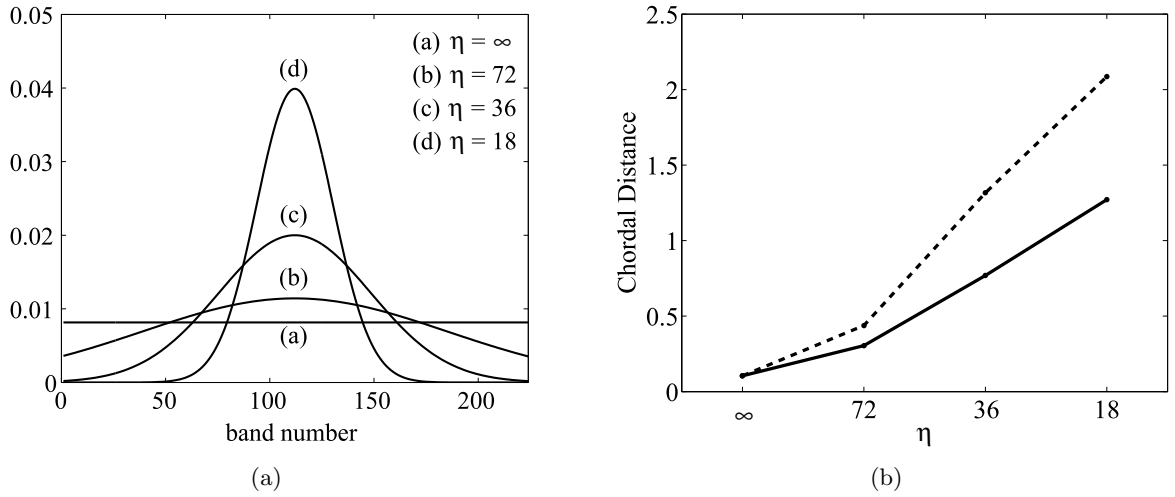
To measure the dissimilarity between the signal subspace and the subspace inferred by SVD, we adopt the *chordal distance* [60, 159], defined as

$$d = \frac{1}{\sqrt{2}} \|\mathbf{U}_p - \mathbf{U}_M\|_F, \tag{2.5}$$

where $\|\cdot\|_F$ denotes the Frobenius norm of a matrix, $\mathbf{U}_p = \mathbf{E}_p\mathbf{E}_p^T$ and $\mathbf{U}_M = \mathbf{E}_M\mathbf{E}_M^T$ are projection matrices onto the subspace of dimension p spanned by, respectively, the first p singular vectors of $\widehat{\mathbf{K}}_y$ and by the columns of $\mathbf{M}$. We note that the *chordal distance* is a measure of the projection error norm, *i.e.*, it is a measure of the errors $(\mathbf{U}_p - \mathbf{U}_M)\mathbf{x}$ for $\|\mathbf{x}\| = 1$ and $\mathbf{x} \in \mathbb{R}^L$. When this distance is zero the two projections are equal.

Fig. 2.1(a) shows the shape of the noise variance σ_i^2 for $\eta \in \{18, 38, 74, \infty\}$. Fig. 2.1(b) presents the chordal distance between the signal subspace and the subspace inferred by SVD, as a function of η and for $p = 5$. Notice the clear increasing of the chordal distance with η, *i.e.,* the chordal distance increases as the noise correlation shape becomes less flat. The degradation of the signal subspace estimate owing to the violation of the white noise assumption is quite clear.

In the example just presented, the subspace dimension was assumed known. However, this



(a) (b)

Figure 2.1: (a) Noise variance in each band for different values of η; (b) *Chordal distance* as a function of parameter η for SNR = 17 dB. (line) subspace inferred by SVD; (dashed line) subspace inferred by MNF.

dimension is unknown in most real applications and must be inferred from data as already referred to. This is a model-order inference problem that, if based only on the eigenvalues of the data correlation matrix, may lead to poor results. This aspect is illustrated in Fig. 2.2, where we have plotted the eigenvalues (circles) of $\widehat{\mathbf{K}}_r$ computed in the experiment above with SNR = 12 dB and $\eta = 18$. Note how difficult is to infer the number of signal eigenvalues for this SNR, because they are masked by the noise eigenvalues. We will see that HySime is able to infer the correct subspace dimension in the present scenario.

2.2.2 Maximum Noise Fraction

Maximum noise fraction (MNF) is a another popular subspace inference tool in remote sensing that takes into account the noise statistics. Nonetheless, it has limitations similar to SVD-based approaches, as we illustrate below.

MNF finds orthogonal directions minimizing the noise fraction (or, equivalently, maximizing the SNR). Assuming that the noise correlation matrix $\mathbf{K}_n$ or an estimate is known, this minimization

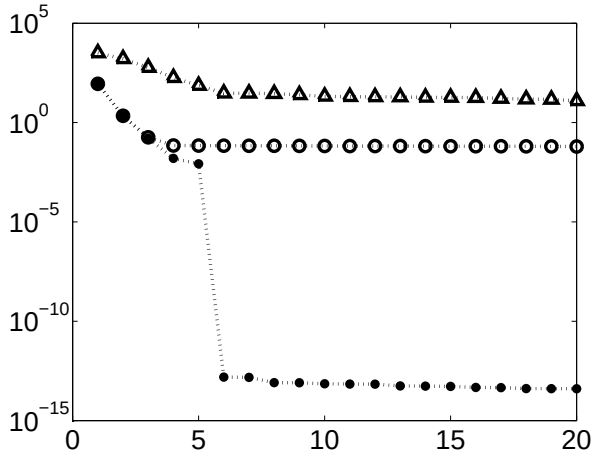


Figure 2.2: (dots) Eigenvalues of $\widehat{\mathbf{K}}_{\mathbf{x}}$; (circles) Eigenvalues of $\widehat{\mathbf{K}}_r$; (triangles) decreasing ratio of $(\mathbf{v}_i^T\mathbf{K}_r\mathbf{v}_i)/(\mathbf{v}_i^T\mathbf{K}_n\mathbf{v}_i)$ [$\mathbf{v}_i$, are the left-hand eigenvectors o $\mathbf{K}_n\mathbf{K}_r^{-1}$]. Simulation parameters: SNR $= 12\,\mathrm{dB}$, $\eta = 18$, and $p = 5$.

consists in finding orthogonal directions minimizing the ratio

$$\frac{\mathbf{v}_i^T\mathbf{K}_n\mathbf{v}_i}{\mathbf{v}_i^T\mathbf{K}_r\mathbf{v}_i}, \tag{2.6}$$

with respect to $\mathbf{v}_i$. This problem is known as the generalized Rayleigh quotient and the solution is given by the left-hand eigenvectors $\mathbf{v}_i$, for $i = 1, \ldots, L$, of $\mathbf{K}_n\mathbf{K}_y^{-1}$ [48].

For i.i.d noise, we have $\mathbf{K}_n = \sigma_n^{-2}\mathbf{I}$ and $\mathbf{K}_r^{-1} = \mathbf{E}\left(\mathbf{\Sigma} + \sigma_{\mathbf{n}}^{\mathbf{2}}\mathbf{I}\right)^{-\mathbf{1}}\mathbf{E}^{\mathbf{T}}$, and therefore MNF and SVD yield the same subspace estimate. However, if the noise is not i.i.d., the directions found by the MNF transform maximize the SNR but do not identify correctly the signal subspace. To illustrate this aspect, we apply the MNF transform to the data set generated in the the previous section.

The dashed line in Fig. 2.1(b) represents the chordal distance between the signal subspace and the subspace inferred by the MNF transform for different values of parameter η and assuming $p = 5$. The chordal distance exhibits a pattern similar to that of the SVD based approach being, however, larger for $\eta \neq \infty$ (white noise case).

Fig. 2.2 (triangles) plots the ratio $(\mathbf{v}_i^T\mathbf{K}_r\mathbf{v}_i)/(\mathbf{v}_i^T\mathbf{K}_n\mathbf{v}_i)$ along direction $\mathbf{v}_i$ by decreasing order.

As in the SVD-based approach, we face a model-order selection problem that is hard to solve without any further information or criterium.

2.3 Signal Subspace Estimation

This section introduce formally a new method to estimate the hyperspectral signal subspace termed *hyperspectral signal identification by minimum error* (HySime). The method starts by estimating the signal and the noise correlation matrices and then it selects the subset of eigenvectors that best represents the signal subspace in the minimum mean squared error sense. The application of this criterium leads to the minimization of a two-term objective function. One term corresponds to the power of the signal projection error and is a decreasing function of the subspace dimension; the other term corresponds to the power of the noise projection and is an increasing function of subspace dimension.

2.3.1 Noise Estimation

Noise estimation is a classical problem in data analysis and particulary in remote sensing. Arguably, in hyperspectral imagery, the simplest noise estimation procedure is the *shift difference* method, also denominated as *nearest neighbor difference* (NND) [61]. This approach assumes that noise samples taken from adjacent pixels are independent and have the same statistics, but the signal component is practically equal. To obtain meaningful noise estimates, the shift difference method shall be applied in homogeneous areas rather than on the entire image. This method has two weaknesses: first, it assumes that adjacent pixels have the same signal information, which is not valid in most hyperspectral data sets; second, to improve the noise estimation, a supervised selection of homogeneous areas must be carried out.

Herein, we follow a multiple regression theory [32, 148] based approach, which outperforms the shift difference method, as it exploits the high correlation existing among hyperspectral bands. Let

us denote by $\mathbf{R} = \left[\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N\right]$ an $L \times N$ matrix holding the N spectral observed vectors of size L. Define the matrix $\mathbf{Z} = \mathbf{R}^T$, the $N \times 1$ vector $\mathbf{z}_i = [\mathbf{Z}]_{:,i}$, where $[\mathbf{Z}]_{:,i}$ stands for the ith column of $\mathbf{Z}$ (*i.e.*, $\mathbf{z}_i$ contains the data read by the hyperspectral sensor at the ith band for all image pixels), and the $N \times (L-1)$ matrix $\mathbf{Z}_{\partial_i} = [\mathbf{z}_1, \ldots, \mathbf{z}_{i-1}, \mathbf{z}_{i+1}, \ldots, \mathbf{z}_L]$.

Assume that $\mathbf{z}_i$ is explained by a linear combination of the remaining $L-1$ bands. Formally, this consists in writing

$$\mathbf{z}_i = \mathbf{Z}_{\partial_i} \boldsymbol{\beta}_i + \boldsymbol{\xi}_i, \tag{2.7}$$

where $\mathbf{Z}_{\partial_i}$ is the explanatory data matrix, $\boldsymbol{\beta}_i$ is the regression vector of size $(L-1) \times 1$, and $\boldsymbol{\xi}_i$ is the modeling error vector of size $N \times 1$. For each $i \in \{1, \ldots, L\}$, the least squares estimator of the regression vector $\boldsymbol{\beta}_i$ is given by

$$\widehat{\boldsymbol{\beta}}_i = \left(\mathbf{Z}_{\partial_i}^T \mathbf{Z}_{\partial_i}\right)^{-1} \mathbf{Z}_{\partial_i}^T \mathbf{z}_i. \tag{2.8}$$

The noise is estimated by

$$\widehat{\boldsymbol{\xi}}_i = \mathbf{z}_i - \mathbf{Z}_{\partial_i} \widehat{\boldsymbol{\beta}}_i, \tag{2.9}$$

and the correlation matrix by $\widehat{\mathbf{K}}_n = \left[\widehat{\boldsymbol{\xi}}_1, \ldots, \widehat{\boldsymbol{\xi}}_N\right]^T \left[\widehat{\boldsymbol{\xi}}_1, \ldots, \widehat{\boldsymbol{\xi}}_N\right] / N$. Notice that the determination of each noise vector $\widehat{\boldsymbol{\xi}}_i$ implies the computation of the pseudo-inverse $\mathbf{Z}_{\partial_i}^{\#} = \left(\mathbf{Z}_{\partial_i}^T \mathbf{Z}_{\partial_i}\right)^{-1} \mathbf{Z}_{\partial_i}^T$, of size $(L-1) \times (L-1)$, for each $i = 1, \ldots, L$. This huge computational complexity can, however, be greatly reduced by taking advantage of the relation between $\mathbf{Z}_{\partial_i}^{\#}$ and $\mathbf{Z}$. Let the $L \times L$ symmetric and positive definite matrices $\mathbf{K}$ and $\mathbf{K}^{-1}$ be partitioned into block matrices as follows

$$\mathbf{K} = \left[\begin{array}{c|c} \mathbf{A} & \mathbf{b} \\ \hline \mathbf{b}^T & c \end{array}\right], \quad \mathbf{K}^{-1} = \left[\begin{array}{c|c} \mathbf{A}' & \mathbf{b}' \\ \hline \mathbf{b}'^T & c' \end{array}\right], \tag{2.10}$$

where $\mathbf{A}$ and $\mathbf{A}'$ are $(L-1 \times L-1)$ matrices, $\mathbf{b}$ and $\mathbf{b}'$ are $(L-1 \times 1)$ vectors, and c and c' are scalars. Since $\mathbf{K}$, $\mathbf{K}^{-1}$, $\mathbf{A}$, $\mathbf{A}^{-1}$, and c' are positive definite, thus

$$\mathbf{A}\mathbf{A}' + \mathbf{b}\mathbf{b}'^{T} = \mathbf{I}_{L-1} \tag{2.11}$$

$$\mathbf{A}\mathbf{b}' + \mathbf{b}c' = \mathbf{0}_{L-1}. \tag{2.12}$$

Replacing $\mathbf{A}^{-1}\mathbf{b}' = -\mathbf{b}'/c'$, derived from (2.12), into expression (2.11), we obtain

$$\mathbf{A}^{-1} = \mathbf{A}' - \mathbf{b}'\mathbf{b}'^{T}/\mathbf{c}'. \tag{2.13}$$

Based on this relation, the inversion of the matrix $\mathbf{Z}_{\partial_i}^T\mathbf{Z}_{\partial_i}$, for $i = 1, \ldots, L$, can be obtained by removing the ith row and the ith column of the matrix $\left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}$ and implementing expression (2.13) with the necessary adjustments.

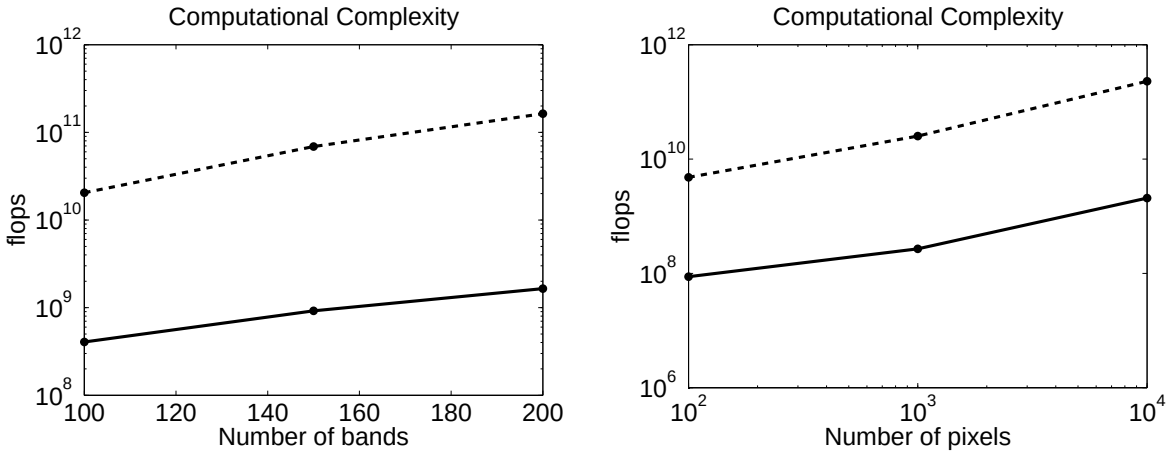
The pseudo-code for the noise estimation is shown in the Algorithm 2.1. Symbol $[\widehat{\mathbf{K}}]_{\partial_i,\partial_i}$ denotes the matrix obtained from $\widehat{\mathbf{K}}$ by deleting the ith row and the ith column, $[\widehat{\mathbf{K}}]_{i,\partial_i}$ denotes the ith row of $[\widehat{\mathbf{K}}]_{:,\partial_i}$, and $[\widehat{\mathbf{K}}]_{\partial_i,i}$ denotes $[\widehat{\mathbf{K}}]_{i,\partial_i}^T$. Steps 2 and 3 compute matrix $\widehat{\mathbf{K}} = \mathbf{Z}^T\mathbf{Z}$ and its inverse, respectively. Steps 5 and 6 estimate, respectively, the regression vector $\widehat{\boldsymbol{\beta}}_i$ and the noise $\widehat{\boldsymbol{\xi}}_i$, for each $i = 1, \ldots, L$.

Algorithm 2.1 : Noise estimation

1: INPUT $\mathbf{R} \equiv [\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N]$
2: $\mathbf{Z} = \mathbf{R}^T$, $\widehat{\mathbf{K}} := \left(\mathbf{Z}^T\mathbf{Z}\right)$;
3: $\mathbf{K}' := \widehat{\mathbf{K}}^{-1}$;
4: **for** $i := 1$ to L **do**
5: $\quad \widehat{\boldsymbol{\beta}}_i := \left([\mathbf{K}']_{\partial_i,\partial_i} - [\mathbf{K}']_{\partial_i,i}[\mathbf{K}']_{i,\partial_i}/[\mathbf{K}']_{i,i}\right)[\widehat{\mathbf{K}}]_{\partial_i,i}$;
$\quad\quad$ {Note that $\partial_i = 1, \ldots i-1, i+1, \ldots L$}
6: $\quad \widehat{\boldsymbol{\xi}}_i := \mathbf{z}_i - \mathbf{Z}_{\partial_i}\widehat{\boldsymbol{\beta}}_i$;
7: **end for**
8: OUTPUT $\widehat{\boldsymbol{\xi}}$; { $\widehat{\boldsymbol{\xi}}$ is a $N \times L$ matrix with the estimated noise}

Table 2.1: Computational complexity of the noise estimation algorithms.

Algorithm 2.1	$4NL^2 + 6L^3$
Algorithm without relation (2.13)	$4NL^2 + 2NL^3 + L^4$



(a) (b)

Figure 2.3: Illustration of the noise estimation algorithm complexity: implementation of Algorithm 2.1 (Solid line); direct implementation without using relation (2.13) (dashed line); (a) computational complexity as a function of the number of bands, for $N = 10^4$; (b) computational complexity as a function of the number of pixels, for $L = 224$.

The main advantage of Algorithm 2.1 is that the computation of $\widehat{\mathbf{K}}$ and of $\mathbf{K}' = \widehat{\mathbf{K}}^{-1}$ are out of the loop *for*. Thus, the computational complexity, *i.e.*, the number of floating point operations (flops), of Algorithm 2.1 is substantially lower than that of an algorithm implementing the multiple regression without using the relation (2.13). Note that the computation of the sample correlation matrix and of its inversion demands, approximately, $2NL^2 + L^3$ flops, whereas the multiple regression algorithm without using the relation (2.13) has to compute L times the above matrices, thus demanding, approximately, $2NL^3 + L^4$ flops. Table 2.1 presents approximated expressions for the number of floating point operations used by each algorithm. For $N \gg L$, Algorithm 2.1 demands, approximately, $L/2$ less flops, what is a significant figure since $L/2$ takes, in many applications, values on the order of 100. Fig. 2.3 presents the computational complexity of the Algorithm 2.1, as a function of the number of pixels and as a function of the number of bands. In the same figure

it is also presented the computational complexity of the noise estimation algorithm without using relation (2.13). Note that the Algorithm 2.1 has lowest complexity as mentioned above.

The next experiment illustrates Algorithm 2.1 working. The input data is a simulated hyperspectral image composed of 10^4 spectral vectors, each one following the linear mixing model (2.1). The abundance fractions are generated according to a Dirichlet distribution and the endmembers signatures are selected from the USGS digital spectral library. The noise is zero-mean independent with variances along the bands following a Gaussian shape [see Fig. 2.1(a)]. The number of endmembers is $p = 5$ and SNR $= 20\,\mathrm{dB}$.

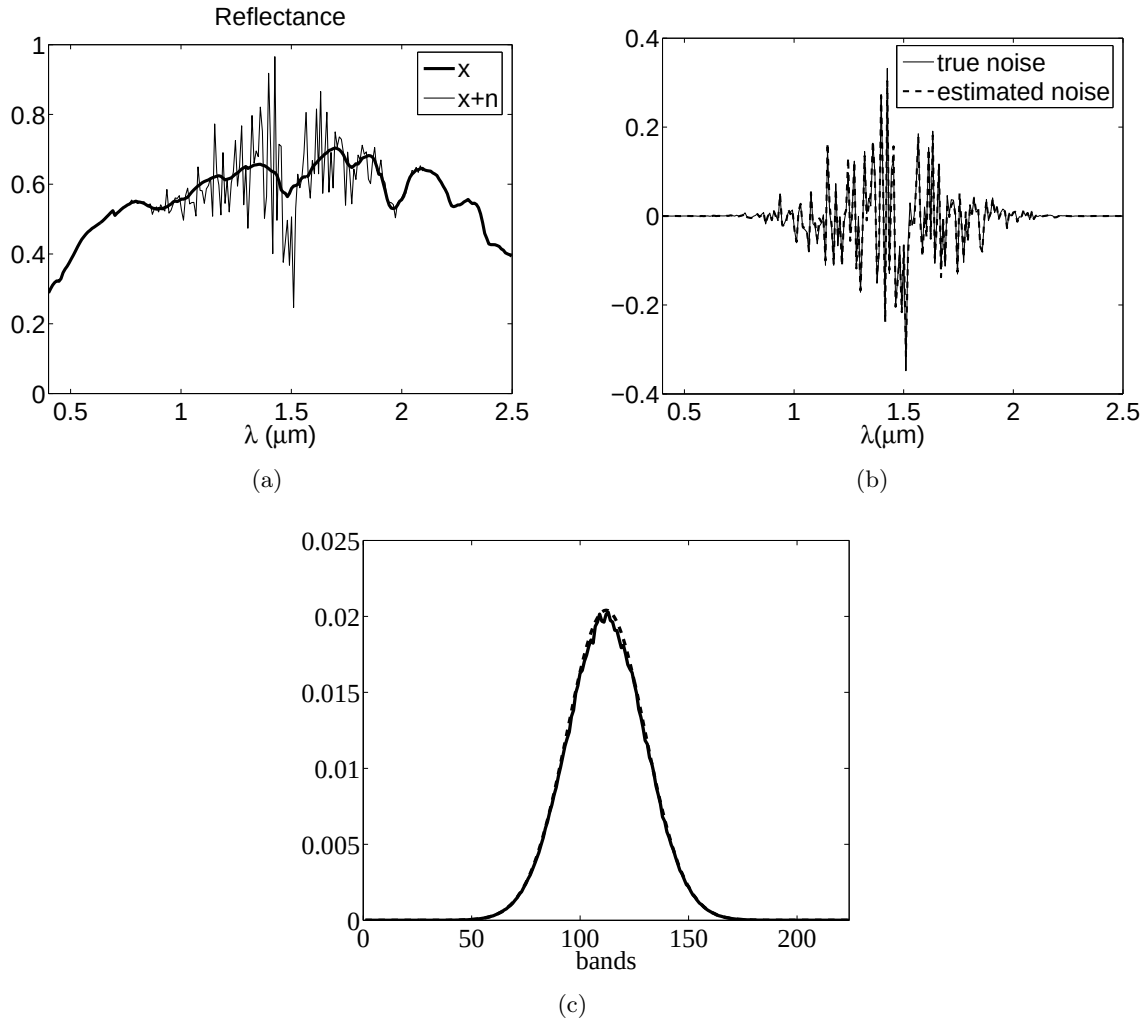
Fig. 2.4(a) shows the noiseless spectral vector $\mathbf{x}$ and the noisy version $\mathbf{x} + \mathbf{n}$. Fig. 2.4(b) shows the true and the estimated noise. The improvement in the SNR (*i.e.*, $E[\|\mathbf{n}\|^2]/E[\|\widehat{\mathbf{x}}-\mathbf{x}\|^2]$) is about 13 dB. Fig. 2.4(c) shows the diagonal of the covariance matrices of noise and its estimate. Note the accuracy of the estimate.

2.3.2 Signal Subspace Inference

This section presents the core structure of HySime. The first step, based on the noise estimation procedure introduced in the previous section, identifies a set of orthogonal directions of which an unknown subset spans the signal subspace. This subset is then determined by seeking the minimum mean squared error between $\mathbf{x}$, the original signal, and a noisy projection of it obtained from the vector $\mathbf{r} = \mathbf{x} + \mathbf{n}$. In the following, we assume that $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \widehat{\mathbf{K}}_n)$, *i.e.*, the noise is zero-mean Gaussian distributed with covariance matrix $\widehat{\mathbf{K}}_n$.

Let $\widehat{\mathbf{K}}_x = [\widehat{\mathbf{x}}_1, \ldots, \widehat{\mathbf{x}}_N]\,[\widehat{\mathbf{x}}_1, \ldots, \widehat{\mathbf{x}}_N]^T/N$ denote the signal sample correlation matrix. The eigendecomposition of $\widehat{\mathbf{K}}_x$ can be written as

$$\widehat{\mathbf{K}}_x = \mathbf{E}\boldsymbol{\Sigma}\mathbf{E}^T, \tag{2.14}$$



(a) (b)

(c)

Figure 2.4: Illustration of the noise estimation; (a) bold line: reflectance signal of a pixel; narrow line: the same pixel with noise; (b) solid line: true noise; dashed line: estimated noise. (c) solid line: diagonal of the estimated noise covariance matrix; dashed line: diagonal of the noise covariance matrix.

where $\mathbf{E} \equiv [\mathbf{e}_1, \ldots, \mathbf{e}_k, \mathbf{e}_{k+1} \ldots, \mathbf{e}_L]$ is a matrix with the eigenvectors, $\mathbf{e}_j$, ordered by the decreasing magnitude of the respective eigenvalues. The space $\mathbb{R}^L$ can be decomposed into two orthogonal subspaces: $\langle \mathbf{E}_k \rangle$ spanned by $\mathbf{E}_k \equiv [\mathbf{e}_1, \ldots, \mathbf{e}_k]$ and $\langle \mathbf{E}_k \rangle^{\perp}$ spanned by $\mathbf{E}_k^{\perp} \equiv [\mathbf{e}_{k+1}, \ldots, \mathbf{e}_L]$, *i.e.*, the orthogonal complement of subspace $\mathbf{E}_k$, where k is the order of the signal subspace.

Let $\mathbf{U}_k = \mathbf{E}_k \mathbf{E}_k^T$ be the projection matrix onto $\langle \mathbf{E}_k \rangle$ and $\widehat{\mathbf{x}}_k \equiv \mathbf{U}_k \mathbf{r}$ be the projection of the observed spectral vector $\mathbf{r}$ onto the subspace $\langle \mathbf{E}_k \rangle$. The first and the second-order moments of $\widehat{\mathbf{x}}_k$ given $\mathbf{x}$ are

$$\begin{aligned} \mathrm{E}\big[\widehat{\mathbf{x}}_k|\mathbf{x}\big] &= \mathbf{U}_k \mathrm{E}\big[\mathbf{r}|\mathbf{x}\big] \\ &= \mathbf{U}_k \mathrm{E}\big[\mathbf{x}+\mathbf{n}|\mathbf{x}\big] \\ &= \mathbf{U}_k \mathbf{x} \\ &\equiv \mathbf{x}_k, \end{aligned} \tag{2.15}$$

$$\begin{aligned} \mathrm{E}\big[(\widehat{\mathbf{x}}_k - \mathbf{x}_k)(\widehat{\mathbf{x}}_k - \mathbf{x}_k)^T|\mathbf{x}\big] &= \mathrm{E}\big[(\mathbf{U}_k\mathbf{r} - \mathbf{U}_k\mathbf{x})(\mathbf{U}_k\mathbf{r} - \mathbf{U}_k\mathbf{x})^T|\mathbf{x}\big] \\ &= \mathrm{E}\big[(\mathbf{U}_k \mathbf{n}\mathbf{n}^T \mathbf{U}_k^T)\big] \\ &= \mathbf{U}_k \widehat{\mathbf{K}}_n \mathbf{U}_k^T. \end{aligned} \tag{2.16}$$

The mean squared error between $\mathbf{x}$ and $\widehat{\mathbf{x}}_k$ is

$$\begin{aligned} \mathrm{mse}(k|\mathbf{x}) &= \mathrm{E}\big[(\mathbf{x} - \widehat{\mathbf{x}}_k)^T(\mathbf{x} - \widehat{\mathbf{x}}_k)|\mathbf{x}\big] \\ &= \mathrm{E}\big[(\underbrace{\mathbf{x} - \mathbf{x}_k}_{\mathbf{b}_k} - \mathbf{U_k n})^T (\underbrace{\mathbf{x} - \mathbf{x}_k}_{\mathbf{b}_k} - \mathbf{U_k n})|\mathbf{x}\big] \\ &= \mathbf{b}_k^T \mathbf{b}_k + \mathrm{tr}(\mathbf{U}_k \widehat{\mathbf{K}}_n \mathbf{U}_k^T). \end{aligned} \tag{2.17}$$

Computing the mean of (2.17) with respect to $\mathbf{x}$, noting that $\mathbf{b}_k = \mathbf{x} - \mathbf{x_k} = \mathbf{U_k^{\perp}} \mathbf{x}$, and using

the properties $\mathbf{U} = \mathbf{U}^T$, $\mathbf{U}^2 = \mathbf{U}$, and $\mathbf{U}^\perp = \mathbf{I} - \mathbf{U}$ of the projection matrices, we get

$$\begin{aligned} \text{mse}(k) &= \text{E}[(\mathbf{U}_k^\perp \mathbf{x})^T (\mathbf{U}_k^\perp \mathbf{x})] + \text{tr}(\mathbf{U}_k \widehat{\mathbf{K}}_n \mathbf{U}_k^T) \\ &= \text{tr}(\mathbf{U}_k^\perp \mathbf{K}_x) + \text{tr}(\mathbf{U}_k \widehat{\mathbf{K}}_n) \\ &= \text{tr}(\mathbf{U}_k^\perp \mathbf{K}_r) + 2\,\text{tr}(\mathbf{U}_k \widehat{\mathbf{K}}_n) + c, \end{aligned} \tag{2.18}$$

where c is an irrelevant constant. The criteria we propose to estimate the signal subspace is the minimization of mse(k) given by (2.18) with the correlation matrix $\mathbf{K}_r$ replaced with the sample correlation matrix $\widehat{\mathbf{K}}_r = \mathbf{R}\mathbf{R}^T/N$, *i.e.*,

$$\widehat{k} = \arg\min_k \left\{ \text{tr}(\mathbf{U}_k^\perp \widehat{\mathbf{K}}_r) + 2\,\text{tr}(\mathbf{U}_k \widehat{\mathbf{K}}_n) \right\}, \tag{2.19}$$

Each term of expression (2.19) has a clear meaning: the first accounts for the projection error power and is a decreasing function of k; the second accounts for the noise power and is an increasing function of k.

The pseudo-code for HySime is shown in the Algorithm 2.2. HySime inputs are the spectral observed vectors and the sample correlation matrix $\widehat{\mathbf{K}}_r$. Step 2 estimates the noise correlation matrix $\widehat{\mathbf{K}}_n$. Step 3 estimates the signal correlation matrix $\widehat{\mathbf{K}}_x$. Step 4 and 5 calculate the eigenvectors of the signal correlation matrix and the mean squared error function. The minimizer of this function is the the estimated signal subspace dimension, $\widehat{k}$.

Algorithm 2.2 : HySime

1: INPUT $\mathbf{R} \equiv [\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N]$, $\widehat{\mathbf{K}}_r \equiv \left(\mathbf{R}\mathbf{R}^T\right)/N$
2: $\widehat{\mathbf{K}}_n := \frac{1}{N}\sum_i \left(\widehat{\mathbf{n}}_i \widehat{\mathbf{n}}_i^T\right)$; $\{\widehat{\mathbf{K}}_n$ is the noise correlation matrix estimates$\}$
3: $\widehat{\mathbf{K}}_x := \frac{1}{N}\sum_i \left((\mathbf{r}_i - \widehat{\mathbf{n}}_i)(\mathbf{r}_i - \widehat{\mathbf{n}}_i^T)\right)$; $\{\widehat{\mathbf{K}}_x$ is the signal correlation matrix estimates$\}$
4: $\mathbf{U}_k := \mathbf{E}_k \mathbf{E}_k^T$; {where $\mathbf{E}_k$ are eigenvectors of $\widehat{\mathbf{K}}_x$}
5: $\widehat{k} := \arg\min_k \left\{ \text{tr}(\mathbf{U}_k^\perp \widehat{\mathbf{K}}_r) + 2\,\text{tr}(\mathbf{U}_k \widehat{\mathbf{K}}_n) \right\}$

Signal Subspace Order Estimation Using the Mean Value of Spectral Vectors (HySime$_m$)

An alternative to the projection $\mathbf{U}_k^{\perp}\widehat{\mathbf{K}}_r$, consists in projecting the sample mean $\overline{\mathbf{r}} \equiv \frac{1}{N}\sum_{i=1}^{N}\mathbf{r}_i$ onto the subspace orthogonal to the signal subspace. The underlying rational for this approach is that spectral vectors are nonnegative and then $\overline{\mathbf{r}}$ accumulates information about every spectral vectors in the data set. Of course, for this approach to work, the projection of $\overline{\mathbf{r}}$ onto any eigenvector $\mathbf{e}_j$, $j = 1, \ldots, k$ must be nonzero. Although we have not any proof of this statement, we believe, supported on practical evidence, that the probability of $\overline{\mathbf{r}}^T\mathbf{e}_j = 0$ is practically zero in real data sets.

The sample mean vector of the data set is

$$\begin{aligned}\overline{\mathbf{r}} &= \frac{1}{N}\sum_{i=1}^{N}\mathbf{r}_i \\ &= \frac{1}{N}\mathbf{M}\sum_{i=1}^{N}\mathbf{s}_i + \frac{1}{N}\sum_{i=1}^{N}\mathbf{n}_i \\ &= \mathbf{c} + \boldsymbol{\omega}, \end{aligned} \tag{2.20}$$

where $\mathbf{c}$ is in the signal subspace and $\boldsymbol{\omega} \sim \mathcal{N}(0, \mathbf{K}_n/N)$. Note that the noise correlation is smaller by a factor of N than that of the first method. Let $\mathbf{c}_k$ be the projection of $\mathbf{c}$ onto $\langle \mathbf{E}_k \rangle$. The estimation of $\mathbf{c}_k$ can be obtained by projecting $\overline{\mathbf{r}}$ onto the signal subspace $\langle \mathbf{E}_k \rangle$, *i.e.*, $\widehat{\mathbf{c}}_k = \mathbf{U}_k\overline{\mathbf{r}}$.

The first and the second-order moments of the estimated error $\mathbf{c} - \widehat{\mathbf{c}}_k$ are respectively,

$$\begin{aligned} \mathrm{E}\big[\mathbf{c} - \widehat{\mathbf{c}}_k\big] &= \mathbf{c} - \mathrm{E}\big[\widehat{\mathbf{c}}_k\big] \\ &= \mathbf{c} - \mathbf{U}_k\mathbf{c} \\ &= \mathbf{c} - \mathbf{c}_k \\ &\equiv \mathbf{b}_k, \qquad (2.21) \\ \mathrm{E}\big[(\mathbf{c} - \widehat{\mathbf{c}}_k)(\mathbf{c} - \widehat{\mathbf{c}}_k)^T\big] &= \mathbf{b}_k\mathbf{b}_k^T + \mathbf{U}_k\widehat{\mathbf{K}}_n\mathbf{U}_k^T/N, \qquad (2.22) \end{aligned}$$

where the bias $\mathbf{b}_k = \mathbf{U}_k^{\perp}\mathbf{c}$ is the projection of $\mathbf{c}$ onto the space $\langle\mathbf{E}_k\rangle^{\perp}$. Therefore, the density of the estimated error $\mathbf{c} - \widehat{\mathbf{c}}_k$ is $\mathcal{N}(\mathbf{b}_k, \mathbf{b}_k^T\mathbf{b}_k + \mathbf{U}_k\widehat{\mathbf{K}}_n\mathbf{U}_k^T/N)$,

The mean squared error between $\mathbf{c}$ and $\widehat{\mathbf{c}}_k$ is given by

$$\begin{aligned} \mathrm{mse}(k) &= \mathrm{E}\big[(\mathbf{c} - \widehat{\mathbf{c}}_k)^T(\mathbf{c} - \widehat{\mathbf{c}}_k)\big] \\ &= \mathrm{tr}\{\mathrm{E}\big[(\mathbf{c} - \widehat{\mathbf{c}}_k)(\mathbf{c} - \widehat{\mathbf{c}}_k)^T\big]\} \\ &= \mathbf{b}_k^T\mathbf{b}_k + \mathrm{tr}(\mathbf{U}_k\widehat{\mathbf{K}}_n\mathbf{U}_k^T/N), \qquad (2.23) \end{aligned}$$

The bias $\mathbf{b}_k$ is not known and an approximation of expression (2.23) is obtained by using the bias estimate $\widehat{\mathbf{b}}_k = \mathbf{U}_k^{\perp}\overline{\mathbf{r}}$. However, $\mathrm{E}\big[\widehat{\mathbf{b}}_k\big] = \mathbf{b}_k$ and $\mathrm{E}\big[\widehat{\mathbf{b}}_k^T\widehat{\mathbf{b}}_k\big] = \mathbf{b}_k^T\mathbf{b}_k + \mathrm{tr}(\mathbf{U}_k^{\perp}\widehat{\mathbf{K}}_n{\mathbf{U}_k^{\perp}}^T/N)$, that is, an unbiased estimate of $\mathbf{b}_k^T\mathbf{b}_k$ is $\widehat{\mathbf{b}}_k^T\widehat{\mathbf{b}}_k - \mathrm{tr}(\mathbf{U}_k^{\perp}\widehat{\mathbf{K}}_n{\mathbf{U}_k^{\perp}}^T/N)$. The criteria for the signal subspace order determination is then

$$\begin{aligned} \widehat{k} &= \arg\min_k\Big\{\widehat{\mathbf{b}}_k^T\widehat{\mathbf{b}}_k + \mathrm{tr}(\mathbf{U}_k\widehat{\mathbf{K}}_n\mathbf{U}_k^T/N) - \mathrm{tr}(\mathbf{U}_k^{\perp}\widehat{\mathbf{K}}_n{\mathbf{U}_k^{\perp}}^T/N)\Big\} \\ &= \arg\min_k\Big\{\overline{\mathbf{r}}^T\mathbf{U}_k^{\perp}\overline{\mathbf{r}} + 2\mathrm{tr}(\mathbf{U}_k\widehat{\mathbf{K}}_n/N)\Big\}. \qquad (2.24) \end{aligned}$$

The pseudo-code for the alternative method, termed HySime_m (m stands for mean), is shown

in the Algorithm 2.3. The core of this algorithm is similar to the Algorithm 2.2, excepts for Step 5, which calculates the sample mean of the data set $\overline{\mathbf{r}}$, and step 6 which refers to expression (2.24).

Algorithm 2.3 :HySime$_m$

1: INPUT $\mathbf{R} \equiv [\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N]$
2: $\widehat{\mathbf{K}}_n := \frac{1}{N}\sum_i \left(\widehat{\mathbf{n}}_i \widehat{\mathbf{n}}_i^T\right)$; {$\widehat{\mathbf{K}}_n$ is the noise correlation matrix estimates}
3: $\widehat{\mathbf{K}}_x := \frac{1}{N}\sum_i \left((\mathbf{r}_i - \widehat{\mathbf{n}}_i)(\mathbf{r}_i - \widehat{\mathbf{n}}_i^T)\right)$; {$\widehat{\mathbf{K}}_x$ is the signal correlation matrix estimates}
4: $\mathbf{U}_k := \mathbf{E}_k \mathbf{E}_k^T$; {where $\mathbf{E}_k$ are eigenvectors of $\widehat{\mathbf{K}}_x$}
5: $\overline{\mathbf{r}} := \frac{1}{N}\sum_{i=1}^{N} \mathbf{r}_i$; {$\overline{\mathbf{r}}$ is the sample mean of the data set}
6: $\widehat{k} := \arg\min_k \left\{ \overline{\mathbf{r}}^T \mathbf{U}_k^{\perp} \overline{\mathbf{r}} + 2\mathrm{tr}(\mathbf{U}_k \widehat{\mathbf{K}}_n / N) \right\}$;

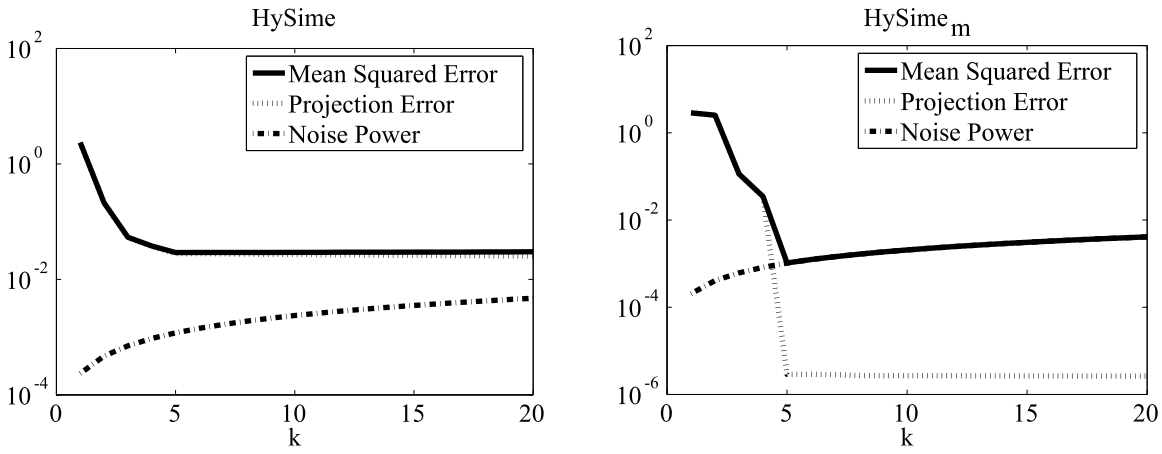
2.4 Evaluation of HySime and HySime$_m$ with Simulated Data

In this section, the proposed methods (HySime and HySime$_m$) are applied to simulated scenes and compared with NWHFC eigen-based Neyman-Pearson detector [32]. As concluded in [32], these algorithms are the state-of-the-art in hyperspectral signal subspace identification, outperforming the information theoretical criteria approaches; namely, the MDL [144, 150] and the AIC [4].

The spectral signatures are selected from the USGS digital spectral library [40]. The abundance fractions are generated according to a Dirichlet distribution defined in expression (2.2). The results presented here are organized into two experiments: in the first experiment, the method is evaluated with respect to the SNR [see def. (2.4)] to the number of endmembers p, and to the spectral noise shape (white and nonwhite). In the second experiment, the methods are evaluated with respect to their ability to detect rare pixels.

Experiment I

Fig. 2.5 shows the evolution of the mean squared error for HySime and HySime$_m$ algorithms as a function of the parameter k, for SNR $= 35\,$dB and $p = 5$. The minimum of the mean squared error occurs at $k = 5$, which is exactly the number of endmembers present in the image. As expected, the



(a) (b)

Figure 2.5: Mean squared error versus k, with SNR = 35 dB, $p = 5$; (a) HySime method; (b) HySime$_m$ method.

projection error power and of noise power display decreasing and increasing behaviors, respectively, as a function of the subspace dimension k.

Table 2.2 presents the signal subspace order estimates yielded by HySime, HySime$_m$ algorithms, and the VD determined by the NWHFC algorithm [32], as a function of the SNR, of the number of endmembers, p, and of the noise shape.

NWHFC algorithm is basically the HFC one [67] preceded by a noise-whitening step, based on the estimated noise correlation matrix. In implementing this step, we got poor results in very high SNRs and colored noise scenarios. For this reason, we have used both the true and estimated noise correlation matrices. The results based on the true correlation matrix are in brackets. We stress that, for the setting of this experiment, both HySime and HySime$_m$ methods yield the same results, whether using the estimated or the true noise correlation matrices.

Another central issue of NWHFC algorithm is the false-alarm probability P_f it is parameterized with. This probability is used in a series of Neyman-Pearson tests, each one designed to detect a different orthogonal signal subspace direction. There is the need, therefore, to specify the false-alarm probability P_f of the tests. Based on the hints given in [32] and in our own results, we choose

$P_f \in \{10^{-3}, 10^{-4}, 10^{-5}\}$.

The figures shown in Table 2.2, based on 50 Monte Carlo runs, have the following behavior:

i) HySime and HySime$_m$ algorithms display similar performance, with a small advantage for former, namely at small SNRs and colored noise;

ii) HySime and NWHFC algorithms parameterized with $P_f = 10^{-3}$ display similar performances at low subspace dimension, say $p \leq 5$, and white noise. This is also true for colored noise and NWHFC working with known noise covariance matrix. However, if the noise statistics is unknown, NWHFC performs much worse than HySime;

iii) HySime performs better that NWHFC for high space dimensions, say $p > 5$.

We conclude, therefore, that HySime algorithm is slight better than HySime$_m$ one , yielding systematically equal or better results than NWHFC algorithm. Another advantage of HySime approach is its adaptiveness, *i.e.*, it does not depend on any tunable parameter.

Experiment II

In this experiment, we set SNR $= 35\,\text{dB}$ and $p = 8$. The first five endmembers are mixed according to a Dirichlet distribution, as in the previous experiment, the sixth, the seventh, and the eighth endmembers are present as pure in 8, 4, and 2 pixels, respectively. Fig. 2.6 shows the mean squared error versus the subspace dimension k for the HySime and HySime$_m$ methods. The respective curves achieve their minima at $k = 8$ and $k = 7$. Thus, the HySime algorithm infers the correct subspace dimension, whereas HySime$_m$ one underestimates it in one unit.

Table 2.3 displays the results of this experiment computed by HySime, HySime$_m$, and NWHFC algorithms. We observe the same pattern of behavior shown in Table 2.2, with HySime method yielding the best performance.

Table 2.2: Signal subspace dimension $\widehat{k}$, based on 50 Monte Carlo runs, as a function of SNR, p, and η (noise shape). Figures in brackets were computed based on the true noise statistics.

	Noise	White ($\eta = 0$)				Gaussian shaped ($\eta = 1/18$)			
SNR	Method	$p = 3$	$p = 5$	$p = 10$	$p = 15$	$p = 3$	$p = 5$	$p = 10$	$p = 15$
50 dB	HySime	**3**	**5**	**10**	**15**	**3**	**5**	**10**	**15**
	HySime_m	3	5	10	15	3	5	10	15
	NWHFC ($P_f = 10^{-3}$)	3 (3)	5 (5)	7 (7)	10 (11)	59 (3)	41 (5)	61 (10)	45 (10)
	NWHFC ($P_f = 10^{-4}$)	3 (3)	5 (5)	7 (7)	8 (8)	48 (3)	33 (5)	54 (10)	34 (10)
	NWHFC ($P_f = 10^{-5}$)	3 (3)	4 (4)	7 (6)	8 (8)	43 (3)	28 (5)	41 (9)	27 (10)
35 dB	HySime	**3**	**5**	**10**	**15**	**3**	**5**	**10**	**15**
	HySime_m	3	5	10	15	3	5	10	15
	NWHFC ($P_f = 10^{-3}$)	3 (3)	4 (4)	7 (7)	9 (9)	9 (3)	10 (5)	12 (10)	10 (10)
	NWHFC ($P_f = 10^{-4}$)	3 (3)	4 (4)	7 (6)	8 (8)	9 (3)	9 (5)	11 (10)	8 (10)
	NWHFC ($P_f = 10^{-5}$)	3 (3)	4 (4)	6 (6)	8 (8)	7 (3)	7 (5)	10 (9)	8 (10)
25 dB	HySime	**3**	**5**	**10**	**14**	**3**	**5**	**10**	**15**
	HySime_m	3	5	9	12	3	5	10	12
	NWHFC ($P_f = 10^{-3}$)	3 (3)	5 (5)	6 (6)	9 (8)	4 (3)	5 (5)	11 (10)	9 (11)
	NWHFC ($P_f = 10^{-4}$)	3 (3)	5 (5)	6 (6)	7 (7)	4 (3)	5 (5)	11 (10)	9 (10)
	NWHFC ($P_f = 10^{-5}$)	3 (3)	4 (4)	5 (5)	7 (7)	4 (3)	5 (5)	11 (9)	8 (10)
15 dB	HySime	**3**	**5**	**8**	**12**	**3**	**5**	**8**	**12**
	HySime_m	3	3	6	8	3	3	5	8
	NWHFC ($P_f = 10^{-3}$)	3 (3)	5 (5)	5 (4)	5 (5)	4 (3)	5 (5)	11 (10)	10 (10)
	NWHFC ($P_f = 10^{-4}$)	3 (3)	4 (4)	3 (3)	3 (2)	4 (3)	5 (5)	11 (10)	8 (10)
	NWHFC ($P_f = 10^{-5}$)	3 (3)	4 (4)	3 (3)	2 (2)	4 (3)	5 (5)	11 (9)	8 (10)

Table 2.3: Signal subspace dimension $\widehat{k}$ as a function parameter η experiment 2 (few pure pixels) with SNR = 35 dB.

Method ($p = 8$)	$\eta = 0$	$\eta = 1/18$
HySime	8	7
HySime_m	7	7
NWHFC ($P_f = 10^{-3}$)	6	13
NWHFC ($P_f = 10^{-4}$)	6	11
NWHFC ($P_f = 10^{-5}$)	6	11

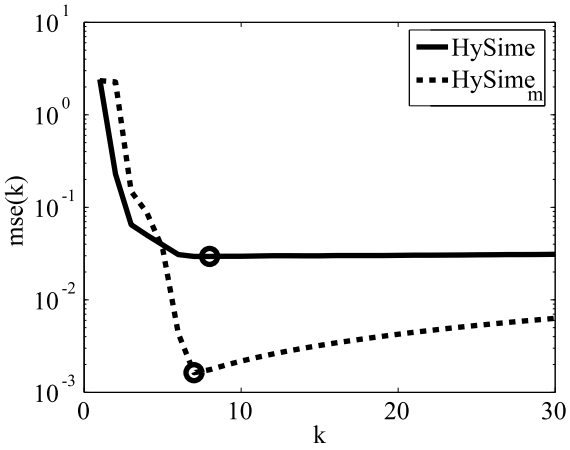
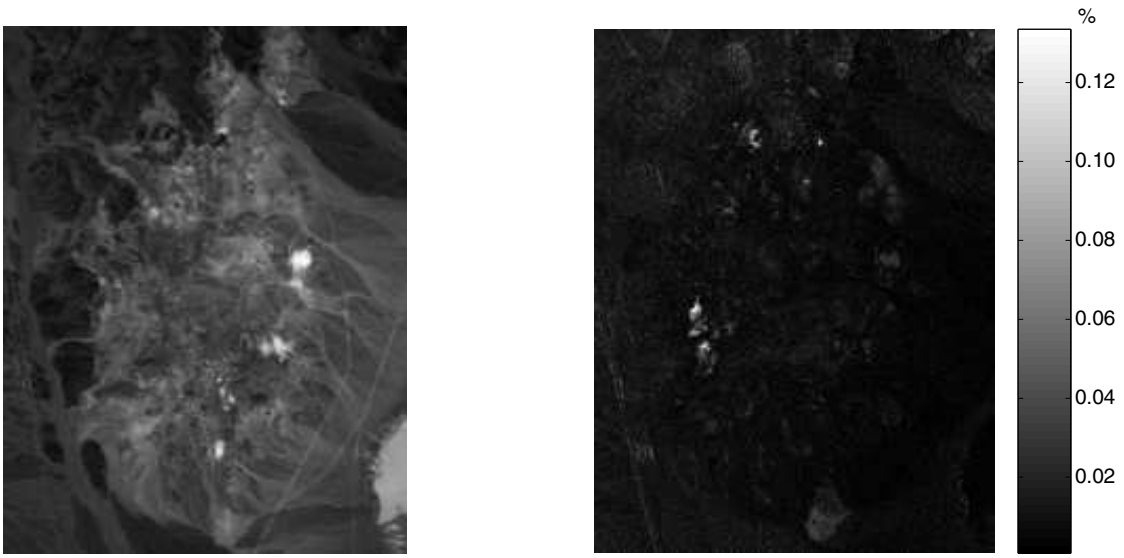


Figure 2.6: Mean squared error versus k, with SNR $= 35\,\text{dB}$, $p = 8$ (second experiment). HySime method(solid line); HySime$_m$ method (dashed line); estimated number of endmembers (circles).

2.5 Experiments with Real Hyperspectral Data

In this section, the proposed methods (HySime and HySime$_m$) are applied to real hyperspectral data collected by the AVIRIS [168] sensor over Cuprite, Nevada[1]. Cuprite is a mining area in southern Nevada with mineral and little vegetation [162]. The Cuprite test site, located approximately 200 Km northwest of Las Vegas is a relatively undisturbed acid-sulfate hidrothermal system near highway 95. The geology and alteration were previously mapped in detail [1, 5]. A geologic summary and a mineral map can be found in [162]. This site has been extensively used for remote sensing experiments over the past years [59, 93] and it has become a standard test site for comparison of unmixing and endmember extraction algorithms. This study is based on a subimage (250×190 pixels and 224 bands) of a data set acquired on the AVIRIS flight of June 19, 1997 [see Fig. 2.7(a)]. AVIRIS instrument covers the spectral region from $0.41\mu m$ to $2.45\mu m$ in 224 bands with a $10nm$ bandwidth. Flying at an altitude of $20km$, it has an IFOV of $20m$ and views a swath over $10km$ wide.

[1] Available at http://aviris.jpl.nasa.gov/html/aviris.freedata.html

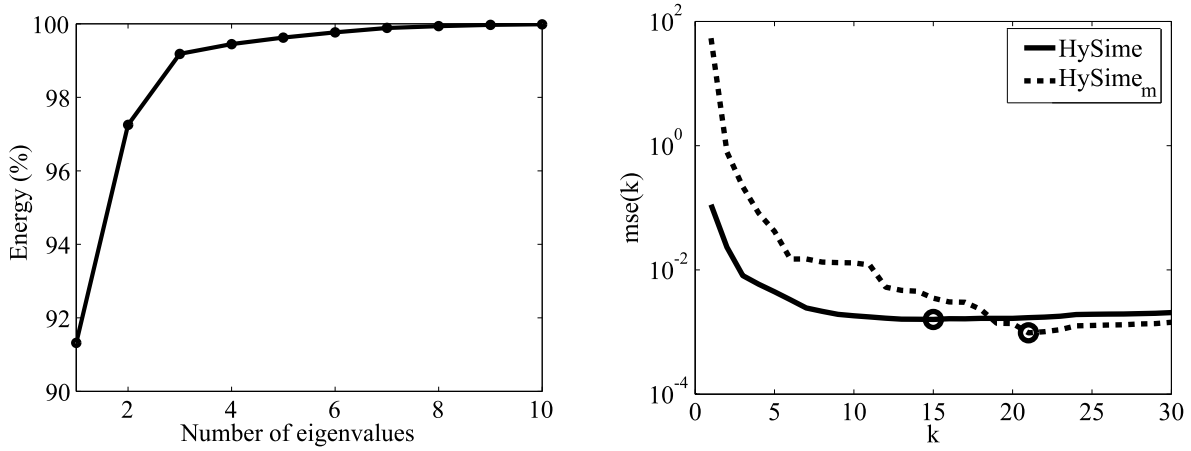


(a) (b)

Figure 2.7: (a) Band 30 (wavelength $\lambda = 667.3nm$) of the subimage of AVIRIS cuprite Nevada data set; (b) percentage of energy in the subspace $\langle \mathbf{E}_{9:15} \rangle$.

The HySime method when applied to this data set estimates $\widehat{k} = 15$, the alternative method HySime_m estimates $\widehat{k} = 21$ [see Fig. 2.8(b)]. According to the ground truth presented in [162], there are 8 materials in this area. These differences are due to i) the presence of rare pixels not accounted for in [162] and ii) spectral variability. The bulk of spectral energy is explained by a small number of eigenvectors. This can be observed from Fig. 2.8(a), where the accumulated signal energy is plotted as a function of the eigenvalue index ordered by decreasing magnitude. The energy contained in the first 8 eigenvalues is 99.95% of the total signal energy. This fact is further confirmed in Fig. 2.7(b), which shows, in gray level and for each pixel, the percentage of energy contained in the subspace $\langle \mathbf{E}_{9:15} \rangle = \langle [\mathbf{e}_9, \ldots, \mathbf{e}_{15}] \rangle$. Note that only a few (rare) pixels contain energy in this subspace. Furthermore, these energies are a very small percentage of the corresponding spectral vector energies (less than 0.14%).

The VD estimated by the NWHFC method [32] ($P_f = 10^{-3}$) on the same data set yields $\widehat{k} = 20$. A lower value of P_f would lead to a lower number of endmembers. According to the ground truth presented in [162], the estimates yielded by HySime and NWHFC methods overestimate the number



(a) (b)

Figure 2.8: (a) Percentage of signal energy as a function of the number of eigenvalues; (b) mean squared error versus k for Cuprite data set: HySime method (solid line); HySime$_m$ method (dashed line); estimated number of endmembers (circles).

of endmembers in the Cuprite data set. The mainly reason for this difference, as we have explained, is the presence of rare pixels present in the data set not accounted for in [162]. A more thorough evaluation of HySime, and NWHFC, would demand further field experiments.

2.6 Conclusions

The huge volumes and rates of data generated by hyperspectral sensors demand expensive processors with very high performance and memory capacities. Dimensionality reduction is, therefore, a relevant first step in the hyperspectral data processing chain. In this chapter, a method to estimate the dimensionality of hyperspectral linear mixtures is proposed. The HySime method estimates the signal and the noise correlation matrices and then selects the subset of singular values that best represents the signal subspace based on the minimum mean squared error criterium. An alternative method, HySime$_m$, exploits the fact that hyperspectral mixtures are nonnegative. This method also estimates the signal and noise correlation matrices and then the signal subspace is selected based on the best representation of the mean value of the hyperspectral image in a least square

sense.

A comparison of the proposed method with NWHFC [32] is conducted. A set of experiments with simulated and real data leads to the conclusion that the HySime algorithm is an effective and useful tool, yielding comparable or better results than the state-of-the-art algorithms.

Chapter 3

Vertex Component Analysis: A Fast Algorithm to Unmix Hyperspectral Data

3.1 Introduction

Hyperspectral vectors are mixtures of the spectral signatures of the endmembers present in the scene. Linear spectral mixture analysis, or linear unmixing, aims at estimating the number of endmembers, their spectral signatures, and their abundance fractions. Usually, this task embodies two steps: endmember extraction to determine the spectral signatures of endmembers followed by inversion to estimate the abundance fractions of each endmember. This chapter presents a new fully automatic method for unsupervised endmember extraction from hyperspectral data, termed *vertex component analysis* (VCA). The algorithm exploits two facts: i) the endmembers are the vertices of a simplex and ii) the affine transformation of a simplex is also a simplex. Briefly, VCA algorithm iteratively projects the spectral vectors onto a direction orthogonal to the subspace spanned by the endmembers already determined. The new endmember signature corresponds to the extreme of the projection, thus the name *vertex component analysis*. The algorithm iterates until all endmembers are exhausted. VCA is a fully automatic algorithm and works with or without dimensionality reduction. It is shown that VCA competes with state-of-the-art methods, while having a computational complexity between one and two orders of magnitude lower than the best

available method.

Under the linear mixing model, the observations from a scene are in a simplex whose vertices correspond to the endmembers [19, 20]. Endmember abundance fractions are determined by the position of the mixed pixel in the simplex. Several approaches have been developed to extract endmembers from data exploiting this geometric feature of hyperspectral mixtures. A comparative study of these algorithms is conducted in [139, 174].

For example, the minimum volume transform (MVT) algorithm [44] determines the simplex of minimum volume containing the data. Convex cone analysis (CCA) [76] is based on the fact that spectral signatures are nonnegative. Thus, CCA finds the boundary points of the convex region. This approaches are complex from the computational point of view.

The method introduced in [10] uses the notion of bundles, *i.e.*, each substance present in the ground surface is represented by a set or bundle of spectra to account for the spectral variability. A simulated annealing algorithm is used to derive the bundles of endmembers. The complexity of this method is even higher than that of MVT type approaches, since the temperature of the simulated annealing algorithm used therein follows a $\log(\cdot)$ law [55] to ensure convergence (in probability) to the desired solution.

Multiple endmember spectral mixture analysis (MESMA) [145] considers many possible mixture models simultaneously in order to produce the lowest margin of root mean square (RMS) error when compared to the apparent surface reflectance of the pixel. The major limitation of this approach is the need of extensive libraries.

Automatic morphological endmember extraction (AMEE) [138] is an algorithm that makes simultaneous use of the spatial and spectral information. The algorithm does not need previous dimensionality reduction and uses the morphological operations erosion and dilatation with different kernel sizes to select the endmembers.

The iterative error analysis (IEA) algorithm [130, 158] does not use dimensionality reduction and is based on sequential constrained unmixing procedures to find the endmembers given by the average score of the vectors with higher degrees of error.

The manual endmember selection tool (MEST) [8, 9] projects the spectral vectors onto a p-dimensional space with PCA, providing the user with a means of exploration in the mixing space to search for $p + 1$ spectra that are acceptable as the spectral signatures of ground components. The supervised nature of the approach is its major limitation.

The exemplar selector module of ORASIS [25, 27] is used to select spectral vectors that best represent the smaller convex cone containing the data. This vectors are then projected onto a lower subspace and a simplex is found by an MVT process.

Aiming at a lower computational complexity, some algorithms such as the pixel purity index (PPI) [19] and the N-FINDR [173] still search for the minimum volume simplex containing the data cloud, but they assume the presence in the data of at least one pure pixel of each endmember. This is a strong requisite that may not hold in some data sets. Note that the presence of pure pixels in the data depends on the sensor spatial resolution and their existence cannot be detect visually. In any case, these algorithms find the set of *most pure pixels* in the data.

PPI algorithm uses the MNF [21] as a preprocessing step to reduce dimensionality and to improve the SNR. The algorithm then projects every spectral vector onto *skewers* (large number of random vectors) [19, 98, 164]. The points corresponding to extremes, for each *skewer* direction, are stored. A cumulative account records the number of times each pixel (*i.e.*, a given spectral vector) is found to be an extreme. The pixels with the highest scores are the purest ones. PPI was, firstly, conceived as supervised tool to identify endmembers. Therefore, the most purest pixels are loaded into a multi-dimensional visualization tool and then endmembers are identified visually as the extreme pixels in the data cloud. Different methods to implement PPI aimed at reducing algorithm complexity are proposed in [98, 164].

N-FINDR [173] is a fully automated algorithm to identify endmembers in a hyperspectral data set. It is based on the fact that in p spectral dimensions, the p-volume defined by a simplex formed by the purest pixels is larger than any other volume defined by any other combination of pixels. The N-FINDR algorithm first reduces the dimensionality of the original data using MNF transform. Then a random choice of a set of pixels is done and the initial volume is calculated. Next, for each pixel of the data set, a trial volume is calculated by replacing each endmember in the set by the current one. If the replacement results in a volume increasing, the replacement is made effective.

N-FINDR and PPI are perhaps the most well known endmember extraction algorithms . These algorithms are also implemented in commercial software packages, which have spread their use in hyperspectral applications. In this chapter VCA is developed and compared with N-FINDR and PPI. We conclude that for comparable performance VCA is faster between one and two orders of magnitude.

The chapter is structured as follows. Section 3.2 describes the fundamentals of the proposed method and Sections 3.3 and 3.4 evaluate the proposed algorithm using simulated and real data, respectively. Section 3.5 ends the chapter by presenting some concluding remarks. Parts of the VCA introduced here were published in [18, 120, 121, 123, 127].

3.2 Vertex Component Analysis Algorithm

Assuming the linear mixing scenario, introduced in Section 1.2.1, each pixel is an $L \times 1$ vector (L is the number of bands) given by

$$\begin{aligned} \mathbf{r} &= \mathbf{x} + \mathbf{n} \\ &= \mathbf{M} \underbrace{\gamma \boldsymbol{\alpha}}_{\mathbf{s}} + \mathbf{n}, \end{aligned} \tag{3.1}$$

where $\mathbf{M} \equiv [\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_p]$ is the mixing matrix ($\mathbf{m}_j$ denotes the jth endmember signature and p is the number of endmembers present in the covered area), $\mathbf{s} \equiv \gamma\boldsymbol{\alpha}$ (γ is a scale factor modeling the illumination variability due to surface topography), $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_p]^T$ is the abundance vector containing the fractions of each endmember and $\mathbf{n}$ models the additive noise.

As mentioned before, owing to physical constraints, $\boldsymbol{\alpha} \in \Delta_p$, *i.e.*, abundance fractions satisfy the full additivity and the positivity constraints [see expression (1.4)]. Each pixel can be viewed as a vector in an L-dimensional Euclidean space, where each channel is assigned to one axis of space. Since the set Δ_p is a simplex, the $S_x = \{\mathbf{x} \in \mathbb{R}^L : \mathbf{x} = \mathbf{M}\boldsymbol{\alpha}, \boldsymbol{\alpha} \in \Delta_p\}$ is also a simplex. However, even assuming that $\mathbf{n} = \mathbf{0}$, the observed vector set belongs to $C_p = \{\mathbf{r} \in \mathbb{R}^L : \mathbf{r} = \mathbf{M}\gamma\boldsymbol{\alpha}, \boldsymbol{\alpha} \in \Delta_p, \gamma \geq 0\}$ which is a convex cone, owing to scale factor γ. For illustration purposes, a simulated scene was generated according to the expression (3.1). Fig. 3.1(a) illustrates a simplex and a cone, projected on a 2D-subspace, defined by a mixture of three endmembers. These spectral signatures (A - Ammonioalunite, B - Biotite, and C - Carnallite) were selected from the U.S. geological survey (USGS) digital spectral library [40]. The simplex boundary is a triangle whose vertices correspond to these endmembers shown in Fig. 3.1(b). Small and medium dots are simulated mixed spectra belonging to the simplex S_x ($\gamma = 1$) and to the cone C_p ($\gamma > 0$), respectively.

The projective projection of the convex cone C_p onto a properly chosen hyperplane is a simplex with vertices corresponding to the vertices of the simplex S_x. This is illustrated in Fig. 3.2. The simplex $S_p = \{\mathbf{y} \in \mathbb{R}^L : \mathbf{y} = \mathbf{r}/(\mathbf{r}^T\mathbf{u}), \mathbf{r} \in C_p\}$ is the projective projection of the convex cone C_p onto the plane $\mathbf{r}^T\mathbf{u} = 1$, where the choice of $\mathbf{u}$ ensures that there are no observed vectors orthogonal to it.

After identifying S_p, the VCA algorithm iteratively projects data onto a direction orthogonal to the subspace spanned by the endmembers already determined. The new endmember signature corresponds to the extreme of the projection. Fig. 3.2 illustrates the two iterations of the VCA algorithm applied to the simplex S_p defined by the mixture of two endmembers. In the first

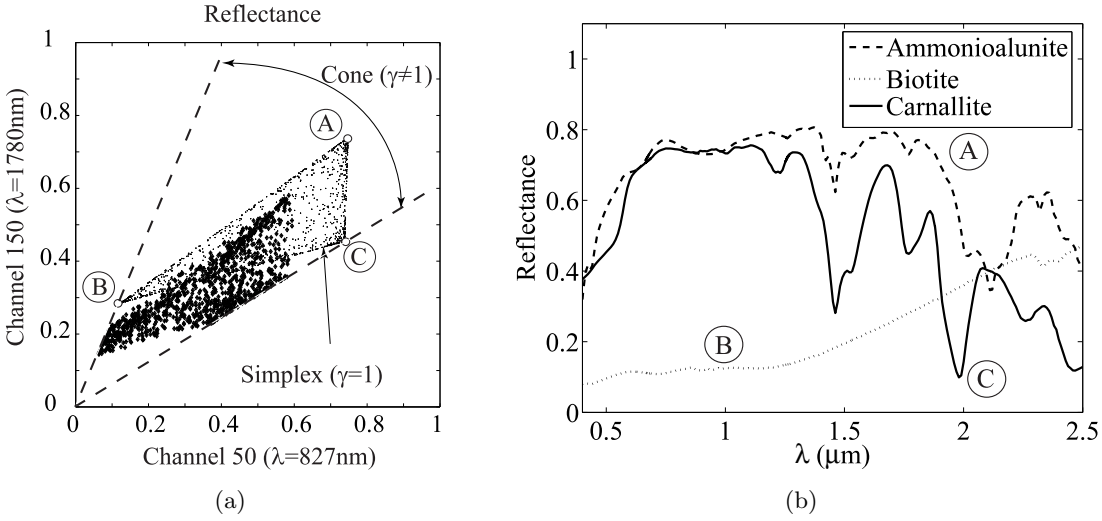


Figure 3.1: (a) 2-D scatter-plot of mixtures of the three endmembers; Circles denote pure materials. (b) Reflectances of Ammonioalunite, Biotite, and Carnallite.

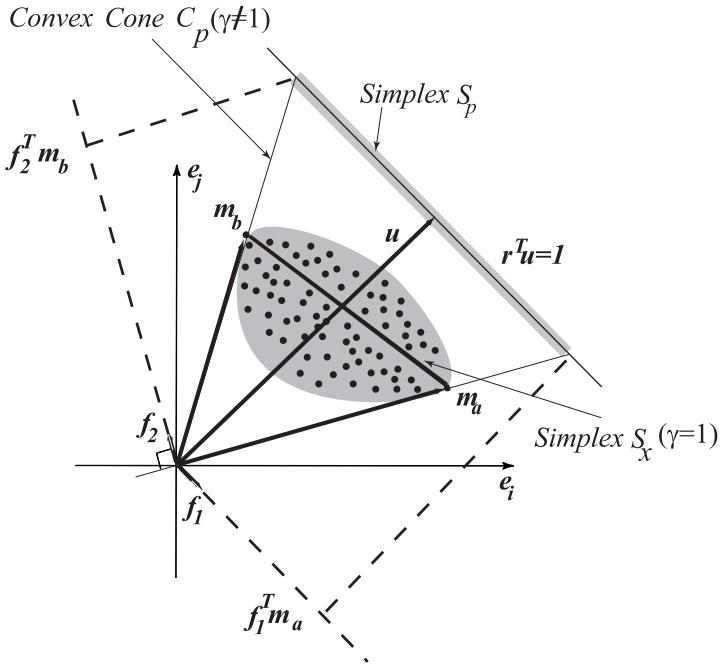


Figure 3.2: Illustration of the VCA algorithm.

iteration, data is projected onto the first direction $\mathbf{f}_1$. The extreme of the projection corresponds to endmember $\mathbf{m}_a$. In the next iteration, endmember $\mathbf{m}_b$ is found by projecting data onto direction $\mathbf{f}_2$, which is orthogonal to $\mathbf{m}_a$. The algorithm iterates until the number of endmembers is exhausted.

3.2.1 Dimensionality Reduction

Although VCA can be applied to data sets without dimensionality reduction, this pre-processing step is usually implemented. Since the dimensionality of data (number of endmembers) is usually much lower than the number of bands, dimensionality reduction leads to significant savings in computational complexity and to SNR improvements. Hysime, introduced in Chapter 2, is a dimensionality reduction method oriented to hyperspectral data.

To illustrate the role of dimensionality reduction, a simulated scene is generated according to expression (3.1) with the three spectral signatures presented in Fig. 3.1(b). The abundance fractions follow a Dirichlet distribution, parameter γ is set to 1, and the noise is zero-mean white Gaussian with covariance matrix $\sigma^2\mathbf{I_L}$, where $\sigma = 0.045$ leading to a SNR = 20 dB [see expression (2.4)]. Fig. 3.3(a) presents a scatter-plot of the simulated spectral mixtures without projection (bands $\lambda = 827nm$ and $\lambda = 1780nm$). Two triangles whose vertices represent the true endmembers (solid line) and the estimated endmembers (dashed line) by the VCA algorithm are also plotted. Fig. 3.3(b) presents a scatter-plot (same bands) of projected data onto the estimated affine set of dimension 2 inferred by the HySime method. Note that by using the HySime method the noise is clearly reduced, leading to a visible improvement on the VCA results.

As discussed before, in the absence of noise, the observed vectors $\mathbf{r}$ lie in a convex cone C_p contained in the subspace $\langle \mathbf{E}_p \rangle$ spanned by the columns of matrix $\mathbf{E}_p \equiv [\mathbf{e}_1, \ldots, \mathbf{e}_p]$, where the singular vectors, $\mathbf{e}_j$, are ordered by the decreasing magnitude of the respective singular values. The VCA algorithm uses the HySime method to identify the subspace $\langle \mathbf{E}_p \rangle$ and then it projects the observed vectors in C_p onto the simplex S_p by computing $\mathbf{y} = \mathbf{r}/(\mathbf{r}^T\mathbf{u})$ (see Fig. 3.2). This

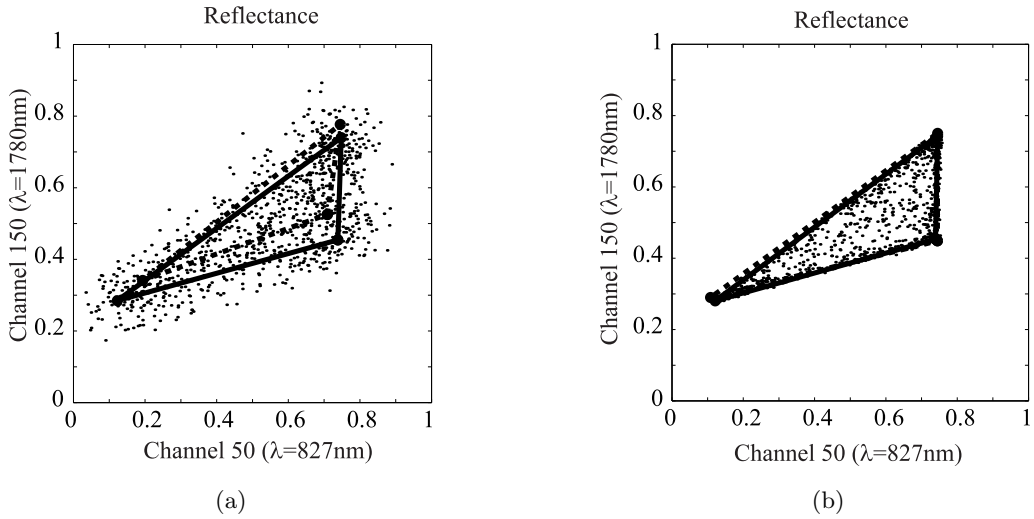


Figure 3.3: Scatter-plot (bands $\lambda = 827nm$ and $\lambda = 1780nm$) of the three endmembers mixture: (a) unprojected data; (b) projected data using HySime. Solid and dashed lines represent simplexes computed from original and estimated endmembers (using VCA), respectively.

simplex is contained in an affine set of dimension $p-1$. Note that the rational underlying the VCA algorithm is still valid if the observed data set is projected onto any subspace $\langle \mathbf{E}_d \rangle \supset \langle \mathbf{E}_p \rangle$ of dimension d, for $p \leq d \leq L$; *i.e.*, the projection of the cone C_p onto $\langle \mathbf{E}_d \rangle$ followed by a projective projection is also a simplex with the same vertices. Of course, the SNR decreases as d increases.

Note, however, that as the SNR decreases the rescaling $\mathbf{r}/(\mathbf{r}^T\mathbf{u})$, to remove the topographic modulation factor, amplifies the noise. Thus, it is preferable to identify directly the affine space of dimension $p-1$ by using the PCA in the HySime algorithm. This phenomenon is illustrated in Fig. 3.4, where data clouds (noiseless and noisy) generated by two signatures are shown. Affine spaces $\langle \mathbf{A}_{p-1} \rangle$ and $\langle \mathbf{A}'_{p-1} \rangle$ identified by PCA of dimension $p-1$ and SVD of dimension p, respectively, followed by projective projection are schematized by straight lines. In the absence of noise, the direction of $\mathbf{m}_a$ is better identified by projective projection onto $\langle \mathbf{A}'_{p-1} \rangle$ ($\widehat{\mathbf{m}}_a$ better than $\widehat{\widehat{\mathbf{m}}}_a$); in the presence of strong noise, the direction of $\mathbf{m}_a$ is better identified by orthogonal projection onto $\langle \mathbf{A}_{p-1} \rangle$ ($\widehat{\widehat{\mathbf{m}}}_a$ better than $\widehat{\mathbf{m}}_a$). As a conclusion, when the SNR is higher than a given threshold SNR_{th}, data is projected onto $\langle \mathbf{E}_p \rangle$ followed by the rescaling $\mathbf{r}/(\mathbf{r}^T\mathbf{u})$; otherwise data is projected onto $\langle \mathbf{A}_{p-1} \rangle$. Based on experimental results, the threshold $\mathrm{SNR}_{th} = 15 + 10\log_{10}(p)\,\mathrm{dB}$

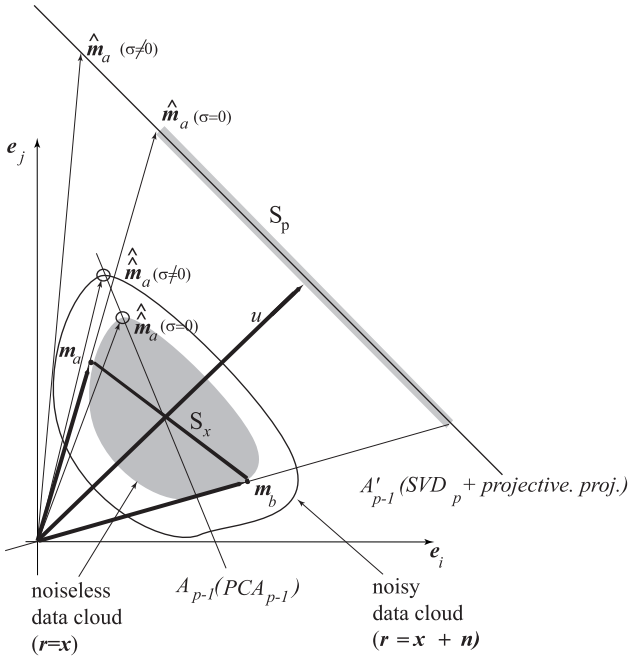


Figure 3.4: Illustration of the noise effect on the dimensionality reduction.

is proposed. Since for zero-mean white noise SNR $= \mathrm{E}\big[\mathbf{x}^T\mathbf{x}\big]/(L\sigma^2)$, then we conclude that at SNR$_{th}$, $\mathrm{E}\big[\mathbf{x}^T\mathbf{x}\big]/(p\sigma^2) = 10^{1.5}L$; *i.e.*, the SNR$_{th}$ corresponds to the fixed value $L \times 10^{1.5}$ of the SNR measured with respect to the signal subspace.

3.2.2 VCA Algorithm Description

The pseudo-code for the VCA method is shown in Algorithm 3.1. Symbols $[\widehat{\mathbf{M}}]_{:,j}$ and $[\widehat{\mathbf{M}}]_{:,i:k}$ stand for the jth column of $\widehat{\mathbf{M}}$ and for the ith to kth columns of $\widehat{\mathbf{M}}$, respectively. Symbol $\widehat{\mathbf{M}}$ stands for the estimated mixing matrix.

Step 2 estimates the number of endmembers with the HySime method developed in Chapter 2. Note that in the simulated data experiments this step is optional since the number of endmembers is known and can be an input of VCA.

Step 4 test if the SNR is higher than SNR$_{th}$ in order to decide whether the data is to be projected

Algorithm 3.1 :Vertex Component Analysis (VCA)

INPUT $\mathbf{R} \equiv [\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N]$
$p :=$ HySime($\mathbf{R}$); {the number of endmembers is estimated with HySime algorithm}
$\text{SNR}_{th} := 15 + 10 \log_{10}(p)$ dB;
if SNR $>$ SNR_{th} **then**
  $d := p$;
  $\mathbf{X} := \mathbf{U}_\mathbf{d}^\mathbf{T}\mathbf{R}$; {$\mathbf{U_d}$ obtained by SVD}
  $\mathbf{u} :=$ mean ($\mathbf{X}$); {$\mathbf{u}$ is a $1 \times d$ vector}
  $[\mathbf{Y}]_{:,j} := [\mathbf{X}]_{:,j} / ([\mathbf{X}]_{:,j}^T \mathbf{u})$; {projective projection}
else
  $d := p - 1$;
  $[\mathbf{X}]_{:,j} := \mathbf{U}_\mathbf{d}^\mathbf{T}([\mathbf{R}]_{:,j} - \overline{\mathbf{r}})$; {$\mathbf{U_d}$ obtained by PCA}
  $\kappa := \arg\max_{j=1\ldots N} \|[\mathbf{X}]_{:,j}\|$;
  $\boldsymbol{\kappa} := [\kappa \,|\, \kappa \,|\, \ldots \,|\, \kappa]$; {$\boldsymbol{\kappa}$ is a $1 \times N$ vector}
  $\mathbf{Y} := \begin{bmatrix} \mathbf{X} \\ \boldsymbol{\kappa} \end{bmatrix}$;
end if
$\mathbf{A} := [\mathbf{e}_u \,|\, \mathbf{0} \,|\, \ldots \,|\, \mathbf{0}]$; {$\mathbf{e}_u := [0, \ldots, 0, 1]^T$ and $\mathbf{A}$ is a $p \times p$ auxiliary matrix}
for $i := 1$ to p **do**
  $\mathbf{w} :=$ randn $(0, \mathbf{I}_p)$; {$\mathbf{w}$ is a zero-mean random Gaussian vector of covariance $\mathbf{I}_p$}
  $\mathbf{f} := \frac{(\mathbf{I}_p - \mathbf{A}\mathbf{A}^\#)\mathbf{w}}{\|(\mathbf{I}_p - \mathbf{A}\mathbf{A}^\#)\mathbf{w}\|}$; {$\mathbf{f}$ is a vector orthonormal to the subspace spanned by $[\mathbf{A}]_{:,1:i}$.}
  $\boldsymbol{\nu} := \mathbf{f}^T\mathbf{Y}$;
  $k := \arg\max_{j=1,\ldots,N} |[\boldsymbol{\nu}]_{:,j}|$; {find the projection extreme.}
  $[\mathbf{A}]_{:,i} := [\mathbf{Y}]_{:,k}$;
  $[indice]_i := k$; {stores the pixel index.}
end for
if SNR $>$ SNR_{th} **then**
  $\widehat{\mathbf{M}} := \mathbf{U_d}[\mathbf{X}]_{:,indice}$; {$\widehat{\mathbf{M}}$ is a $L \times p$ estimated mixing matrix}
else
  $\widehat{\mathbf{M}} := \mathbf{U_d}[\mathbf{X}]_{:,indice} + \overline{\mathbf{r}}$; {$\widehat{\mathbf{M}}$ is a $L \times p$ estimated mixing matrix}
end if

onto a subspace of dimension p or $p-1$. In the first case the projection matrix $\mathbf{U_d}$ is obtained by SVD from $\mathbf{R}\mathbf{R}^T/N$. In the second case the projection is obtained by PCA from $(\mathbf{R}-\overline{\mathbf{r}})(\mathbf{R}-\overline{\mathbf{r}})^T/N$ (recall that $\overline{\mathbf{r}}$ is the sample mean of $[\mathbf{R}]_{:,i}$, for $i = 1, \ldots, N$).

Steps 6 and 11 ensure that the inner product between any vector $[\mathbf{X}]_{:,j}$ and vector $\mathbf{u}$ is non-negative, a crucial condition for the VCA algorithm to work correctly. The chosen value of $\kappa = \arg\max_{j=1\ldots N} \|[\mathbf{X}]_{:,j}\|$ ensures that the colatitude angle between $\mathbf{u}$ and any vector $[\mathbf{X}]_{:,j}$ is between $0°$ and $45°$, then avoiding numerical errors that otherwise would occur for angles near $90°$.

Step 16 initializes the auxiliary matrix $\mathbf{A}$, which stores the projection of the estimated endmembers signatures. Assume that there exists at least one pure pixel of each endmember in the input sample $\mathbf{R}$ (see Fig. 3.2). Each time the loop *for* is executed, a vector $\mathbf{f}$ orthonormal to the space spanned by the columns of the auxiliary matrix $\mathbf{A}$ is randomly generated and $\mathbf{y}$ is projected onto $\mathbf{f}$. Knowing that pure endmembers occupy the vertices of a simplex, then $a \leq \mathbf{f}^T[\mathbf{Y}]_{:,i} \leq b$, for $i = 1, \ldots, N$, where values a and b correspond only to pure pixels. The endmember signature corresponding to $\max\big(\big[\,|a|\,|b|\,\big]\big)$ is stored. The next time loop *for* is executed, $\mathbf{f}$ is orthogonal to the space spanned by the signatures already determined. Since $\mathbf{f}$ is the projection of a zero-mean Gaussian independent random vector onto the orthogonal space spanned by the columns of $[\mathbf{A}]_{:,1:i}$, then the probability of $\mathbf{f}$ being null is zero. Note that the underling reason for generating a random vector is only to get a non null projection onto the orthogonal space generated by the columns of $\mathbf{A}$. Fig. 3.2 shows the input samples and the chosen pixels, after the projection $\boldsymbol{\nu} = \mathbf{f}^T\mathbf{Y}$. Then a second vector $\mathbf{f}$ orthonormal to the endmember a is generated and the second endmember is stored. Finally, steps 26 and 28 compute the columns of matrix $\widehat{\mathbf{M}}$, which contain the estimated endmembers signatures in the L-dimensional space.

3.3 Evaluation of VCA with Simulated Data

In this section, a comparison of VCA, PPI, and N-FINDR algorithms is conducted. N-FINDR and PPI were implemented according to [173] and [19], respectively. Regarding PPI, the number of *skewers* must be *large* [21, 26, 138, 139, 164]. Based on Monte Carlo runs, it is concluded that the minimum number of *skewers* beyond which there is no unmixing improvements is about 1000. All experiments are based on simulated scenes from which the signature endmembers and their abundance fractions are known. The estimated endmembers are the columns of $\widehat{\mathbf{M}} \equiv [\widehat{\mathbf{m}}_1, \widehat{\mathbf{m}}_2, \ldots, \widehat{\mathbf{m}}_p]$ and the estimated abundance fractions are given by $\widehat{\mathbf{S}} = \widehat{\mathbf{M}}^{\#}[\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N]$.

To evaluate the performance of the three algorithms, the following vectors of angles are computed: $\boldsymbol{\varepsilon}_{\mathbf{m}} \equiv [\varepsilon_{\mathbf{m}_1}, \varepsilon_{\mathbf{m}_2}, \ldots, \varepsilon_{\mathbf{m}_p}]^T$ and $\boldsymbol{\varepsilon}_{\mathbf{s}} \equiv [\varepsilon_{\mathbf{s}_1}, \varepsilon_{\mathbf{s}_2}, \ldots, \varepsilon_{\mathbf{s}_p}]^T$ with,

$$\varepsilon_{\mathbf{m}_i} \equiv \arccos\left(\frac{(\mathbf{m}_i)^T \widehat{\mathbf{m}}_i}{\|\mathbf{m}_i\| \, \|\widehat{\mathbf{m}}_i\|}\right), \tag{3.2}$$

$$\varepsilon_{\mathbf{s}_i} \equiv \arccos\left(\frac{([\mathbf{S}]_{i,:})^T [\widehat{\mathbf{S}}]_{i,:}}{\|[\mathbf{S}]_{i,:}\| \, \|[\widehat{\mathbf{S}}]_{i,:}\|}\right), \tag{3.3}$$

i.e., $\varepsilon_{\mathbf{m}_i}$ is the angle between vectors $\mathbf{m_i}$ and $\widehat{\mathbf{m}}_i$ (ith endmember signature estimate) and $\varepsilon_{\mathbf{s}_i}$ is the angle between vectors $[\mathbf{S}]_{i,:}$ and $[\widehat{\mathbf{S}}]_{i,:}$ (vectors of $\mathbb{R}^N$ formed by the ith rows of matrices $\widehat{\mathbf{S}}$ and $\mathbf{S} \equiv [\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_N]$, respectively). The *symmetric Kullback distance* [95], a relative entropy-based distance, is another error measure used to compare the similarity between signatures, also known as spectral information divergence (SID) [30]. SID is given by

$$SID_{\mathbf{m}_i, \widehat{\mathbf{m}}_i} \equiv D_{KL}(\mathbf{m}_i \,|\widehat{\mathbf{m}}_i) + D_{KL}(\widehat{\mathbf{m}}_i \,|\mathbf{m}_i), \tag{3.4}$$

where $D_{KL}(\mathbf{m}_i\,|\widehat{\mathbf{m}}_i)$ is the relative entropy of $\mathbf{m}_i$ with respect to $\widehat{\mathbf{m}}_i$ given by

$$D_{KL}(\mathbf{m}_i\,|\widehat{\mathbf{m}}_i) \equiv \sum_{j=1}^{L} p_j \log\left(\frac{p_j}{q_j}\right), \tag{3.5}$$

and $p_j = m_{ij}/\sum_{k=1}^{L} m_{ik}$ and $q_j = \widehat{m}_{ij}/\sum_{k=1}^{L} \widehat{m}_{ik}$.

Based on $\boldsymbol{\varepsilon}_{\mathbf{m}}$, $\boldsymbol{\varepsilon}_{\mathbf{s}}$, and $\boldsymbol{\varepsilon}_{SID} \equiv [SID_{\mathbf{m}_1,\widehat{\mathbf{m}}_1}, SID_{\mathbf{m}_2,\widehat{\mathbf{m}}_2}, \ldots, SID_{\mathbf{m}_p,\widehat{\mathbf{m}}_p}]^T$, the following root mean square error distances are estimated:

$$\mathrm{rms}_{\mathrm{SAE}} = \left(\frac{1}{p}\mathrm{E}\big[\|\boldsymbol{\varepsilon}_{\mathbf{m}}\|^2\big]\right)^{1/2}, \tag{3.6}$$

$$\mathrm{rms}_{\mathrm{SID}} = \left(\frac{1}{p}\mathrm{E}\big[\|\boldsymbol{\varepsilon}_{\mathbf{SID}}\|^2\big]\right)^{1/2}, \tag{3.7}$$

$$\mathrm{rms}_{\mathrm{AFAE}} = \left(\frac{1}{p}\mathrm{E}\big[\|\boldsymbol{\varepsilon}_{\mathbf{s}}\|^2\big]\right)^{1/2}. \tag{3.8}$$

The first two quantities measure distances between $\widehat{\mathbf{m}}_i$ and $\mathbf{m}_i$, for $i = 1, \ldots, p$; the third is similar to the first, but for the estimated abundance fractions. Herein the subscripts SAE and AFAE stand for signature angle error and abundance fraction angle error, respectively. Mean values in expressions (3.6), (3.7), and (3.8) are approximated by sample means based on one hundred of Monte Carlo runs.

In all experiments, the spectral signatures are selected from the U.S. geological survey (USGS) digital spectral library [40]. Fig. 3.1(b) shows three of these endmember signatures. Abundance fractions are generated according to a Dirichlet distribution given by expression (2.2), parameter γ is Beta (β_1, β_2) distributed, *i.e.*,

$$p_{beta}(\gamma) = \frac{\Gamma(\beta_1+\beta_2)}{\Gamma(\beta_1)\Gamma(\beta_2)}\gamma^{\beta_1-1}(\gamma-1)^{\beta_2-1}, \tag{3.9}$$

which is also a Dirichlet distribution with just one component. The expected value and the variance

of γ are $\mu_\gamma = E[\gamma] = \beta_1/(\beta_1+\beta_2)$ and $\sigma_\gamma^2 = \beta_1\beta_2/\big((\beta_1+\beta_2)^2(\beta_1+\beta_2+1)\big)$, respectively. The Dirichlet density, besides enforcing positivity and full additivity constraints, displays a wide range of shapes, depending on the parameters of the distribution. This flexibility underlies its choice in these simulations.

The results next presented are organized into five experiments: in the first experiment, the algorithms are evaluated with respect to the SNR and to the absence of pure pixels, where the SNR is defined in expression (2.4),

$$\mathrm{SNR} \equiv 10\log_{10}\frac{\mathrm{E}\big[\mathbf{x}^T\mathbf{x}\big]}{\mathrm{E}\big[\mathbf{n}^T\mathbf{n}\big]}.$$

In the case of zero-mean noise with covariance $\sigma^2\mathbf{I}_L$ and Dirichlet abundance fractions, one obtains

$$\mathrm{SNR} = 10\log_{10}\frac{\mathrm{tr}\left[\mathbf{M}\mathbf{K}_s\mathbf{M}^T\right]}{L\sigma^2}, \tag{3.10}$$

where

$$\mathbf{K}_s \equiv \sigma_\gamma^2\mathrm{E}\big[\boldsymbol{\alpha}\boldsymbol{\alpha}^T\big] = \sigma_\gamma^2\frac{\boldsymbol{\theta}\boldsymbol{\theta}^T + \mathrm{diag}\,(\boldsymbol{\theta})}{\big(\sum_{i=1}^p\theta_i\big)\big(1+\sum_{i=1}^p\theta_i\big)}, \tag{3.11}$$

and $\boldsymbol{\theta} = \big[\theta_1 \ldots \theta_p\big]^T$. For example, assuming abundance fractions equally distributed, $\mathrm{SNR} \simeq 10\log_{10}\sigma_\gamma^2\sum_{i=1}^p\big(\sum_{j=1}^p m_{ij}^2/p\big)/(L\sigma^2)$ for $\theta p \ll 1$ and $\mathrm{SNR} \simeq 10\log_{10}\sigma_\gamma^2\big(\sum_{i=1}^p\big(\sum_{j=1}^p m_{ij}\big)^2/p^2\big)/(L\sigma^2)$ for $\theta p \gg 1$ are obtained after some algebra.

In the second experiment, the performance is measured as a function of the parameter γ, which models fluctuations on the illumination due to the surface topography. In the third experiment, the number of pixels of the scene varies with the size of the covered area: as the number of pixels increases, the likelihood of having pure pixels also increases, improving the performance of the unmixing algorithms. In the fourth experiment, the algorithms are evaluated as a function of the number of endmembers present in the scene; finally, in the fifth experiment, the number of floating

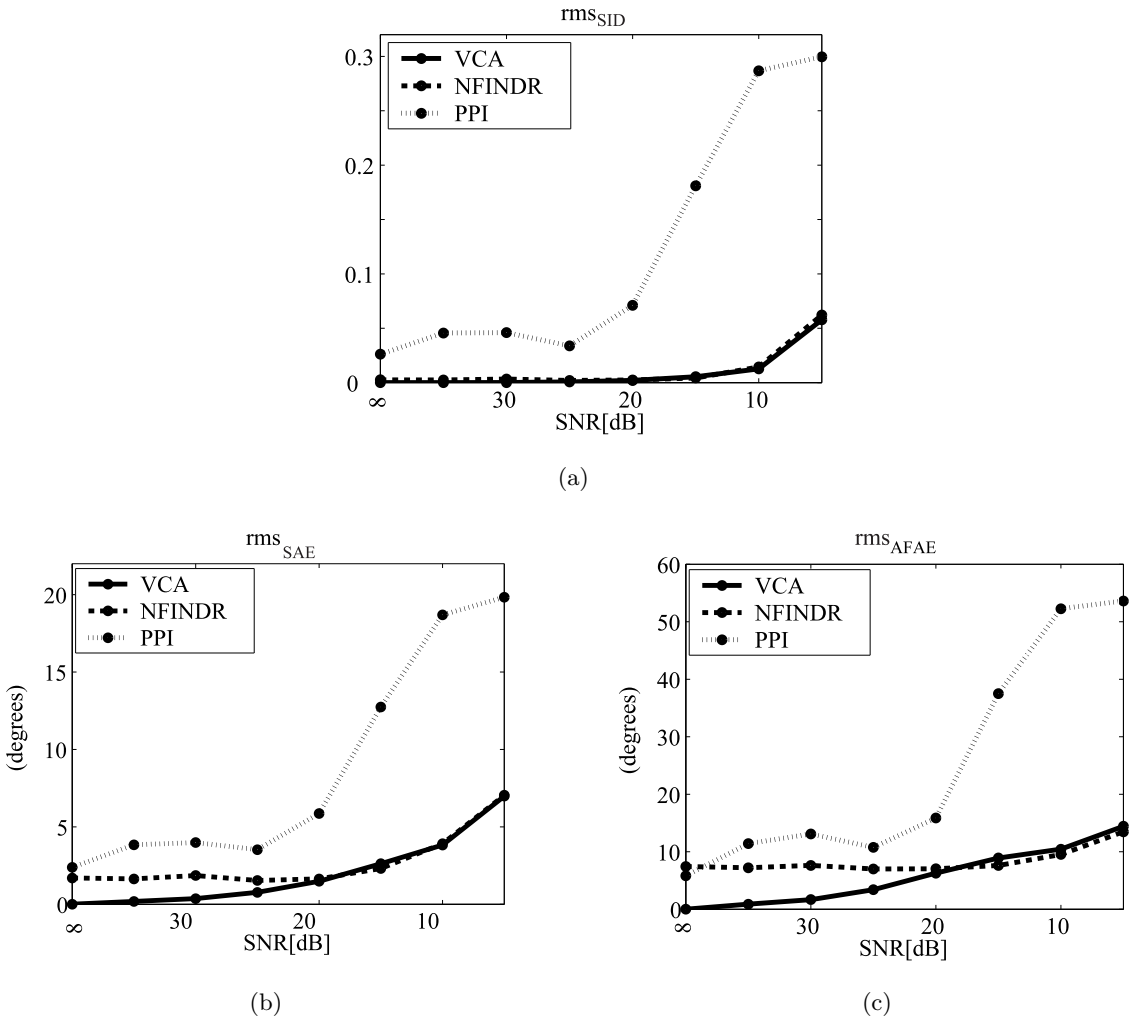


Figure 3.5: Performance measures as a function of SNR, for $N = 1000$, $p = 3$, $L = 224$, $\theta_1 = \theta_2 = \theta_3 = 1/3$, $\beta_1 = 20$,and $\beta_2 = 1$: (a) rms_{SID} as a function of SNR; (b) rms_{SAE} as a function of SNR; (c) rms_{AFAE} as a function of SNR.

point operations (flops) is measured, to compare the computational complexity of the VCA, N-FINDR, and PPI algorithms.

Experiment I

In this experiment, the hyperspectral scene has one thousand pixels and the abundance fractions are Dirichlet distributed with $\theta_j = 1/3$, for $j = 1, 2, 3$; parameter γ is Beta distributed with $\beta_1 = 20$ and $\beta_2 = 1$ implying $\text{E}[\gamma] = 0.952$ and $\sigma_\gamma = 0.05$.

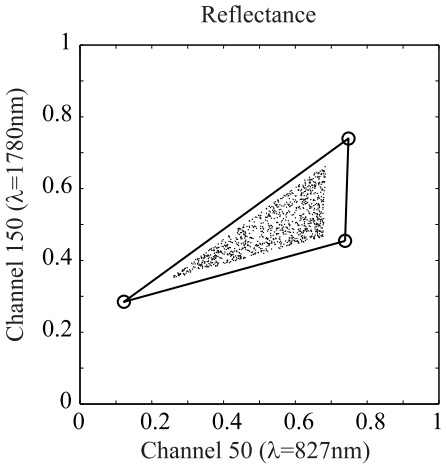
Fig. 3.5 shows the performance results as a function of the SNR. As expected, the presence of noise degrades the performance of all algorithms. In terms of rms_{SID}, VCA and N-FINDR algorithms have identical performances, whereas PPI displays the worst result. In terms of rms_{SAE} and rms_{AFAE}, shown in Fig. 3.5(b) and Fig. 3.5(c), when SNR is less than 20 dB, the VCA algorithm exhibits the best performance. Note that for noiseless scenes, only VCA has $\text{rms}_{\text{SAE}} = 0$.

Fig. 3.6 shows the performance results as a function of the SNR in the absence of pure pixels. Spectral data without pure pixels is obtained by rejecting pixels with any abundance fraction smaller than 0.2. Fig. 3.6(a) shows the obtained scatter plot. VCA and N-FINDR display similar results, being both better than PPI. Note that the performance is almost independent of the SNR and is uniformly worse than that displayed with pure pixels and SNR = 5 dB in the first experiment. We conclude that this family of algorithms is more affected by the lack of pure pixels than by low SNR.

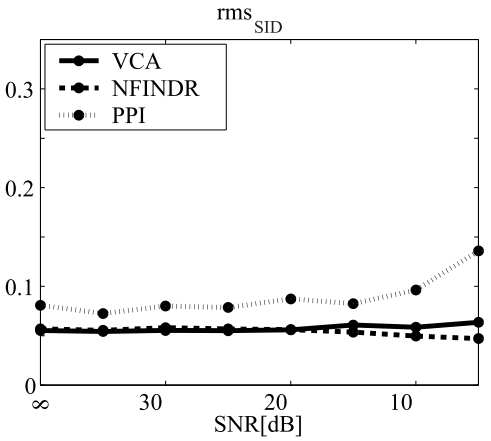
Since rms_{SID}, rms_{SAE}, and rms_{AFAE} disclose similar pattern of behavior, only rms_{SAE} is presented in the remaining experiments.

Experiment II

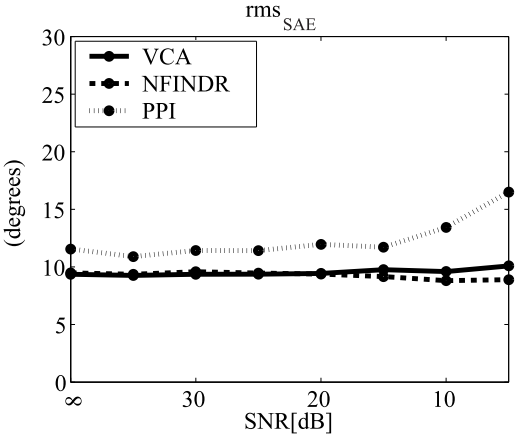
In this experiment, abundance fractions are generated as in the first one, the SNR is set to 20 dB, and the parameter γ is Beta distributed with $\beta_2 = 1$ and β_1 in the interval $\{2, \ldots, 28\}$. This corresponds to vary $\text{E}[\gamma]$ from 0.66 to 0.96 and σ_γ from 0.23 to 0.03. By varying parameter β_1, the severity of topographic modulation is also modified. Fig. 3.7 illustrates the effect of topographic modulation on the performance of the three algorithms. When β_1 grows (σ_γ reduces) the performance improves. This is expected, since the simplex identification is more accurate when the topographic modulation is smaller. PPI algorithm displays the worst performance for $\sigma_\gamma < 0.1$. VCA and N-FINDR algorithms have identical performances when β_1 takes higher values ($\sigma_\gamma < 0.045$), otherwise VCA algorithm has the best performance. VCA is more robust to topo-



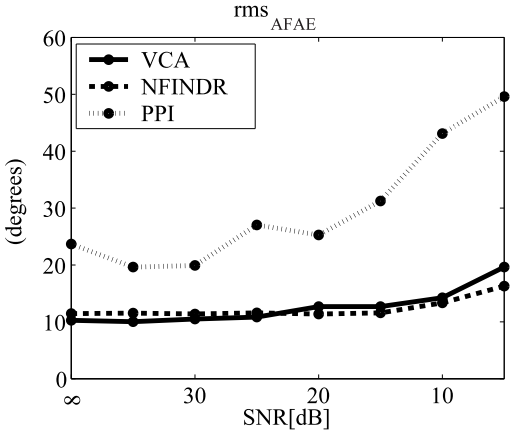
(a)



(b)



(c)



(d)

Figure 3.6: Illustration of the absence of pure pixels, for $N = 1000$, $p = 3$, $L = 224$, $\theta_1 = \theta_2 = \theta_3 = 1/3$, $\beta_1 = 20$, and $\beta_2 = 1$: (a) Scatter-plot (bands $\lambda = 827nm$ and $\lambda = 1780nm$), with abundance fraction smaller than 0.2 rejected and $\gamma = 1$; (b) rms_{SID} as a function of SNR; (c) rms_{SAE} as a function of SNR; (d) rms_{AFAE} as a function of SNR.

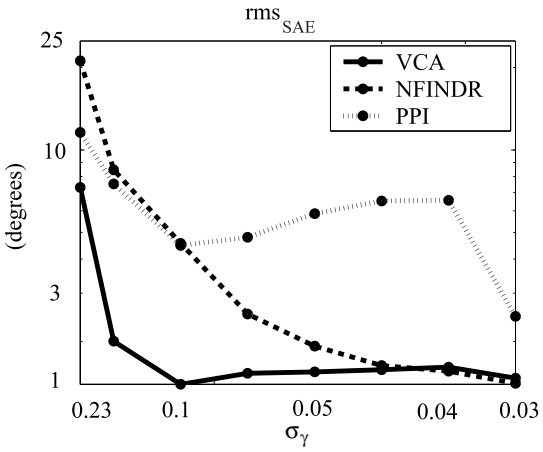


Figure 3.7: Robustness to the topographic modulation, for $N = 1000$, $p = 3$, $L = 224$, $\theta_1 = \theta_2 = \theta_3 = 1/3$, SNR $= 20\,\mathrm{dB}$, and $\beta_2 = 1$: $\mathrm{rms}_{\mathrm{SEA}}$ as a function of the σ_γ^2 (variance of γ).

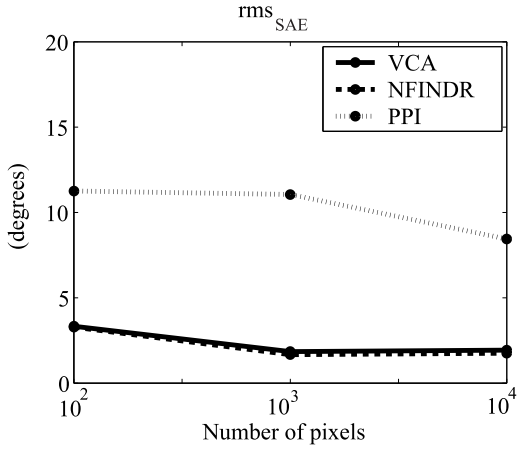


Figure 3.8: $\mathrm{rms}_{\mathrm{SAE}}$ as a function of the number of pixels in a scene, for $p = 6$, $L = 224$, $\theta_1 = \theta_2 = \theta_3 = 1/3$, SNR $= 20\,\mathrm{dB}$, $\beta_2 = 20$, and $\beta_2 = 1$.

graphic modulation, because it seeks for the extreme projections of the simplex, whereas N-FINDR seeks for the maximum volume, which is more sensitive to fluctuations of γ.

Experiment III

In this experiment, the number of pixels is varied, the abundance fractions are generated as in the first one, and the SNR $= 20\,\mathrm{dB}$. Fig. 3.8 shows that VCA and N-FINDR exhibit identical results, whereas, the PPI algorithm displays the worst result. Note that the behavior of the three algorithms is quasi-independent of the number of pixels.

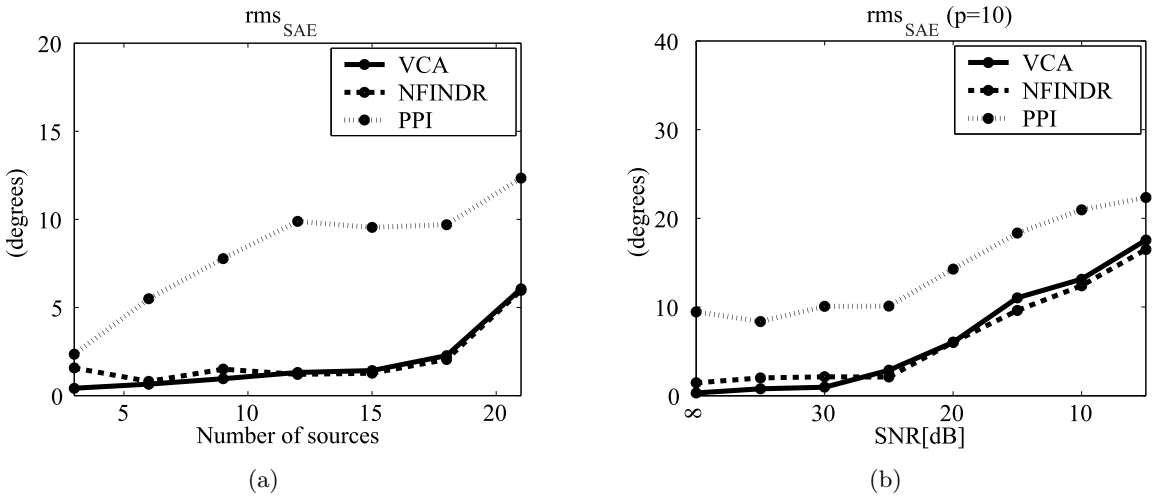


Figure 3.9: Impact of the number of endmembers, for $N = 1000$, $L = 224$, $\theta_1 = \theta_2 = \theta_3 = 1/3$, SNR $= 30\,\text{dB}$, $\beta_2 = 20$, and $\beta_2 = 1$: (a) rms_{SEA} as a function of the number of endmembers; (b) rms_{SEA} as a function of the SNR with $p = 10$.

Experiment IV

In this experiment, the number of signatures vary from $p = 3$ to $p = 21$, the scene has one thousand pixels, and the SNR $= 30\,\text{dB}$. Fig. 3.9(a) shows that VCA and N-FINDR performances are comparable, while PPI displays the worst result. The rms_{SAE} increases slightly as the number of endmembers present in the scene increases. The rms_{SAE} as a function of the SNR with $p = 10$ is also plotted, see Fig. 3.9(b). Compared with Fig. 3.5(b) it is concluded that when the number of endmembers increases the performance of the algorithms slightly decreases.

Experiment V

In this experiment, the number of floating point operations is measured, in order to compare the computational complexity of the VCA, PPI, and N-FINDR algorithms. Herein, the scenarios of the second and third experiments are used. Table 3.1 presents approximated expressions for the number of floating point operations used by each algorithm. These expressions do not account for the computational complexities involved in the computations of the sample covariance $(\mathbf{R} - \overline{\mathbf{r}})(\mathbf{R} - \overline{\mathbf{r}})^T/N$, nor in the eigendecomposition. The reason is that these operations, compared with

Table 3.1: Computational complexity of VCA, N-FINDR, and PPI algorithms.

Algorithm	Complexity (flops)
VCA	$2p^2N$
N-FINDR	$p^{\delta+1}N$
PPI	$2psN$

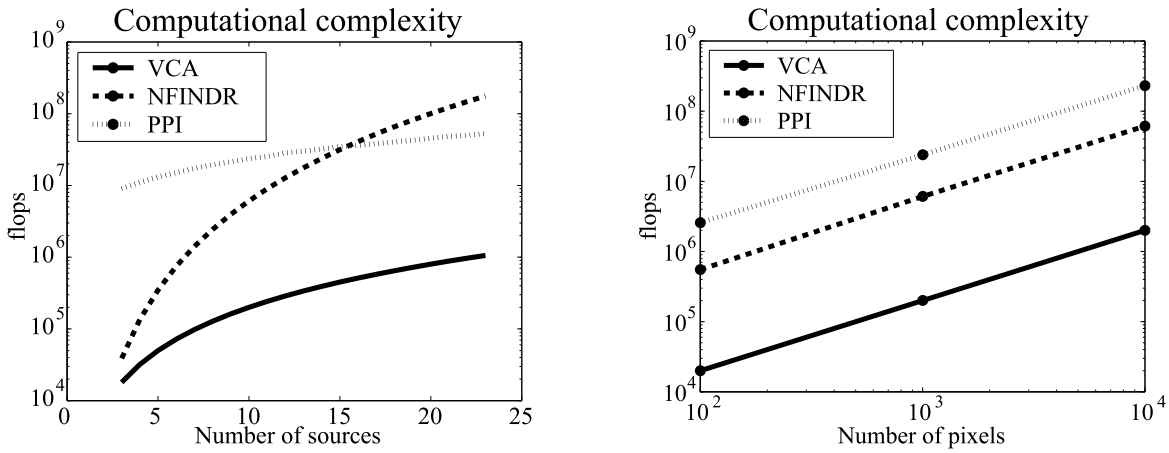
the remaining steps of VCA, PPI, and N-FINDR algorithms, have a negligible computational cost since:

1. The computation of $(\mathbf{R}-\overline{\mathbf{r}})(\mathbf{R}-\overline{\mathbf{r}})^T/N$ has a complexity of $2NL^2$ flops. However, in practice one does not need to use the complete set of N hyperspectral vectors. If the scene is noiseless, only $p-1$ linearly independent vectors would be enough to infer the exact subspace $\langle\mathbf{E}_{p-1}\rangle$. In the presence of noise, however, a larger set should be used. For example in a 1000×1000 hyperspectral image, it was found that only 1000 random samples are enough to get a very good estimate of $\langle\mathbf{E}_{p-1}\rangle$. Even a sample size of 100 leads to good results in this respect.

2. Concerning the eigendecomposition of $(\mathbf{R}-\overline{\mathbf{r}})(\mathbf{R}-\overline{\mathbf{r}})^T/N$ (or the SVD of $\mathbf{R}\mathbf{R}^T/N$), one only needs to compute $p-1$ (or p) eigenvectors corresponding to the largest $p-1$ eigenvalues (or p single values). For these partial eigendecomposition, the PCA algorithm is used [82], (or SVD analysis [149]) whose complexity is negligible compared with the remaining operations.

The VCA algorithm projects all data (N vectors of size p) onto p orthogonal directions. N-FINDR computes pN times the determinant of a $p \times p$ matrix, whose complexity is p^δ, with $2.3 < \delta < 2.9$ (see, *e.g.*, [84]). Assuming that $N \gg p > 2$, the VCA complexity is lower than that of N-FINDR. Concerning PPI, given that the number of *skewers* (s) is much higher than the usual number of endmembers, the PPI complexity is much higher than that of VCA. The conclusion is that the VCA algorithm has always the lowest complexity.

Fig. 3.10 plots the flops for the three algorithms after data projection. On Fig. 3.10(a), the



(a) (b)

Figure 3.10: Computational complexity measured in number of floating point operations (flops): (a) as a function of the number of sources; (b) as a function of the number of pixels.

abscissa is the number of endmembers in the scene, whereas on Fig. 3.10(b), the abscissa is the number of pixels. Note that for five endmembers, the VCA computational complexity is one order of magnitude lower than that of the N-FINDR algorithm. When the number of endmembers is higher than 15, the VCA computational complexity is, at least, two orders of magnitude lower than PPI and N-FINDR algorithms.

The results presented in this section are based on abundance fractions with symmetric Dirichlet distribution. The same pattern of behavior was, however, found for any other abundance fraction distribution tested. Fig. 3.11 shows the results for abundance fractions $\alpha_j = x_j^\beta / \sum_{l=1}^{p} x_i^\beta$ for $j = 1, \ldots, p$, with x_j uniformly distributed in [0, 1] and $\beta = 1/2$ (a), $\beta = 1$ (b), $\beta = 2$ (c), and $\beta = 3$ (d). Dirichlet density with $\theta_1 = \theta_2 = \theta_3 = 1/3$ is included (e) for comparison purposes. The VCA algorithm display always the lowest rms_{SAE}. The performance increases for all algorithms as β increases, because the likelihood of having pure pixels increases with β.

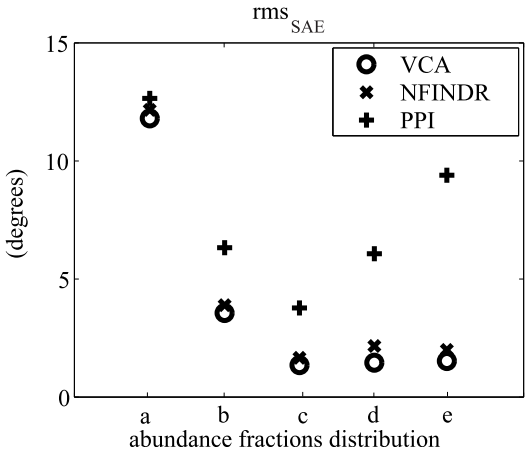


Figure 3.11: rms_{SAE} as a function of different abundance fractions distributions with $N = 1000$, $p = 3$, $L = 224$, SNR $= 20\,\text{dB}$, $\beta_2 = 20$, and $\beta_2 = 1$. From left to right: (a) $\beta = 1/2$; (b) $\beta = 1$; (c) $\beta = 2$; (d) $\beta = 3$; (e) Dirichlet density ($\theta_1 = \theta_2 = \theta_3 = 1/3$).

3.4 Experiments with Real Hyperspectral Data

In this section, the VCA algorithm is applied to real hyperspectral data collected by the AVIRIS sensor over Cuprite, Nevada (see Section 2.5 for more details). In order to compare results with a signature library, the reflectance image after atmospheric correction is used.

To estimate the number of endmembers present in the processed area, the HySime algorithm is applied to the data set. The study conducted in Section 2.5 leads to the conclusion that the bulk of spectral energy is due to 8 materials.

To determine the type of projection applied by VCA, an estimate of the SNR is computed

$$\text{SNR} \simeq 10 \log_{10} \frac{P_{R_p} - (p/L) P_R}{P_R - P_{R_p}}, \tag{3.12}$$

where $P_R \equiv \text{E}\left[\mathbf{r}^T \mathbf{r}\right]$ and $P_{R_p} \equiv \text{E}\left[\mathbf{r}^T \mathbf{U}_d \mathbf{U}_d^T \mathbf{r}\right]$ in the case of SVD and $P_{R_p} \equiv E\left[\mathbf{r}^T \mathbf{U}_d \mathbf{U}_d^T \mathbf{r}\right] + \bar{\mathbf{r}}^T \bar{\mathbf{r}}$ in the case of PCA.

A visual comparison between VCA results on the Cuprite data set and the ground truth presented in [161], shows that the first component [see Fig. 3.12(a)] is predominantly Alunite, the second component [see Fig. 3.12(b)] is Sphene, the third component [see Fig. 3.12(c)] is Bud-

dingtonite, the fourth component [see Fig. 3.12(d)] is Montmorillonite, the fifth, seventh, and the eighth components [see Fig. 3.12(e), 3.12(g), and 3.12(h)] are Kaolinite, the sixth component [see Fig. 3.12(f)] is predominantly Nontronite.

To confirm the classification based on the estimated abundance fractions, a comparison between the estimated VCA endmember signatures and the laboratory spectrum [40] is presented in Fig. 3.13. The signatures provided by VCA are scaled by a factor in order to minimize the mean square error between them and the respective library spectra. The estimated signatures are close to the laboratory spectra. The larger mismatches occur for Buddingtonite and Kaolinite (#1) signatures, but only on a small percentage of the total bands.

Table 3.2 compares the spectral angles between extracted endmembers and laboratory reflectances for the VCA, N-FINDR, and PPI algorithms. The first column shows the laboratory substances with smaller spectral angle distance with respect to the signature extracted by the VCA algorithm; the second column shows the respective angle. The third and the fourth columns are similar to the second one, except when the closest spectral substance is different from the correspondent VCA one. In these cases, the name of the substance is presented. The displayed results follow the pattern of behavior shown in the simulations, where VCA performs better than PPI and better or similarly to N-FINDR.

Table 3.2: Spectral angle distance (in degrees) between extracted endmembers and laboratory reflectances for VCA, N-FINDR, and PPI algorithms.

Substance	VCA	N-FINDR	PPI
Alunite or Montmorillonite	3.9	3.9	4.3
Sphene	3.1	Barite (2.7)	Pyrope (3.9)
Buddingtonite	4.2	4.1	3.9
Montmorillonite	3.1	3.0	2.9
Kaolinite #1	5.7	5.3	Dumortierite (5.3)
Nontronite or Kaolinite	3.4	4.8	4.7
Kaolinite #2	3.5	Montmor. (4.2)	3.5
Kaolinite #3	4.2	4.3	5.0

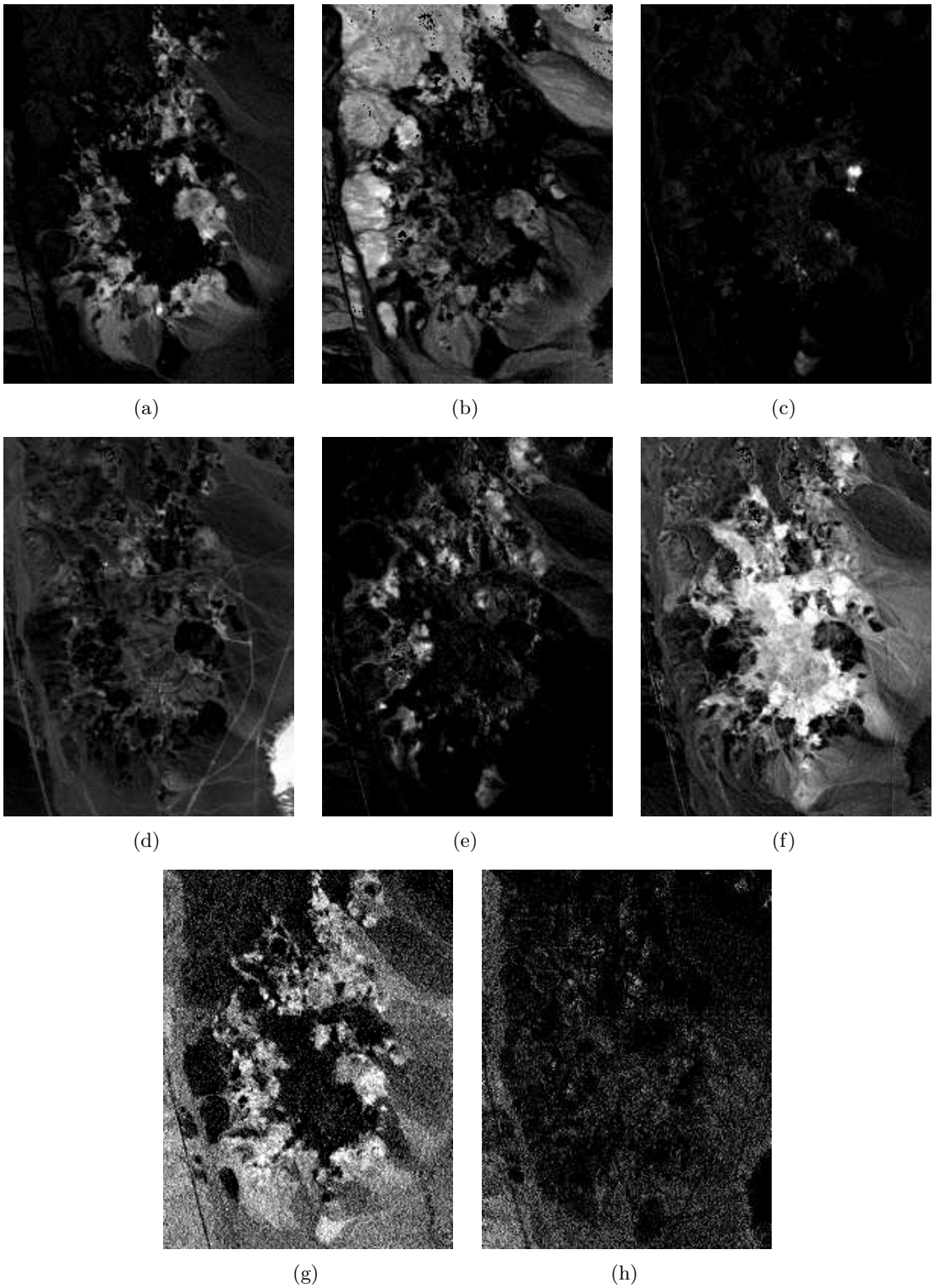


Figure 3.12: Eight abundance fractions estimated with VCA algorithm: (a) Alunite or Montmorillonite; (b) Sphene; (c) Buddingtonite; (d) Montmorillonite; (e) Kaolinite #1; (f) Nontronite or Kaolinite; (g) Kaolinite #2; (h) Kaolinite #3.

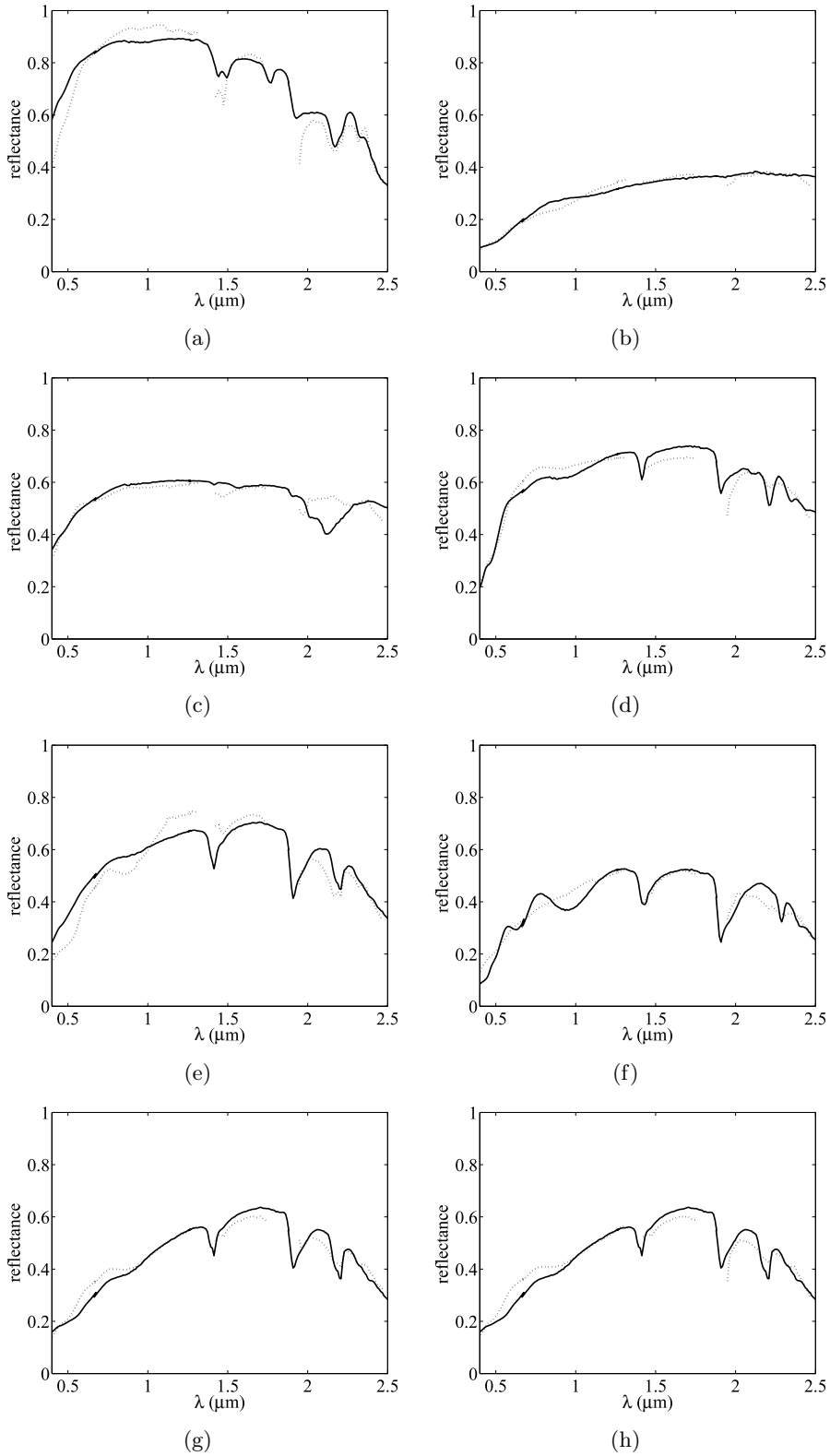


Figure 3.13: Comparison of the extracted signatures (dotted line) with the U.S.G.S spectral library (solid line): (a) Alunite or Montmorillonite; (b) Sphene; (c) Buddingtonite; (d) Montmorillonite; (e) Kaolinite #1; (f) Nontronite or Kaolinite; (g) Kaolinite #2; (h) Kaolinite #3.

3.5 Conclusions

In this chapter a new algorithm to unmix linear mixtures of hyperspectral sources, termed *vertex component analysis* (VCA) has been proposed. VCA algorithm is unsupervised and is based on the geometry of hyperspectral data sets. It exploits the fact that endmembers occupy the vertices of a simplex.

VCA assumes the presence of pure pixels in the data and iteratively projects data onto a direction orthogonal to the subspace spanned by the endmembers already determined. The new endmember signature corresponds to the extreme of the projection. The algorithm iterates until the number of endmembers is exhausted. The number of endmembers is determined by the Hysime method introduced in Chapter 2.

A comparison of VCA with PPI and N-FINDR algorithms is conducted. Several experiments with simulated data lead to the conclusion that VCA performs better than PPI and better than or similar to N-FINDR. However, VCA has the lowest computational complexity among these three algorithms[1]. Savings in computational complexity ranges between one and two orders of magnitude. This conclusion has great impact when the data set has a large number of pixels. VCA was also applied to real hyperspectral data. The results achieved show the effectiveness of VCA on the unmixing procedure.

[1] After this study, a fast iterative implementation of PPI [34, 140] and an improved version of N-FINDR [141] were proposed in order to reduce the computational complexity of those algorithms.

Chapter 4

The Role of Independent Component Analysis on Hyperspectral Data Unmixing

4.1 Introduction

In Chapter 3 the linear unmixing problem was addressed under a geometrical point of view. The approach therein proposed, termed VCA, unmixes hyperspectral mixtures exploiting the fact that the endmembers correspond to the vertices of the simplex containing the hyperspectral vectors. Implicit in VCA, as in other geometrical approaches, is the assumption that there are at least one pure pixel per endmember in the data. This is a strong requisite that may not hold in many data sets. In any case, VCA finds the set of most pure pixels in the data.

Under the linear mixing model and assuming that the number of endmembers and their spectral signatures are known, hyperspectral unmixing is a linear problem, which can be addressed, for example, under the ML setup [151], the constrained least squares approach [31], the spectral signature matching [112], the spectral angle mapper (SAM)[176], and the subspace projection methods [31, 35, 68]. The *maximum a posteriori* (MAP) probability framework [133] and the projection pursuit [77, 81] have also been applied to hyperspectral data.

In most cases the number of endmembers and their signatures are not known. Independent component analysis (ICA) is an unsupervised source separation process that finds a linear decom-

position of the observed data yielding statistically independent components. It has been applied with success to blind source separation, to feature extraction, and to unsupervised recognition [41, 74].

Given that hyperspectral data are, under given circumstances, linear mixtures, ICA comes to mind as a possible tool to unmix this class of data. In fact, the application of ICA to hyperspectral data has been proposed in [11], where endmember signatures are treated as sources and the mixing matrix is composed by the abundance fractions, and in [24, 36, 38, 50, 86, 88, 89, 103, 133, 165], where sources are the abundance fractions of each endmember. The first approach faces two difficulties: i) The number of samples are limited to the number of channels; ii) The process of pixel selection, playing the role of mixed sources, is not straightforward. The second approach also faces difficulties since the sum of the abundance fractions is constant, implying statistical dependence among abundances (*i.e.*, among sources). This dependence violates a key assumption of ICA, which is the source statistical independence. The applicability of ICA to hyperspectral images is thus compromised. In addition, hyperspectral data are immersed in noise, which degrades the ICA performance.

Independent factor analysis (IFA) [6, 117] was introduced as a method for recovering independent hidden sources from their observed noisy mixtures. IFA implements two steps: first, source densities and noise covariance are estimated from the observed data by ML; second, sources are reconstructed by an optimal nonlinear estimator. Although IFA is a well suited technique to unmix independent sources under noisy observations, the dependence among abundance fractions in hyperspectral imagery compromises, as in the ICA case, the IFA performance.

This chapter addresses the impact of the source dependence on unmixing hyperspectral data with ICA and IFA algorithms. It is shown that these algorithms do not correctly unmix hyperspectral data. We put in evidence that the unmixing matrix minimizing the mutual information might be very far from the true one. Nevertheless, some abundance fractions might be well separated,

mainly in the presence of strong signature variability, large number of endmembers, and high SNR. At the end of the chapter, a method based on the source entropy is proposed, to sort the output of ICA or IFA algorithms according to the likelihood of being correctly separated. Chapter 5 presents a new direction to blindly unmix hyperspectral data, where abundance fraction dependence is taken into account.

The chapter is organized as follows. Section 4.2 presents a brief overview of the ICA and IFA algorithms. Section 4.3 illustrates the performance of IFA and of some well known ICA algorithms with simulated hyperspectral data. Section 4.4 studies the ICA and IFA limitations in unmixing hyperspectral data. Section 4.5 presents results of ICA based on real data. Section 4.6 concludes the chapter with some remarks. Parts of this chapter were published in [118, 119, 122, 124, 128].

4.2 Independent Component Analysis and Independent Factor Analysis

ICA [41, 73, 74] is an unsupervised source separation process, which has been applied to linear blind separation problems [12, 29, 75]. The goal of ICA is to recover independent sources, given only sensor observations that are unknown linear mixtures of the unobserved independent sources.

Let $\mathbf{r}$ be an $L \times 1$ observation vector, such that

$$\mathbf{r} = \mathbf{M}\mathbf{s}, \tag{4.1}$$

where $\mathbf{M}$ is an unknown $L \times p$ ($L \geq p$) mixing matrix and $\mathbf{s} \equiv [s_1\ s_2\ ...\ s_p]^T$ is an unknown random data vector of mutually independent sources with unknown distributions, although, at most, one can be Gaussian distributed. ICA finds a $p \times L$ separating matrix $\mathbf{W}$, such that

$$\mathbf{y} = \mathbf{W}\mathbf{r} = \mathbf{P}\mathbf{G}\mathbf{s}, \tag{4.2}$$

where $\mathbf{y}$ is a vector of independent components and $\mathbf{P}$ and $\mathbf{G}$ are permutation and scale matrices, respectively.

ICA looks for a linear representation that maximizes a nongaussianity measure [100]. A commonly objective function used in ICA algorithms is the mutual information [41] of vector $\mathbf{y} \equiv [y_1\ y_2\ \ldots\ y_p]^T$ given by

$$\begin{aligned} I(y_1, y_2, \ldots, y_p) &\equiv \sum_j H(y_j) - H(\mathbf{y}) \\ &= \int p(\mathbf{y}) \log \frac{p(\mathbf{y})}{\prod_j p(y_j)} \mathrm{d}\mathbf{y}. \end{aligned} \tag{4.3}$$

From expression (4.3), we can see that the mutual information of a vector $\mathbf{y}$ is the *Kullback-Leibler* distance between the densities $p(\mathbf{y})$ and $\prod_j p(y_j)$. This distance is zero if and only if $\mathbf{y}$ has independent components [95].

Assume that the spectral vectors are in a subspace of dimension p and that $\mathbf{W}$ is a nonsingular matrix. Then

$$I(y_1, y_2, \ldots, y_p) = \sum_j H(y_j) - H(\mathbf{r}) - \log|\det \mathbf{W}|, \tag{4.4}$$

where $H(y_j)$, $H(\mathbf{y})$, and $H(\mathbf{r})$ are the entropy of random variable y_j, of random vector $\mathbf{y}$, and of random vector $\mathbf{r}$, respectively (see [12, 29]).

Most ICA algorithms find the separating matrix by minimizing expression (4.4), or an equivalent objective function, with respect to $\mathbf{W}$. The *Negentropy* (see [75, 100]), an entity closely related with the mutual information, which is also used as an objective function to obtain $\mathbf{W}$, is defined as

$$J(\mathbf{y}) \equiv H(\mathbf{y}_{gauss}) - H(\mathbf{y}), \tag{4.5}$$

where $\mathbf{y}_{gauss}$ is a Gaussian random vector with the same mean and covariance as $\mathbf{y}$ [43]. *Negentropy*

is nonnegative and is equal to zero if, and only if, $\mathbf{y}$ has Gaussian distribution. Assuming that components y_j, for $j = 1, \ldots, p$, are uncorrelated, it follows that

$$I(y_1, y_2, \ldots, y_p) = J(\mathbf{y}) - \sum_j J(y_j), \tag{4.6}$$

which means that finding maximum of *Negentropy* directions (*i.e.*, maximizing $\sum_j J(y_j)$ with respect to $\mathbf{W}$), is equivalent to minimize the mutual information.

Well known ICA methods are *FastICA* [75], *Jade* [29], and the Bell and Sejnowski algorithm [12]. *FastICA* is based on a fixed-point procedure and uses the absolute value of kurtosis as a measure of nongaussianity, while *Jade* uses the fourth-order cross-cumulants of the data to separate sources. The Bell and Sejnowski algorithm uses the stochastic gradient ascent learning rule to minimize the mutual information.

IFA [6] was proposed as a method for recovering independent hidden sources from their observed mixtures immersed in additive noise. IFA implements two steps: first, source densities and noise covariance are estimated from the observed data by ML; second, sources are reconstructed by an optimal nonlinear estimator. IFA assumes the observation model

$$\mathbf{r} = \mathbf{M}\mathbf{s} + \mathbf{n}, \tag{4.7}$$

where sources are independent with unknown distributions (at most one is Gaussian) and $\mathbf{n}$ is Gaussian noise with covariance $\mathbf{C}_n$, a matrix not necessarily diagonal. To make the model analytically tractable each source density is modeled by a mixture of Gaussians (MOG) [117]. An expectation-maximization (EM) algorithm [46, 113] is applied to compute the ML estimate of the noise covariance and the Gaussian mixture parameters.

The classical PCA [82] seeks for a linear decomposition that best represents the data in the least-squares sense. PCA finds a linear transformation $\mathbf{y} = \mathbf{W}\mathbf{r}$, where each row vector of $\mathbf{W}$

corresponds to the normalized orthogonal eigenvector of the data covariance matrix. While PCA only uses the second-order statistics, ICA looks for components that are statistically independent rather than uncorrelated; thus, it requires statistics of orders higher than the second [100].

It is known that, owing to physics of the acquisition process, the abundance fractions are dependent ($\alpha_j \in \Delta_p$). Thus the application of ICA and IFA in hyperspectral unmixing yields incorrect results. To shed light into the limitation of ICA/IFA approach, let us compute the ML estimate of $\mathbf{W}$:

$$\mathbf{W}_{\mathrm{ML}} = \underset{([\mathbf{W}]_{i,:})\mathbf{y}\in\Delta_p}{\arg\min} \frac{1}{N}\sum_{i=1}^{N} \log p(\mathbf{y}|\mathbf{W}) \tag{4.8}$$

$$\simeq \underset{([\mathbf{W}]_{i,:})\mathbf{y}\in\Delta_p}{\arg\min} \int p(\mathbf{y}|\mathbf{W}_0) \log \frac{p(\mathbf{y}|\mathbf{W}_0)}{p(\mathbf{y}|\mathbf{W})} \mathrm{d}\mathbf{y}, \tag{4.9}$$

where $\mathbf{W}_0$ stands for the true unmixing matrix and the weak law of large numbers [171] has been used in (4.9). Note that, according to expression (4.9), $\mathbf{W}_0$ is given by the minimization of the *Kullback-Leibler* distance between $p(\mathbf{y}|\mathbf{W}_0)$ and $p(\mathbf{y}|\mathbf{W})$. This is similar to ICA rationale, although with a remarkable difference: the minimization is constrained and $p(\mathbf{y}|\mathbf{W})$ is not independent. There is no hope, therefore, that $\mathbf{W}$ given by (4.9) and (4.8) are equivalent. However, due to degradation mechanisms normally found in hyperspectral data sets, namely, signature variability, abundance constraints, topography modulation, and additive noise, the constraint $([\mathbf{W}]_{i,:})\mathbf{y} \in \Delta_p$ is no longer valid, and a much more complex and difficult model shall be considered.

The application of ICA to hyperspectral data, proposed in several works [11, 24, 36, 38, 50, 86, 88, 89, 103, 133, 165], relies on the fact that those degradations mechanisms might attenuate the source dependence on hyperspectral data. For example, signature variability introduces randomness among sources and thereby attenuates their statistical dependencies. This can be understood by computing the correlation factor between sources $s_i = \psi_i\alpha_i$ and $s_j = \psi_j\alpha_j$. Assuming that ψ_i is

independent from ψ_j and that $\boldsymbol{\psi}$ is independent from $\boldsymbol{\alpha}$, the magnitude of the correlation factor between s_i and s_j is given by

$$\begin{aligned} |\varrho_{s_i s_j}| &= \frac{\left|\mathrm{E}[\psi_i]\mathrm{E}[\psi_j]\right| \left|C_{\alpha_i\alpha_j}\right|}{\sqrt{\mathrm{E}[\psi_i^2]\mathrm{E}[\alpha_i^2] - \mathrm{E}^2[\psi_i]\mathrm{E}^2[\alpha_i]}\sqrt{\mathrm{E}[\psi_j^2]\mathrm{E}[\alpha_j^2] - \mathrm{E}^2[\psi_j]\mathrm{E}^2[\alpha_j]}} \\ &\leq \frac{\left|\mathrm{E}[\psi_i]\mathrm{E}[\psi_j]\right|}{\sqrt{\mathrm{E}[\psi_i^2]\mathrm{E}[\psi_j^2]}} \frac{\left|C_{\alpha_i\alpha_j}\right|}{\sigma_{\alpha_i}\sigma_{\alpha_j}} \qquad (4.10) \\ &\leq |\varrho_{\alpha_i\alpha_j}|, \qquad (4.11) \end{aligned}$$

where $\mathrm{E}[\psi_k^2] \geq \mathrm{E}^2[\psi_k]$ was invoked to obtain the right-hand side of expressions (4.10) and (4.11). We conclude then that signature variability does not increase source correlation. Of course, decorrelation does not imply independence. It is, however, plausible that increasing decorrelation means increasing independence. In the next section we give experimental evidence that hyperspectral sources are not independent even in presence of strong degradations mechanisms.

4.3 ICA and IFA Evaluation with Simulated Data

In this section ICA (*FastICA*, *Jade*, and Bell and Sejnowski) and the IFA algorithms are applied to simulated data. The study consider that hyperspectral observations are described by the generative model presented in (1.8), here reproduced,

$$\begin{aligned} \mathbf{r} &= \mathbf{M}\underbrace{\gamma\boldsymbol{\psi}\boldsymbol{\alpha}}_{\mathbf{s}} + \underbrace{\gamma\sum_{j=1}^{p}\alpha_j\boldsymbol{\eta}_j + \mathbf{n}'}_{\mathbf{n}} \\ &= \mathbf{M}\mathbf{s} + \mathbf{n}. \qquad (4.12) \end{aligned}$$

This model takes into account the degradation mechanisms normally found in hyperspectral applications, namely, signature variability [69, 152, 166], abundance constraints, topography modulation, and noise. Four experiments are conducted: the first models a canonical ICA scenario, where the

abundance fractions are independent; the second models an ideal hyperspectral scenario, where only constraints (1.4) are enforced, meaning that abundance fractions are dependent; the third models a real hyperspectral scenario, where abundance fractions are generated according to (4.12), thus modeling abundance fraction dependence, signature variability, topography modulation, and noise; the fourth model sources according to (4.12) and the mean magnitude of the cross-correlation factor between sources and their estimates by *FastICA* algorithm is evaluated as a function of the SNR, of the signature variability, and of the number of sources. The cross-correlation factor is adopted as a performance measure because ICA and IFA unmixes abundance fractions up to a constant factor.

In all experiments, the scene dimension is of 30×30 pixels and endmember signatures were extracted from a hyperspectral subimage of the Indian Pine Test Site in northwestern Indiana, acquired by an AVIRIS instrument in June 1992[1]. Noisy channels and water absorption channels were removed (channels $1-4$, $107-113$, and $150-166$). Concerning atmospheric correction, this image has been processed in order to remove the path radiance and the light scattered by the interaction between the surface and the atmosphere.

Table 4.1 presents the name of the substances extracted and the angle between pairs of signatures, *i.e.*, $\phi_{ij} = \arccos\left(\frac{(\mathbf{m}_i)^T \mathbf{m}_j}{\|\mathbf{m}_i\| \, \|\mathbf{m}_j\|}\right)$. The lowest and the highest angles are, approximately, 4° (between second and seventh signatures) and 48° (between fourth and ninth signatures), respectively. Endmember 9 has the highest angle with respect to the closest endmember. In the first three experiments, three endmembers were selected (see Fig. 4.1); in the fourth experiment, the number of endmembers is varied from three to ten.

[1]Available at http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/

Table 4.1: Substances extracted from the data set and angle between pairs of radiance spectrum (in degrees).

substance	number	1	2	3	4	5	6	7	8	9	10
Corn-notill	1	0									
Grass/Trees	2	31	0								
Oats	3	19	12	0							
Grass/Pasture	4	36	6	17	0						
Hay-windrowed	5	18	17	7	22	0					
Wheat	6	27	5	8	10	13	0				
Bldg-Grass-Tree-drives	7	34	4	14	4	19	7	0			
Stone-steel towers	8	5	35	23	40	22	31	38	0		
House roof	9	15	43	32	48	31	39	45	11	0	
Steel	10	6	29	18	35	16	26	32	9	17	0

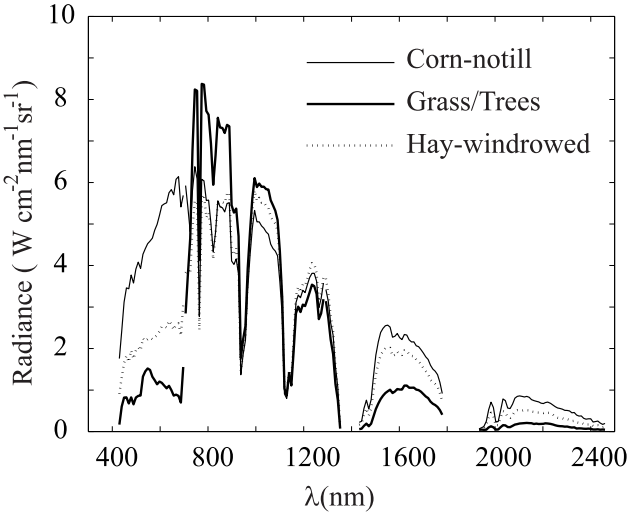


Figure 4.1: Radiance spectrum of the first second and fifth substances used in experiments I, II, and III.

Experiment I

In this experiment, the abundance fractions are mutually independent, each one following a Beta distribution given by expression (3.9). Mean values are set to $\mu_1 = 0.4$, $\mu_2 = 0.3$, and $\mu_3 = 0.3$. In the remaining experiments, abundance fractions follow a joint Dirichlet distribution given by expression (2.2), which constrains the abundance fractions to be in the simplex Δ_p. On the other hand, as noted in [116], the Dirichlet density is adjusted to model proportions.

Table 4.2 presents the sample mean of the cross-correlation coefficients between the abundance fractions and their estimates and the sample cross-correlation coefficients between the endmember signatures and their estimates. These coefficients are computed from 256 Monte Carlo runs. It can be seen that under this condition (independent abundance fractions) IFA, *FastICA*, and *Jade* algorithms work very well, whereas the Bell and Sejnowski algorithm performs a little worse.

Experiment II

In this experiment, abundance fractions are dependent, following a Dirichlet distribution with parameters $\theta_1 = 0.4$, $\theta_2 = 0.3$, and $\theta_3 = 0.3$. It is clear that none of the algorithms correctly unmixes the original dependent data. IFA and *FastICA* algorithms only estimate two abundance fractions because they implement a pre-processing step to whiten the observed data and to reduce the dimension (note that the data is in the simplex Δ_p due to constraint $\sum_j \alpha_j = 1$).

Experiment III

In this experiment, abundance fractions are dependent following a Dirichlet distribution with parameters $\theta_1 = 0.4$, $\theta_2 = 0.3$, and $\theta_3 = 0.3$ for each endmember. Scale ψ_i, controlling signature variability, is uniformly distributed with in the interval $[0.9\ ,\ 1.1]$; parameter γ is Beta distributed with parameters $\beta_1 = 1$, $\beta_2 = 0.8$. Noise in expression (4.12) is zero-mean white Gaussian with variance σ^2 in each band such that the SNR [defined in expression (2.4)] is set to 30 dB. As in

Table 4.2: Sample cross-correlation between abundance fractions and the correspondent estimates ($\overline{\varrho_\alpha}$) and sample cross-correlation between endmember signatures and the correspondent estimates ($\overline{\varrho_m}$). Results based on 256 Monte Carlo runs.

Algorithm		IFA		*FastICA*		*Jade*		Bell Sejnowski	
	source	$\overline{\varrho_\alpha}$	$\overline{\varrho_m}$	$\overline{\varrho_\alpha}$	$\overline{\varrho_m}$	$\overline{\varrho_\alpha}$	$\overline{\varrho_m}$	$\overline{\varrho_\alpha}$	$\overline{\varrho_m}$
Experiment I	1	0.9767	0.9725	0.9978	0.9971	0.9994	0.9997	0.7817	0.7119
	2	0.9926	0.9777	0.9970	0.9921	0.9985	0.9952	0.8553	0.7071
	3	0.9797	0.9765	0.9895	0.9921	0.9911	0.9954	0.5738	0.5803
Experiment II	1	-	-	-	-	0.1978	0.1623	0.3332	0.8836
	2	0.8310	0.9259	0.7902	0.7123	0.4778	0.5794	0.8352	0.9812
	3	0.4913	0.9144	0.3281	0.7366	0.2482	0.5529	0.4490	0.9815
Experiment III	1	0.9755	0.4534	0.8780	0.6432	0.9304	0.6965	0.9727	0.7843
	2	0.9670	0.8619	0.9641	0.7631	0.9752	0.7668	0.9692	0.7830
	3	0.9065	0.8618	0.8454	0.7729	0.8981	0.7645	0.8778	0.6238

the previous experiment, ICA and IFA algorithms do not correctly unmix the three abundance fractions. IFA, however, yields the best results being able to approximately unmix two abundance fractions.

Experiment IV

In this experiment, the cross-correlation factor between abundance fractions and their estimates is computed as a function of the SNR, the number of endmembers, and the signature variability. Two abundance fraction distributions are considered: i) symmetric Dirichlet distributions ($\theta_k = 1/p$, $k = 1, \ldots, p$); ii) asymmetric Dirichlet distributions $\big[\theta_k = 5/(p+8)$, $k = 1, 2$; $\theta_k = 1/(p+8)$, $k = 3, \ldots, p\big]$. Signature variability is controlled by the distribution of the scale random vector ψ_i. In this experiment it is assumed that ψ_i is uniformly distributed within the interval $[\eta, 1]$, where $\eta \geq 0$. Thus $\eta = 1$ means absence of variability, whereas $\eta = 0$ means maximum variability.

Although IFA was conceived to recover independent sources from linear mixtures immersed in noise, this algorithm is not considered in this experiment, because IFA computational complexity increases exponentially with the number of endmembers. In this experiment only the *FastICA* algorithm is applied.

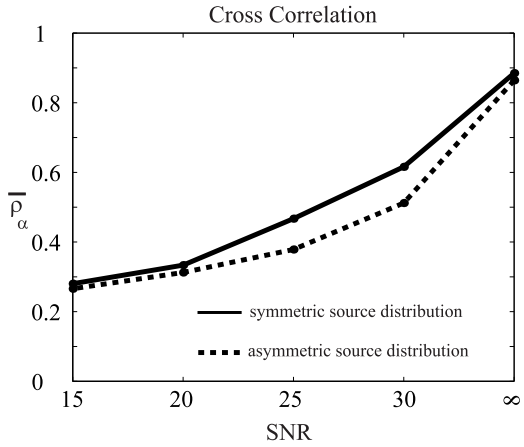


Figure 4.2: Mean magnitude of the cross-correlation factors between abundance fractions and their estimates $\overline{\varrho_\alpha}$ as a function of the SNR, for $p = 10$ and $\eta = 0.8$ (results are based on 256 Monte Carlo runs).

Fig's. 4.2, 4.3, and 4.4 present the sample mean cross-correlation factors, $\overline{\varrho_\alpha}$, between each abundance fraction and the correspondent estimate based on 256 Monte Carlo runs.

Fig. 4.2 shows an increasing $\overline{\varrho_\alpha}$ as a function of the SNR $= \{15, 20, 25, 30, \infty\}$ dB, for $p = 10$ and $\eta = 0.8$. Asymmetry of abundance fraction distributions affects little the unmixing results. Fig. 4.3(a) shows $\overline{\varrho_\alpha}$ as a function of the number of endmembers, for $p = 3, \ldots, 10$, SNR $= \{30, \infty\}$ dB, and $\eta = 0.6$. As the number of endmembers increases, the statistical dependence among sources decreases and a better performance of ICA algorithms is expected. This trend can be observed in Fig. 4.3(a) at least for high SNR. On Fig. 4.3(b) the sample mean of the magnitude of the cross-correlation factor is shown for each source separately, with SNR $= 30$ dB, $\eta = 0.6$, and symmetric source distribution. In this figure, it can be seen that there is always endmembers correctly unmixed and others incorrectly unmixed, regardless of the number of endmembers. Fig. 4.4(a) shows $\overline{\varrho_\alpha}$ as a function of signature variability, for SNR $= \{30, \infty\}$ dB, and $p = 10$. The unmixing performance is quasi-constant for $0.1 < \eta < 0.6$ and takes higher values in noiseless scenes. As η approaches to 1, meaning smaller signature variability and higher statistical dependence among sources, ICA performance decays as expected. Fig. 4.4(b) presents the $\overline{\varrho_\alpha}$ for each source separately, with SNR $= 30$ dB, $p = 10$, and asymmetric source distribution. Second, fifth,

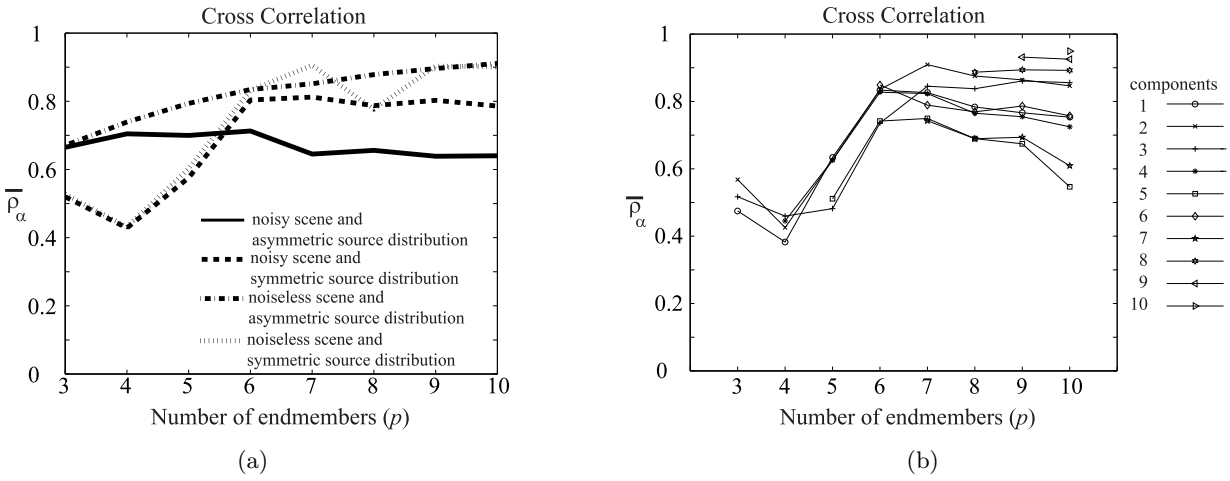


Figure 4.3: Mean magnitude of the cross-correlation factors between abundance fractions and their estimates (results are based on 256 Monte Carlo runs): (a) $\overline{\varrho_\alpha}$ as a function of the number of endmembers p in the scene for $\eta = 0.6$; (b) $\overline{\varrho_\alpha}$ as a function of p, for each individual source ($\eta = 0.6$, symmetric sources, SNR $= 30\,\mathrm{dB}$).

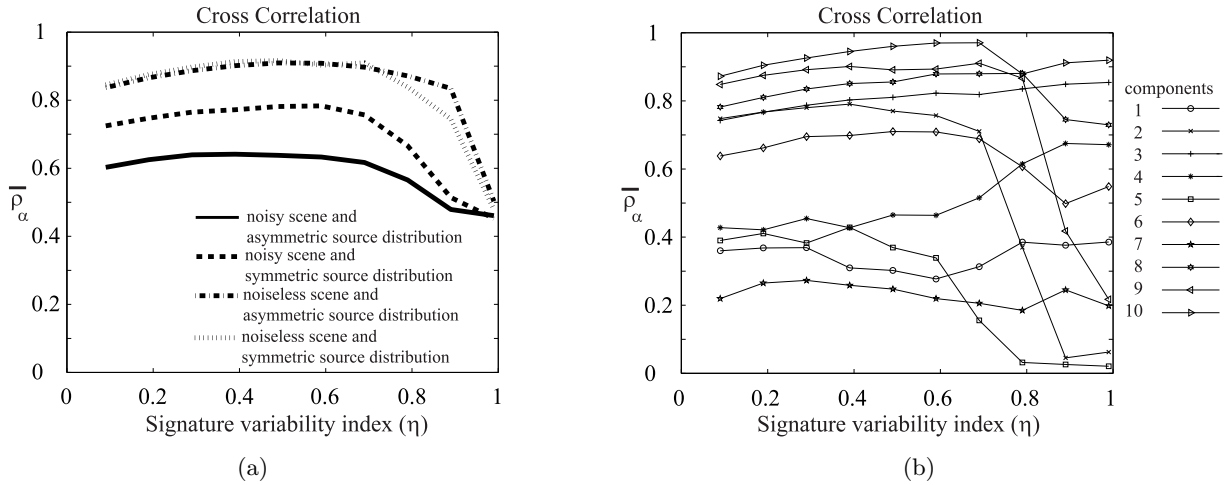


Figure 4.4: Mean magnitude of the cross-correlation factors between abundance fractions and their estimates (results are based on 256 Monte Carlo runs): (a) $\overline{\varrho_\alpha}$ as a function of parameter η, for $p = 10$; (b) $\overline{\varrho_\alpha}$ as a function of parameter η, for all sources ($p = 10$, asymmetric sources, SNR $= 30\,\mathrm{dB}$).

and seventh endmembers are clearly incorrectly unmixed. Note that with respect to table 4.1, endmember 7, which jointly with endmember 2 form the closest pair, shows the worst unmixing result, whereas endmember 9, which has the highest angle with respect to the closest endmember, shows a good unmixing result.

The pattern of behavior exhibited in experiments I to IV was systematically replicated regardless of the source statistics. The conclusions are the following:

1. the accuracy of ICA applied to hyperspectral data tends to increase with the signature variability, the number of sources, and the SNR;

2. there are, however, always endmembers incorrectly unmixed, regardless of the unmixing scenario. [124].

In the next sections the estimation of the unmixing matrix is considered. This study is based on the minimization of the mutual information, which gives some evidence on the reasoning underlying ICA and IFA limitations in unmixing hyperspectral data.

4.4 Limitations of ICA and IFA in Unmixing Hyperspectral Data

In this section, ICA and IFA limitations in unmixing hyperspectral data are studied. To this end, the behavior of the mutual information in the neighborhood of the true unmixing matrix is characterized. The aim is to show that when sources are dependent, minimizing mutual information might lead to a unmixing matrix far from the true unmixing matrix.

A brief introduction to ICA and IFA was presented in Section 4.2. Given an unknown linear mixture of the unobserved independent sources $\mathbf{r} = \mathbf{M}\mathbf{s}$, ICA and IFA look for a matrix $\mathbf{W}$ that maximizes a nongaussianity measure of the vector $\mathbf{y} = \mathbf{W}\mathbf{r}$. An objective function commonly used in ICA algorithms is the mutual information [41].

Assuming that the det($\mathbf{W}$) is constant, the minimization of the mutual information reduces to finding [see expression (4.3)]

$$\widehat{\mathbf{W}} = \arg\min_{\mathbf{W}} I(y_1, y_2, \ldots, y_p, \mathbf{W}) = \arg\min_{\mathbf{W}} \sum_j H(y_j), \tag{4.13}$$

where

$$H(y_j) \equiv -\int_{-\infty}^{+\infty} p_{Y_j}(u) \log p_{Y_j}(u)\ du, \tag{4.14}$$

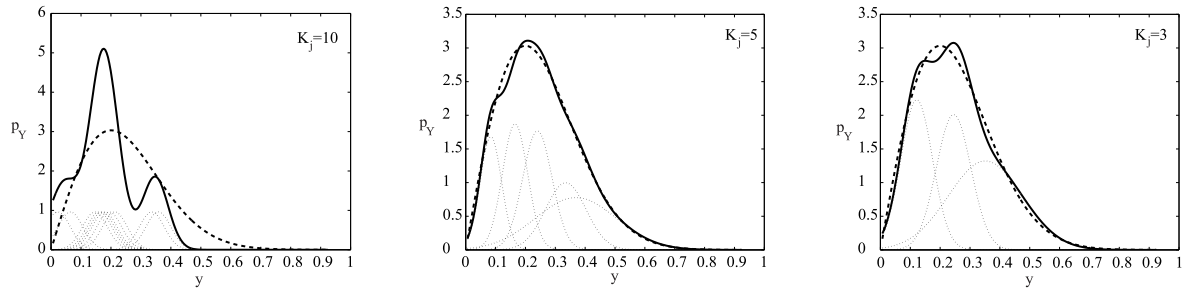
and p_{Y_j} is the probability density function of y_j. To compute expression (4.14), p_{Y_j}, for $j = 1, \ldots, p$, needs to be estimated. Herein this estimate is obtained by fitting a MOG to samples of y_j [114]:

$$p_{Y_j}(y_j) = \sum_{q_j=1}^{K_j} \epsilon_{q_j} \left(2\pi \det(\Sigma_{q_j})\right)^{-1/2} \exp\left[-\frac{1}{2}(y_j - \mu_{q_j})^T \Sigma_{q_j}^{-1} (y_j - \mu_{q_j})\right], \tag{4.15}$$

where K_j is the number of Gaussian modes to fit the jth source and ϵ_{q_j}, μ_{q_j}, and Σ_{q_j} are the weight, the mean, and the covariance of the q_jth Gaussian mode, respectively. The number of Gaussian modes, and respective parameters (means, covariances and weights) are obtained via the minimum description length (MDL) based EM algorithm [49]. The entropy (4.14) is computed via numerical integration.

Fig. 4.5 shows an example of a Rayleigh probability density function fitted with a Gaussian mixture using the MDL-EM algorithm [49]. Fig. 4.5(a),(b), and (c) presents the probability density functions obtained with, respectively, the initial parameters, the solution, and the solution constrained to three Gaussian modes.

The maxima of $I(y_1, y_2, \ldots, y_p, \mathbf{W})$ with respect to $\mathbf{W}$ is not very sensitive to the shape of p_{Y_j} (see Chap. 9, [74]). For example [6] uses only three Gaussian modes to fit any density shape.



(a) (b) (c)

Figure 4.5: Rayleigh probability density function (dashed line) fitted with a Gaussian mixture (solid line) using the MDL-EM algorithm [49]: (a) initial parameters (ten Gaussian modes); (b) solution with five Gaussian modes; (c) solution constrained to three Gaussian modes.

Herein, however, all modes given by the MDL-EM algorithm [49] are used, because the interest is, not only in the unmixing matrix $\mathbf{W}$, but also in computing the mutual information $I(y_1, y_2, \ldots, y_p, \mathbf{W})$ as a function of $\mathbf{W}$.

4.4.1 Experimental Results

In the next experiments, the behavior of the mutual information $I(y_1, y_2, \ldots, y_p, \mathbf{W})$ is studied for $\mathbf{W}$ in the neighborhood of the true unmixing matrix $\mathbf{W} = \mathbf{M}^{-1}$. In all experiments the constraint $\det(\mathbf{W}) = 1$ is assumed. This setting does not constraint the unmixed results, as they are defined up to a constant.

Experiment I

The first experiment considers independent abundance fractions with uniform distribution to test the setup under canonical mixing conditions. This experiment assumes $p = 3$ (number of endmembers), $L = 3$ (number of bands), $\mathbf{M} = \mathbf{I}_p$, $\boldsymbol{\psi} = \mathbf{I}_p$, $\gamma = 1$, and $\mathbf{n} = \mathbf{0}_L$.

Fig. 4.6(a) shows the mutual information as a function of ϕ_1 and ϕ_2 in a gray scale (ϕ_1 and ϕ_2 define a rotation in $\mathbb{R}^3$). This angles are termed as azimuth and elevation angles, respectively. The minimum is global and occurs for $\phi_1 = 0$ and $\phi_2 = 0$, *i.e.*, $\mathbf{W} = \mathbf{I}_p$.

As mentioned above, abundance fractions in hyperspectral data are not independent. In order to test ICA with this constraint, in the next experiment, the abundance fractions are generated according to a Dirichlet distribution [see expression 2.2] parameterized with $\mu_1 = 1/3$, $\mu_2 = 1/3$, and $\mu_3 = 1/3$ (recall that μ_j is the expected value of the jth abundance fraction).

Experiment II

In this experiment the following settings are considered: $p = 3$, $L = 3$, $\mathbf{M} = \mathbf{I}_p$, $\boldsymbol{\psi} = \mathbf{I}_p$, $\gamma = 1$, $\mathbf{n} = \mathbf{0}_L$.

In Fig. 4.6(b) the mutual information as a function of angles ϕ_1 and ϕ_2 is presented. No ICA algorithm could ever correctly unmix the original dependent data, since ($\widehat{\phi}_1 \simeq -\pi/4$, $\widehat{\phi}_2 \simeq -\pi/5$), far from the true unmixing matrix, *i.e.*, ($\phi_1 = 0$, $\phi_2 = 0$).

Experiment III

In this experiment the abundance fractions are dependent and Dirichlet distributed ($\mu_1 = 1/3$, $\mu_2 = 1/3$, $\mu_3 = 1/3$). The remaining parameters are $p = 3$, $L = 3$, $\mathbf{M} = \mathbf{I}_p$, $\mathbf{n} = \mathbf{0}_L$, ψ_i uniformly distributed in the interval [0.9 , 1.1], and γ Beta distributed with parameters $\beta_1 = 1$ and $\beta_2 = 0.8$.

Fig. 4.6(c) presents the mutual information as a function of angles ϕ_1 and ϕ_2. Although there is a local minimum at $\phi_1 = 0$ and $\phi_2 = 0$, the absolute minimum occurs at $\widehat{\phi}_1 \simeq -\pi/4$ and $\widehat{\phi}_2 \simeq -\pi/5$.

Experiment IV

This experiment is similar to experiment III, but now the zero-mean white Gaussian noise is added. Herein, three scenarios are considered: SNR = {15, 20, 25} dB.

Fig. 4.6 [(d), (e), and (f)] shows the obtained mutual information for SNR = {25, 20, 15} dB, respectively. The surface exhibits a random pattern as the noise level increases. The global minimum occurs at $\widehat{\phi}_1 \simeq -\pi/5$, $\widehat{\phi}_2 \simeq -\pi/4$ for SNR = 25 dB (see Fig. 4.6(d)), $\widehat{\phi}_1 \simeq 10\pi/36$, $\widehat{\phi}_2 \simeq -10\pi/36$ for SNR = 20 dB (see Fig. 4.6(e)), and $\widehat{\phi}_1 \simeq 10\pi/36$, $\widehat{\phi}_2 \simeq -2\pi/9$ for SNR = 15 dB (see Fig. 4.6(f)).

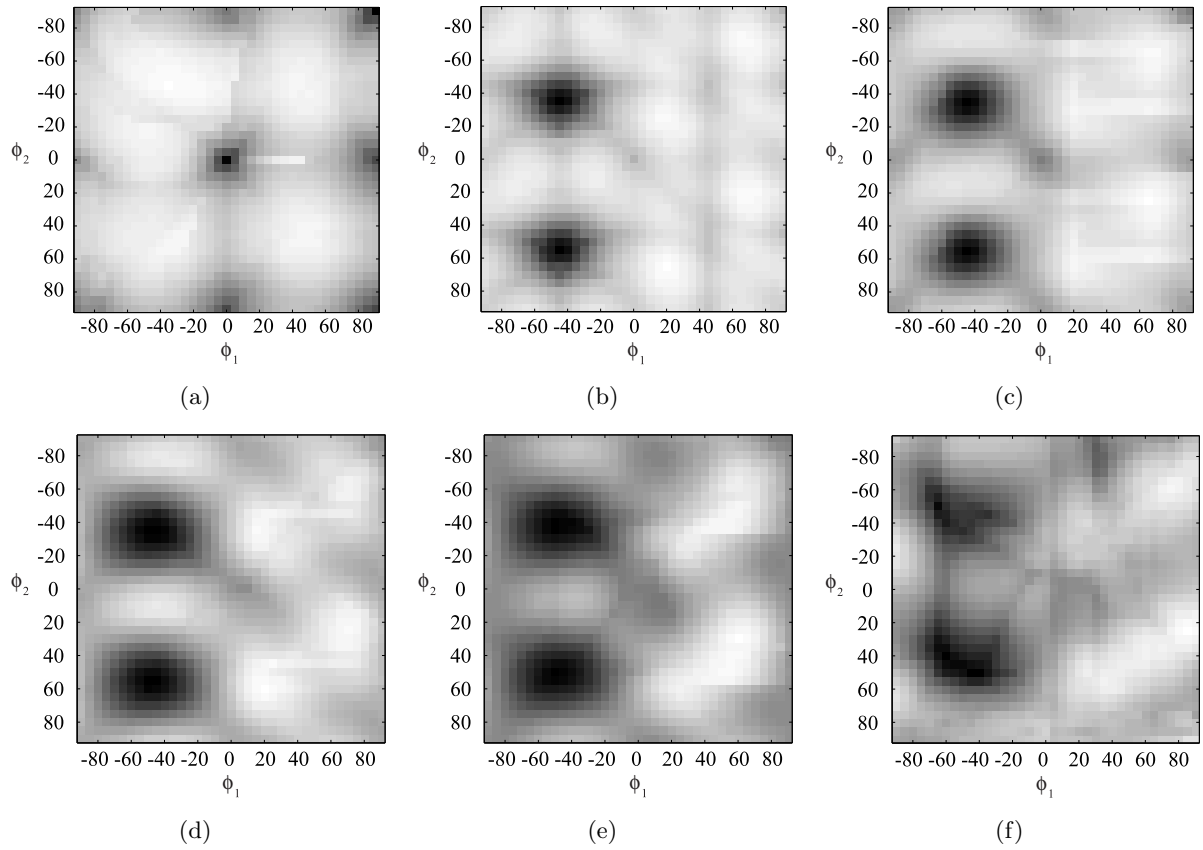
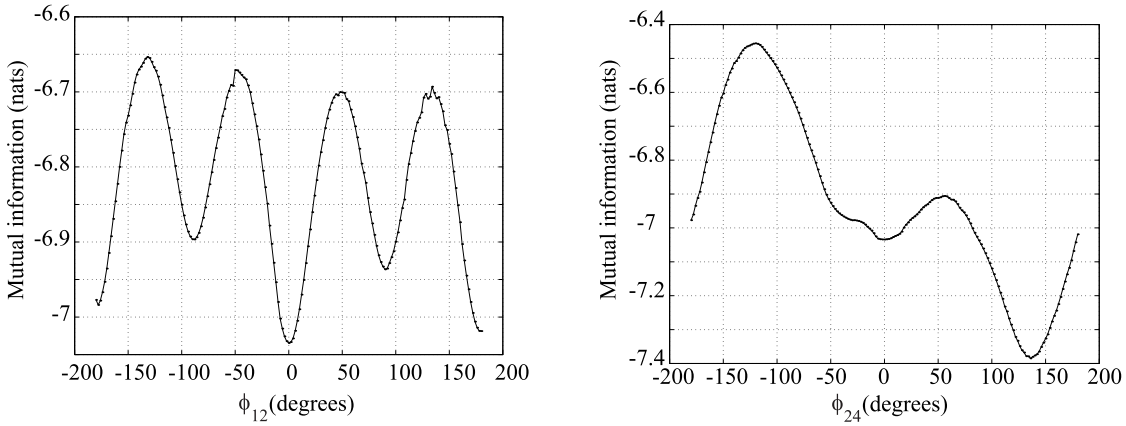


Figure 4.6: Mutual information as a function of parameters ϕ_1 (azimuth angle) and ϕ_2 (elevation angle): (a) three independent sources; (b) dependent sources; (c) dependent sources with illumination perturbations and signature variability; (d) noisy scene (SNR = 25 dB); (e) noisy scene (SNR = 20 dB); (f) noisy scene (SNR = 15 dB).

All these global minima are far from ($\phi_1 = 0$, $\phi_2 = 0$). Note, however, that the presence of a local minimum at $\widehat{\phi}_2 \simeq 0$ and $\widehat{\phi}_2 \simeq 0$.

Experiment V

This experiment considers ten endmembers with asymmetric Dirichlet distributions ($\mu_1 = 5/18$, $\mu_2 = 5/18$, $\mu_k = 1/18$; $k = 3, \ldots, 10$). The remaining parameters are $\mathbf{M} = \mathbf{I}_p$, $\mathbf{n} = \mathbf{0}_L$, ψ_i uniformly distributed in the interval [0.9 , 1.1], and γ Beta distributed with parameters $\beta_1 = 1$ and $\beta_2 = 0.8$. The mixture is immersed in zero-mean white Gaussian noise, corresponding to SNR = 20 dB.



(a) (b)

Figure 4.7: Ten dependent components are mixed with noise added: (a) Mutual information (up to a constant) as a function of rotation angle ϕ_{12} angle between first and second components; (b) Mutual information (up to a constant) as a function of rotation angle ϕ_{24} angle between second and fourth components.

Fig. 4.7(a) shows the mutual information (up to a constant) as a function of the rotation matrix

$$\mathbf{W} = \begin{bmatrix} \cos\phi_{12} & \sin\phi_{12} & & & \mathbf{0} \\ -\sin\phi_{12} & \cos\phi_{12} & & & \\ & & 1 & & \\ & & & \ddots & \\ \mathbf{0} & & & & 1 \end{bmatrix},$$

where ϕ_{12} stands for the angle between first and second components. Note that the minimum occurs at $\widehat{\phi}_{12} \simeq 0$, which means that first and second component could be correctly unmixed; Fig. 4.7(b) shows the mutual information as a function of ϕ_{24}. It can be observed a local minima at the origin but the global minimum is at $\widehat{\phi}_{24} \simeq -\pi/4$. The conclusion is that, under these conditions, although might exist local minima corresponding to the true unmixing matrix, the global minimum of the mutual information might be very far from true one.

The pattern of behavior described in experiments II, III, IV, and V was systematically observed in a series of experiments with different abundance fraction distributions. The conclusion is that in linear hyperspectral data unmixing, the unmixing matrix $\mathbf{W}$ minimizing the mutual information might be very far from the true one, at least for a few number of endmembers. This is in agreement with conclusions drawn in Section 4.3.

4.5 FastICA Algorithm Applied to Real Hyperspectral Data

In this section, a subimage of the hyperspectral data set from the Indian Pine Test Site in north-western Indiana acquired by an AVIRIS in June 1992 is considered. The data set is composed of 220 spectral channels with 10nm bandwidth acquired in the $0.4 - 2.5\mu$m region. It contains 145×145 pixels ($21,025$ pixels) with a ground pixel resolution of 17 meters [70]. This region contains a mixture of agriculture and forestry. However, due to the early season date of data collection, the cultivated land appears to have very little canopy cover. There is a major dual lane highway (U.S. 52 and U.S. 231), a rail line crossing near the top, a major secondary road (Jackson Highway) near the middle, several other county roads, and houses (Fig. 4.8(a) shows band 29 of the data set). The ground truth of the region [96] classifies the ground covered area into sixteen classes and ignores many small variations within fields that can be seen in the image data [see Fig. 4.8(a)]. With respect to atmospheric correction, this data set has been processed to remove path radiance and the light scattered by interaction between surface and the atmosphere. The corrected image is still, however, in radiance units.

A PCA pre-processing step was implemented to whiten the observed data and to reduce the dimension: every pixel vector $\mathbf{r}$ is transformed into a vector of dimension 16 with zero-mean and identity covariance matrix: $\mathbf{x} = \mathbf{U}_{16}\big(\mathbf{r} - \mathrm{E}[\mathbf{r}]\big)$, such that $\mathrm{E}[\mathbf{x}] = 0$ and $\mathrm{E}[\mathbf{x}\mathbf{x}^T] = \mathbf{I}$, where $\mathbf{I}$ the identity matrix.

Fig. 4.8 presents the first sixteen components extracted by the *FastICA* algorithm. The first, second, fourth, tenth, and eleventh components correspond to houses or man made materials that exist in a few locations in the scene [see Fig's. 4.8(c), 4.8(e), 4.8(k), and 4.8(l)]. Note that the maximum contrast of sources occurs in house locations, stone-steel towers, Jackson highway, and rail line. The third and fifth components are a mixture of grass with pasture and trees, respectively [see Fig's. 4.8(d) and 4.8(f)]. The sixth component [see Fig. 4.8(g)] represents a mixture of grass, soybeans, and corn. Apparently, the seventh component is hay-windrowed [see Fig. 4.8(h)]. The ninth component represents vegetation mowed [see Fig. 4.8(j)]. Components presented in Fig's. 4.8(b), 4.8(i), 4.8(m), and 4.8(n) do not represent any class of the available ground truth [96]. Fig's. 4.8(o), 4.8(p), and 4.8(q) are mainly noise. The conclusion is that only six sources are unmixed and the other ten are a mixture of several materials present in the scene or are manly noise. This is in accordance with the main findings based on simulated data.

A pertinent question is, What sources are correctly unmixed? To address this question, the entropy of each component is computed, normalized to unit variance, following the procedure described in Section 4.4. The underlying idea is that a mixed source tends to exhibit higher entropy, the maximum value being $H_{max} = 1.419$ nats achieved by a Gaussian source. Table 4.3 shows the entropy for each component shown in Fig. 4.8 b)-q). By inspection of Table 4.3, two subsets can be identified: the first, with smaller values of entropy, corresponds to components correctly unmixed [see Fig's. 4.8 c)-h)]; the second, with larger values of entropy, corresponds to components with mixed sources [see Fig's. 4.8 b) and i)-q)]. The larger values, near H_{max}, correspond to the last three components which are mostly noise.

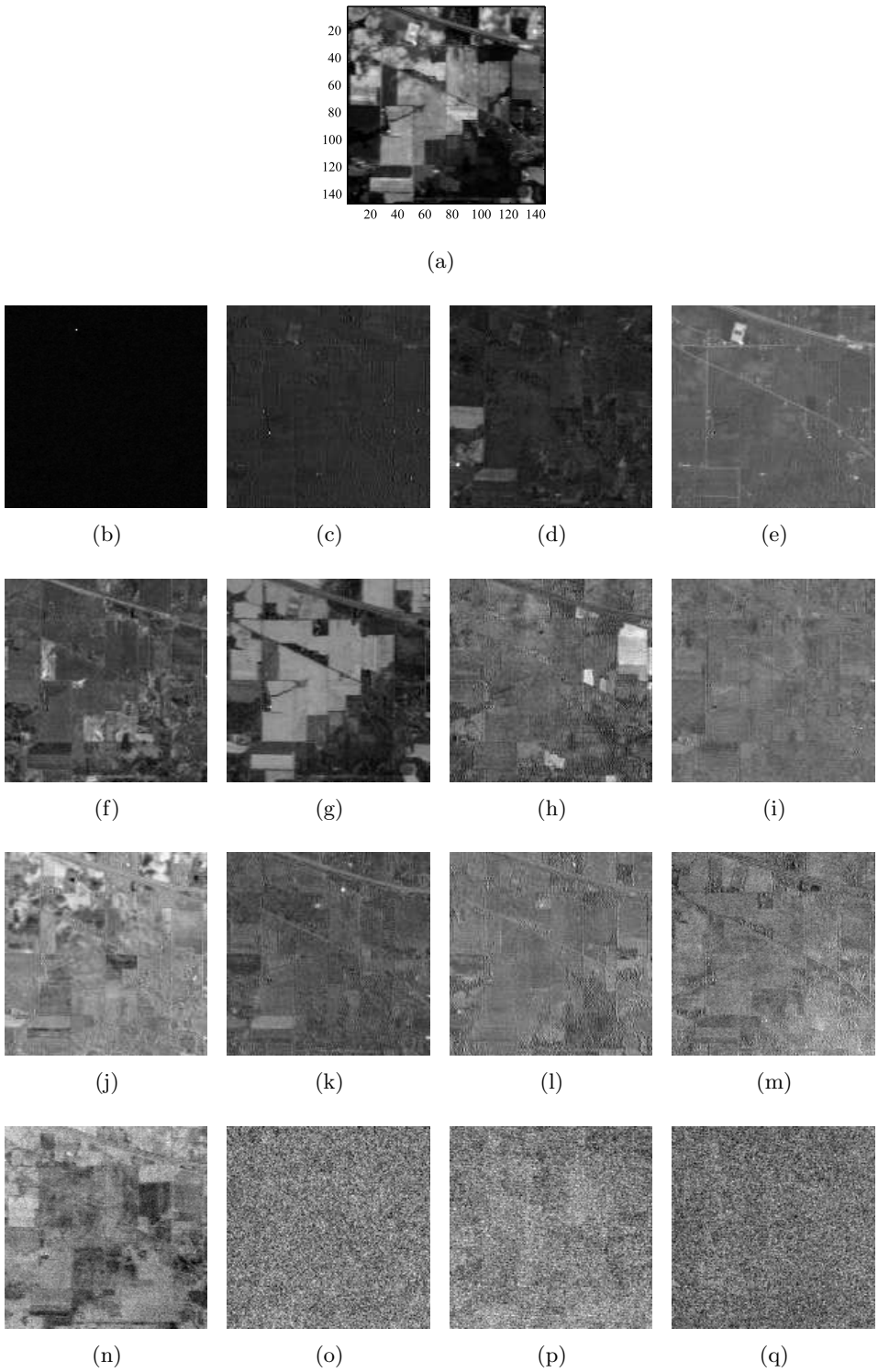


Figure 4.8: (a) Subimage of Indiana Pines test site, band 29 ($\lambda = 667.3nm$) and independent components extracted from data set, with *FastICA* algorithm: (b) not classified; (c)houses; (d) grass/pasture; (e) stone steel towers, rail line, and highways; (f) grass/trees; (g) grass, soybeans, and corn; (h) hay-windrowed; (i) not classified; (j) vegetation mowed; (k) houses; (l) houses; (m) not classified; (n) not classified; (o) noise; (p) noise; (q) noise.

Table 4.3: Entropy of each component extracted with *FastICA* algorithm.

Component	b)	c)	d)	e)	f)	g)	h)	i)
Entropy	1.4020	1.3290	1.1834	1.2002	1.3399	1.1300	1.2811	1.4031
Unmixed		$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	
Component	j)	k)	l)	m)	n)	o)	p)	q)
Entropy	1.3781	1.3945	1.3842	1.4069	1.4172	1.4188	1.4187	1.4189
Unmixed								

4.6 Conclusions

Blind hyperspectral linear unmixing aims at estimating the number of reference substances (also called endmembers), their spectral signatures, and their fractions at each pixel (called abundance fractions), using only the observed data (mixed pixels). Geometric approaches have been used whenever pure pixels are present in data [19, 44, 120, 164, 173]. In most cases, however, pure pixels cannot be found in the data. In such cases, unmixing procedures become a difficult task.

In the recent past, ICA has been proposed as a tool to unmix hyperspectral data [11, 24, 36, 38, 86, 88, 103, 133, 165]. ICA consists in finding a linear decomposition of data into statistically independent components. IFA extends ICA concepts when noise is present. Crucial assumptions of ICA and IFA are that abundance fractions are independent and that each pixel is a linear mixture of endmember signatures weighted by the correspondent abundance fractions. Concerning hyperspectral data, the second assumption is valid whenever the multiple scattering among the distinct endmembers is negligible and the surface is partitioned according to the abundance fractions. The first assumption, however, is not valid due to physical constraints on the acquisition process.

This chapter addresses the impact of the abundance fraction (sources) dependence on unmixing hyperspectral data with ICA/IFA. The study considers simulated and real hyperspectral data, where hyperspectral observations are described by the generative model (1.8) that includes degradation mechanism such as signature variability, abundance constraints, topography modulation, and noise.

IFA and three well known ICA algorithms were tested on simulated data. The main findings in the Chapter are the following:

i) ICA/IFA performance increases with the SNR;

ii) ICA/IFA performance tends to increase with the signature variability and/or with the number of endmembers. The underlying reason is that by increasing the signature variability and/or the number of endmembers the statistical dependence among endmembers is attenuated;

iii) There are always endmembers incorrectly unmixed, regardless the unmixing scenario.

In order to assess the impact of hyperspectral abundance fraction dependence on the ICA/IFA algorithms, the behavior of the mutual information of the unmixed sources is studied in the neighborhood of the true unmixed data. The conclusion is that in hyperspectral linear unmixing, the unmixing matrix minimizing the mutual information might be very far from the true one, at least for a few number of endmembers.

Finally ICA and IFA algorithms are tested in a subimage of the hyperspectral data set from the Indian Pine test site in northwestern Indiana acquired by an AVIRIS in June 1992. According to the available ground truth of the region, only 6 sources are correctly unmixed and 10 are incorrectly unmixed. This is in line with the conclusion drawn from simulated data. A method based on the source entropy to sort the output of ICA or IFA algorithms according to the likelihood of being correctly separated was proposed.

Chapter 5

Dependent Component Analysis

5.1 Introduction

The previous chapter addressed the impact of the source dependence on unmixing hyperspectral data with ICA/IFA concepts. It was therein shown that the most well known blind source separation algorithms do not correctly unmix hyperspectral data. It was also concluded that the unmixing matrix minimizing the mutual information might be very far from the true one.

This chapter presents a new direction to blindly unmix hyperspectral data, termed *dependent component analysis* (DECA), where abundance fractions are modeled by a mixture of Dirichlet densities, thus enforcing source nonnegativity and additivity constraints. DECA is in the vein of works [6, 117] replacing independent sources represented by MOGs with mixtures of Dirichlet (MODs) sources. Compared with the geometric-based approaches, the advantage of DECA is that there is no need to have pure pixels in the observations.

This chapter is organized as follows. Section 5.2 presents the algorithm to estimate the Dirichlet parameters of a MOD. Section 5.3 describes the fundamentals of the proposed method (DECA). Section 5.4 and Section 5.5 illustrate aspects of the performance of DECA approach with experimental data and real data, respectively. Section 5.6 concludes the chapter with some remarks. Parts of the approach presented in this chapter were published in [17, 128].

5.2 Learning Dirichlet Sources

This section describes the algorithm introduced in [47] to compute the ML estimates of the parameters of a Dirichlet density and of a MOD densities. An approach based on the majorization maximization (MM) [131] perspective leading to a similar algorithm can bee found in [116].

5.2.1 Dirichlet Parameters Estimation

Let $\mathbf{S} = [S_1, \ldots, S_p]^T$ be a p-dimensional random variable, with $\mathbf{s} = [s_1, \ldots, s_p]^T$ representing one particular outcome $\mathbf{S}$. Lets assume that $\mathbf{S}$ follows the Dirichlet distribution

$$D(\mathbf{s}|\boldsymbol{\theta}) = \frac{\Gamma(\sum_{j=1}^{p} \theta_j)}{\prod_{j=1}^{p} \Gamma(\theta_j)} \prod_{j=1}^{p} s_j^{\theta_j - 1}, \tag{5.1}$$

where $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_p]^T$ is the density parameter vector, $s_i \geq 0$, and $\sum_i s_i = 1$. Given a set of N i.i.d. samples $\mathcal{S} = \{\mathbf{s}^{(1)}, \ldots, \mathbf{s}^{(N)}\}$, the normalized log-likelihood of $\boldsymbol{\theta}$ is

$$\begin{aligned}
\mathcal{L}_N(\boldsymbol{\theta}) &\equiv \frac{1}{N} \log\, D_S(\mathcal{S}|\boldsymbol{\theta}) \\
&= \frac{1}{N} \log\, D_S(\mathbf{s}^{(1)}, \mathbf{s}^{(2)}, \ldots, \mathbf{s}^{(N)}|\boldsymbol{\theta}) \\
&= \frac{1}{N} \log \left(\prod_{i=1}^{N} D_S(\mathbf{s}^{(i)}|\boldsymbol{\theta}) \right) \\
&= \frac{1}{N} \sum_{i=1}^{N} \log D_S(\mathbf{s}^{(i)}|\boldsymbol{\theta}) \\
&= \mathbb{T}\left[\log D_S(\mathbf{s}|\boldsymbol{\theta})\right], \qquad (5.2)
\end{aligned}$$

where $\mathbb{T}[\mathbf{x}] \equiv 1/N \sum_{i=1}^{N} \mathbf{x}^{(i)}$ (*i.e.*, $\mathbb{T}[\mathbf{x}]$ is the sample average of $\mathbf{x}$). The ML estimates $\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \mathcal{L}_N(\boldsymbol{\theta})$ cannot be found analytically. The usual choice for obtaining the ML estimates of the parameters is the EM framework [46, 113], which relies on the so-called incomplete data and missing data. Let us denote $\mathcal{S}$ as incomplete data and denote $\mathcal{Z} = \left\{z^{(1)}, \ldots, z^{(N)}\right\}$ as the missing data, which we assume to be i.i.d. and independent of $\mathcal{S}$. Further assume that $z^{(i)} \in \mathbb{R}$ is Gamma

distributed, *i.e.*,

$$p(z) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} z^{\alpha-1} e^{-\beta z}, \tag{5.3}$$

with the inverse scale parameter $\beta = 1$ and the shape parameter $\alpha = \sum_{j=1}^{p} \theta_j$. The complete log-likelihood is, therefore,

$$\begin{aligned} \mathcal{L}_C(\boldsymbol{\theta}) &= \frac{1}{N} \log \left[p_{S,Z}(\mathcal{S}, \mathcal{Z}|\boldsymbol{\theta}) \right] \\ &= -\sum_{j=1}^{p} \log \Gamma(\theta_j) + \mathbb{T} \left[\sum_{j=1}^{p} (\theta_j - 1) \log s_j \right] + \mathbb{T} \left[\Big(\sum_{j=1}^{p} \theta_j - 1 \Big) \log z - z \right]. \end{aligned} \tag{5.4}$$

The EM algorithm iterates between the E-step and the M-step [46, 113].

- E-step:

  Computes the conditional expectation of the complete log-likelihood, given the samples and the current estimate $\widehat{\theta}^{(t)}$. The result is the so-called Q-function

$$\begin{aligned} Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{(t)}) &\equiv \mathrm{E}\Big[\log p_{S,Z}(\mathcal{S}, \mathcal{Z}|\boldsymbol{\theta}) | \, \mathcal{S}, \widehat{\boldsymbol{\theta}}^{(t)} \Big] \\ &= -\sum_{j=1}^{p} \log \Gamma\left(\theta_j\right) + \mathbb{T} \left[\sum_{j=1}^{p} \left(\theta_j - 1\right) \log s_j \right] \\ &\quad + \mathbb{T} \left[\Big(\sum_{j=1}^{p} \theta_j - 1 \Big) \underbrace{\mathrm{E}[\log z | \, \widehat{\boldsymbol{\theta}}^{(t)}]}_{\mathrm{psi} \ \sum_{j=1}^{p} \widehat{\theta}_j^{(t)}} \right] + c, \end{aligned} \tag{5.5}$$

  where c is a constant not depending on $\boldsymbol{\theta}$ and psi$(\cdot)$ is the logarithmic derivative of the Gamma function, also know as the digamma function.

- M-step: Updates the parameter estimates according to

$$\widehat{\boldsymbol{\theta}}^{(t+1)} = \arg\max_{\boldsymbol{\theta}} \left\{ Q\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{(t)}\right) \right\}. \tag{5.6}$$

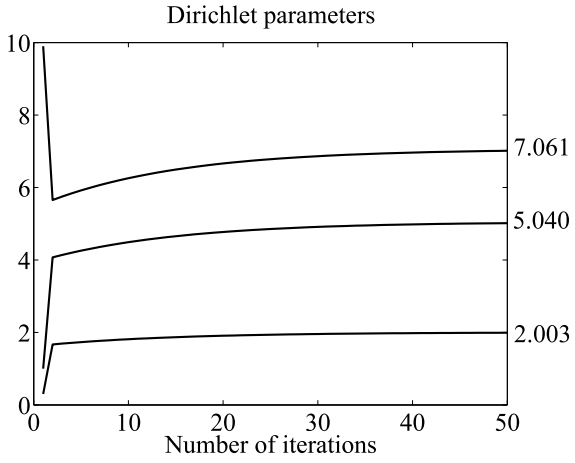


Figure 5.1: Dirichlet parameters estimation.

Maximization of expression (5.6) is still a hard optimization problem. Instead of computing $\boldsymbol{\theta}^{(t+1)}$, we maximize $Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{(t)})$ with respect to θ_j, for $j = 1, \ldots, p$. Thus, we have

$$\begin{aligned} &\textbf{for } j \;=\; 1 \text{ to } p \\ &\qquad \widehat{\theta}_j^{(t+1)} = \arg\max_{\theta_j} \left\{ Q\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{(t)}\right) \right\} \\ &\qquad\qquad = \text{psi}^{-1}\left(\text{psi}\left(\sum_{l=1}^{p} \widehat{\theta}_l^{(t)} \right) + \mathbb{T}\left[\log \widehat{s}_j^{(t)} \right] \right) \\ &\textbf{end}, \end{aligned} \tag{5.7}$$

where $\text{psi}^{-1}(\cdot)$ denotes the inverse of the psi function. Computation details of function $\text{psi}^{-1}(\cdot)$ are described in Appendix B.

The resulting algorithm is of the generalized expectation-maximization class (GEM) [46, 113], as Q is not maximized but only increased.

The obtained algorithm works very well for all range of $\boldsymbol{\theta}$. To illustrate its behavior, we generated 10^5 samples of random variable $\mathbf{s} = [s_1, s_2, s_3]^T$ with $\boldsymbol{\theta} = [2, 5, 7]$. The ML estimates $\widehat{\boldsymbol{\theta}}^{(0)}$ was randomly initialized. Fig. 5.1 illustrates the evolution of $\boldsymbol{\theta}$ as function of the number of iterations. Very accurate estimates are obtained after fifty iterations.

5.2.2 MOD Parameters Estimation

The previous section describes the ML estimation of the parameters of a Dirichlet distribution given a set of samples. Herein, we address the estimation of the parameters of a MOD distribution.

Let the random variable $\mathbf{S} = [S_1, \ldots, S_p]^T$ follow a K-component Dirichlet finite mixture given by

$$p_S(\mathbf{s}|\boldsymbol{\theta}) = \sum_{q=1}^{K} \epsilon_q D(\mathbf{s}|\boldsymbol{\theta}_q), \tag{5.8}$$

where $\epsilon_1, \ldots, \epsilon_K$ are the mixing probabilities satisfying $\sum_{q=1}^{K} \epsilon_q = 1$ and $\epsilon_q \geq 0$ for $q = 1 \ldots, K$, and $\boldsymbol{\theta}_q$ is the set of parameters defining the q-component. The complete set of parameters $\boldsymbol{\theta}$ needed to specify the mixture includes the mixing probabilities $\epsilon_1, \ldots, \epsilon_K$ and the Dirichlet parameters $\boldsymbol{\theta}_q = \{\theta_{q1}, \ldots, \theta_{qp}\}$, for $q = 1, \ldots, K$, *i.e.*, $\boldsymbol{\theta} = \{\epsilon_1, \ldots, \epsilon_K, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K\}$.

Given a set of N i.i.d. samples $\mathcal{S} = \{\mathbf{s}^{(1)}, \ldots, \mathbf{s}^{(N)}\}$, the log-likelihood of $\boldsymbol{\theta}$ is

$$\begin{aligned} \mathcal{L}_N(\boldsymbol{\theta}) &= \frac{1}{N} \log p_S(\mathbf{s}|\boldsymbol{\theta}) \\ &= \mathbb{T}\left[\log \sum_{q=1}^{K} \epsilon_q D(\mathbf{s}|\boldsymbol{\theta}_q)\right]. \end{aligned} \tag{5.9}$$

As in the previous section, we resort to the EM framework because the ML estimate of complete set of parameters $\boldsymbol{\theta}$ can not be found analytically [114]. In this setup, $\mathcal{S}$ denotes the incomplete data and denote $\mathcal{Z} = \left\{\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(N)}\right\}$ the missing data, a set of N labels indicating which component has produced each sample. Each label $\mathbf{z}^{(i)} = [z_1^{(i)}, \ldots, z_K^{(i)}]$ is a binary K-vector, where only one z_q^i is set to one, indicating which component produced the i-sample. The complete log-likelihood is then

$$\begin{aligned} \mathcal{L}_C(\boldsymbol{\theta}) &= \frac{1}{N} \log \left[p_{S,Z}(\mathcal{S}, \mathcal{Z}|\boldsymbol{\theta})\right] \\ &= \mathbb{T}\left[\sum_{q=1}^{K} z_q \log \epsilon_q D(\mathbf{s}|\boldsymbol{\theta}_q)\right]. \end{aligned} \tag{5.10}$$

The E-step of the EM algorithm computes the Q-function given by

$$\begin{aligned} Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{(t)}) &\equiv \mathrm{E}\left[\log p_{S,Z}(\mathcal{S}, \mathcal{Z}|\boldsymbol{\theta})|\, \mathcal{S}, \widehat{\boldsymbol{\theta}}^{(t)}\right] \\ &= \mathbb{T}\left[\sum_{q=1}^{K} \underbrace{\mathrm{E}\left[z_q|\, \mathbf{s}, \widehat{\boldsymbol{\theta}}^{(t)}\right]}_{\beta_q^{(t)}(\mathbf{s})} \log\left[\epsilon_q^{(t)} D(\mathbf{s}|\boldsymbol{\theta}_q^{(t)})\right]\right], \end{aligned} \tag{5.11}$$

where

$$\begin{aligned} \beta_q^{(t)}(\mathbf{s}) &\equiv \mathrm{E}\left[z_q|\, \mathbf{s}, \widehat{\boldsymbol{\theta}}^{(t)}\right] \\ &= \frac{\widehat{\epsilon}_q^{(t)} D\left(\mathbf{s}|\widehat{\boldsymbol{\theta}}_q^{(t)}\right)}{\sum_{l=1}^{K} \widehat{\epsilon}_l^{(t)} D\left(\mathbf{s}|\widehat{\boldsymbol{\theta}}_l^{(t)}\right)}. \end{aligned} \tag{5.12}$$

The M-step, updates the parameter estimates according to

$$\widehat{\boldsymbol{\theta}}^{(t+1)} = \arg\max_{\boldsymbol{\theta}} \left\{ Q\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{(t)}\right)\right\}. \tag{5.13}$$

As in the previous section, we resort to a GEM approach resulting in the following learning rules for the mixing probabilities and for the mixture of Dirichlet source parameters:

$$\epsilon_q^{(t)} = \mathbb{T}\left[\beta_q^{(t)}(\mathbf{s})\right], \tag{5.14}$$

$$\widehat{\theta}_{qj}^{(t+1)} = \mathrm{psi}^{-1}\left(\mathrm{psi}\left(\sum_{l=1}^{p} \widehat{\theta}_{ql}^{(t)}\right) + \frac{\mathbb{T}\left[\beta_q^{(t)}(\mathbf{s}) \log \widehat{s}_j^{(t)}\right]}{\mathbb{T}\left[\beta_q^{(t)}(\mathbf{s})\right]}\right), \tag{5.15}$$

for $q = 1 \ldots, K$ and $j = 1 \ldots, p$, respectively. The learning rule (5.14) maximizes Q-function with respect to $\epsilon_q^{(t)}$ whereas expression (5.15) ensures that the Q-function does not decrease (see [47] for details).

Section 5.4 illustrates aspects of this algorithm with simulated data.

5.3 Statistical Modeling and Unmixing Algorithm

Assuming a noiseless linear observation model ($\mathbf{r} = \mathbf{M}\mathbf{s}$), the spectral vectors are in a $(p-1)-$ dimensional simplex in $\mathbb{R}^L$, due to the abundance constraints, *i.e.*, $\mathbf{s} \in \Delta_p$ [see expression (1.4)]. As mentioned before, if $p \ll L$, it is worthy to project the observed spectral vectors onto the subspace signal. Let $\mathbf{E}_p$ be a unitary matrix spanning the signal subspace. The coordinates of $\mathbf{x}$ with respect to $\mathbf{E}_p$ are

$$\begin{aligned} \mathbf{x} &\equiv \mathbf{E}_p^T \mathbf{r} \\ &= \mathbf{E}_p^T \mathbf{M}\mathbf{s} \\ &= \mathbf{A}\mathbf{s}, \end{aligned} \tag{5.16}$$

where $\mathbf{A}$ is a $p \times p$ square mixing matrix and $\mathbf{x} = [x_1, x_2, \ldots, x_p]^T$ is a $p \times 1$ vector.

Herein, we assume that signature variability and topographic modulation are absent. Then, the full additivity constraint implies that $s_p = 1 - \sum_{j=1}^{p-1} s_j$, *i.e.* only $p-1$ components of $\mathbf{s}$ are free. Recall that $\mathbf{s} = [s_1, \ldots, s_p]^T \in \Delta_p$.

Let's assume that $\mathbf{W} \equiv \mathbf{A}^{-1}$ exists. Then, the abundance estimates is $\widehat{\mathbf{s}} = \mathbf{W}\mathbf{x}$. Consider that each pixel $\mathbf{x}$ represents one particular outcome of a p-dimensional random variable $\mathbf{X} = [X_1, \ldots, X_p]^T$. Given a set of N i.i.d. samples $\mathcal{X} = \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}\}$, then, we may write the likelihood of $\mathbf{W}$ and $\boldsymbol{\theta}$ as

$$\begin{aligned} \mathcal{L}_N(\mathbf{W}, \boldsymbol{\theta}) &\equiv \frac{1}{N} \log p_X(\mathcal{X}|\mathbf{W}, \boldsymbol{\theta}) \\ &= \mathbb{T}\left[\log p_X(\mathbf{x}|\mathbf{W}, \boldsymbol{\theta})\right] \\ &= \mathbb{T}\left[\log p_S(\mathbf{s}|\boldsymbol{\theta})\right] + \log|\det \mathbf{W}|, \end{aligned} \tag{5.17}$$

where we have used the fact that $p_X(\mathbf{x}) = p_S(\mathbf{s})|\det(\mathbf{W})|$. Assuming that the abundance fractions

follow a K-component Dirichlet finite mixture given by expression (5.8), it follows that

$$\mathcal{L}_N(\mathbf{W}, \boldsymbol{\theta}) = \mathbb{T}\left[\log \sum_{q=1}^{K} \epsilon_q D(\mathbf{s}|\boldsymbol{\theta}_q)\right] + \log(|\det \mathbf{W}|). \tag{5.18}$$

Again, the ML estimate of $\left(\widehat{\mathbf{W}}, \widehat{\boldsymbol{\theta}}\right) = \arg\max_{\mathbf{W},\boldsymbol{\theta}} \mathcal{L}_N(\mathbf{W}, \boldsymbol{\theta})$ can not be found analytically.

5.3.1 Unmixing and Mixture Estimation with the EM Framework

Herein, as in Section 5.2.2, the EM framework is adopted to infer the parameters of a mixture. Let $\mathcal{X}$ and $\mathcal{Z}$ denote the incomplete and missing data, respectively. Thus the complete log-likelihood is

$$\begin{aligned} \mathcal{L}_C(\mathbf{W}, \boldsymbol{\theta}) &= \frac{1}{N} \log\left[p_{X,Z}(\mathcal{X}, \mathcal{Z}|\boldsymbol{\theta})\right] \\ &= \mathbb{T}\left[\sum_{q=1}^{K} z_q \log \epsilon_q D(\mathbf{s}|\boldsymbol{\theta}_q)\right] + \log\left(|\det \mathbf{W}|\right). \end{aligned} \tag{5.19}$$

The E-step of the EM algorithm computes the Q-function

$$\begin{aligned} Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{(t)}) &\equiv \mathrm{E}\left[\log p_{X,Z}(\mathcal{X}, \mathcal{Z}|\boldsymbol{\theta})|\, \mathcal{X}, \widehat{\boldsymbol{\theta}}^{(t)}\right] \\ &= \mathbb{T}\left[\sum_{q=1}^{K} \underbrace{\mathrm{E}\left[z_q|\, \mathbf{s}, \widehat{\boldsymbol{\theta}}^{(t)}\right]}_{\beta_q^{(t)}} \log\left[\epsilon_q^{(t)} D\left(\mathbf{s}|\boldsymbol{\theta}_q^{(t)}\right)\right]\right] \\ &\quad + \log\left(|\det \mathbf{W}|\right), \end{aligned} \tag{5.20}$$

where $\beta_q^{(t)}$ is given by expression (5.12) and the M-step estimates the parameters according to expressions (5.14) and (5.15).

Since $\partial Q/\partial \mathbf{W} = 0$ is not a linear equation and cannot be solved analytically, an iterative gradient type learning rule is derived for the unmixing matrix $\mathbf{W}$:

$$\mathbf{W}^{(t+1)} = \mathbf{W}^{(t)} + \tau^{(t)} \left(\frac{\partial Q}{\partial \mathbf{W}}\right)^{(t)}, \tag{5.21}$$

where $\tau^{(t)}$ determines the learning rate on iteration t and

$$\left(\frac{\partial Q}{\partial \mathbf{w}_j}\right)^{(t)} = \mathbb{T}\left[\sum_{q=1}^{K}\left[\beta_q^{(t)}\frac{(\widehat{\theta}_{qj}^{(t)}-1)}{\widehat{s}_j}\mathbf{x}^T - \beta_q^{(t)}\frac{(\widehat{\theta}_{qp}^{(t)}-1)}{\widehat{s}_p}\mathbf{x}^T\right]\right] + \left[\mathbf{W}^{-T}\right]_{j,:} - \left[\mathbf{W}^{-T}\right]_{p,:} \quad , \tag{5.22}$$

where $\mathbf{w}_j$, for $j = 1, \ldots, p-1$ denotes the jth row of matrix $\mathbf{W}$ and $\left[\mathbf{W}^{-T}\right]_{j,:}$ denotes the jth row of the inverse of $\mathbf{W}$ transposed. Rule (5.22) is of the type considered by Lee *et al.* in [100] and by Attias in the IFA framework [6]. The term $\left[\mathbf{W}^{-T}\right]_{j,:} - \left[\mathbf{W}^{-T}\right]_{p,:}$ in expression (5.22) results from the constraint

$$\widehat{s}_p = 1 - \sum_{j=1}^{p-1}\widehat{s}_j, \tag{5.23}$$

which is equivalent to

$$\widehat{\mathbf{w}}_p\mathbf{x} = \mathbf{w}_m\mathbf{x} - \sum_{j=1}^{p-1}\widehat{\mathbf{w}}_j\mathbf{x}, \tag{5.24}$$

where $\mathbf{w}_m$ is a $1 \times p$ vector obtained from p linear independent samples $\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(p)}\}$, *i.e.*,

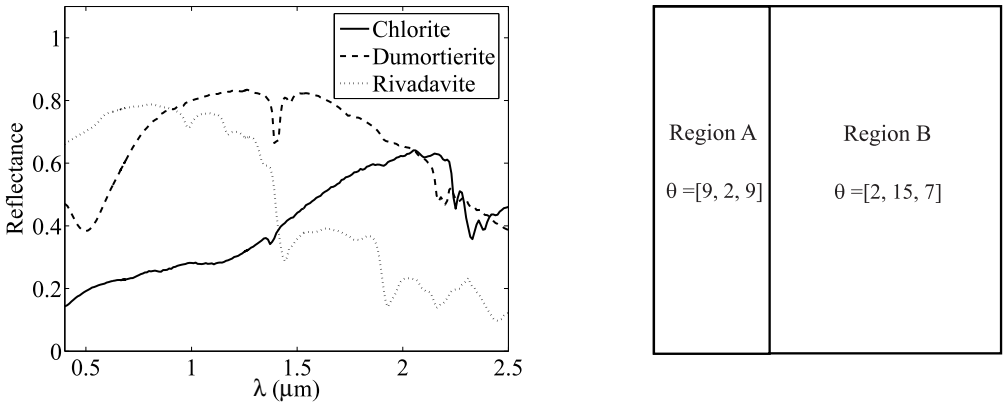
$$\mathbf{w}_m = \left(\left[\mathbf{x}^{(1)} \ldots \mathbf{x}^{(p)}\right]^{-T}\mathbf{1}_p\right)^T. \tag{5.25}$$

Algorithm 5.1 presents the pseudo-code aimed at the maximization of the rule (5.19). It implements a cyclic maximizer algorithm, which splits the estimation of $\mathbf{W}$ and $\boldsymbol{\theta}$ into block maximization operations. The estimation of $\mathbf{W}$ uses a gradient ascendent method with adaptative steps. The estimation of $\boldsymbol{\theta}$ is based on the algorithm [47] described on Section 5.2. The resulting scheme is in the vein of works [6, 117], replacing independent sources represented by MOGs with mixture of Dirichlet sources.

Given the set of N observations $\mathbf{x}$, $\mathbf{W}$ is initialized such that $\widehat{s}_j^{(i)} = \mathbf{w}_j\mathbf{x}^{(i)} > 0$, for $j = 1, \ldots, p$ and $i = 1, \ldots, N$.

Algorithm 5.1 :DECA

initialize τ, $\widehat{\mathbf{W}}$, $\widehat{\boldsymbol{\theta}}$, and L_{best}
while $L_N - L_{best} > threshold$ **do**
  $\widehat{\mathbf{s}} := \widehat{\mathbf{W}}^{(t)}\mathbf{x}$;
  for $q := 1$ to K **do**
    $D\left(\widehat{\mathbf{s}}|\widehat{\boldsymbol{\theta}}_q^{(t)}\right) := \frac{\Gamma\ \sum_{j=1}^{p}\widehat{\theta}_{qj}^{(t)}}{\prod_{j=1}^{p}\Gamma\ \widehat{\theta}_{qj}^{(t)}}\prod_{j=1}^{p}\widehat{s}_j^{\widehat{\theta}_{qj}^{(t)}-1}$;
    $\beta_q^{(t)} := \frac{\widehat{\epsilon}_q^{(t)}D\ \widehat{\mathbf{s}}|\widehat{\boldsymbol{\theta}}_q^{(t)}}{\sum_{l=1}^{K}\widehat{\epsilon}_l^{(t)}D\ \widehat{\mathbf{s}}|\widehat{\boldsymbol{\theta}}_l^{(t)}}$;
    $\widehat{\epsilon}_q^{(t)} := \mathbb{T}\left[\beta_q^{(t)}\right]$;
  end for
  for $j := 1$ to p **do**
    $\theta aux_{qj} := \text{psi}^{-1}\left(\text{psi}\left(\sum_{l=1}^{p}\widehat{\theta}_{ql}^{(t)}\right) + \frac{\mathbb{T}\left[\beta_q^{(t)}\log\widehat{s}_j^{(t)}\right]}{\mathbb{T}\left[\beta_q^{(t)}\right]}\right)$;
    $\frac{\partial Q}{\partial \mathbf{w}_j} := \mathbb{T}\left[\sum_{q=1}^{K}\left(\beta_q^{(t)}\frac{\widehat{\theta}_{qj}^{(t)}-1}{\widehat{s}_j}\mathbf{x}^T - \beta_q^{(t)}\frac{\widehat{\theta}_{qp}^{(t)}-1}{\widehat{s}_p}\mathbf{x}^T\right)\right] + \left[\widehat{\mathbf{W}}^{-T}\right]_j - \left[\widehat{\mathbf{W}}^{-T}\right]_p$;
  end for
  $\mathbf{W}aux := \widehat{\mathbf{W}}^{(t)} + \tau\frac{\partial Q}{\partial \mathbf{W}}$;
  $L_N := \mathbb{T}\left[\log\sum_{q=1}^{K}\widehat{\epsilon}_q^{(t)}D(\mathbf{s}|\boldsymbol{\theta}aux_q)\right] + \log\left(|\det\mathbf{W}aux|\right)$;
  if $L_{best} < L_N$ **then**
    $L_{best} = L_N$;
    $\widehat{\mathbf{W}}^{(t+1)} := \mathbf{W}aux$;
    $\widehat{\boldsymbol{\theta}}^{(t+1)} := \boldsymbol{\theta}aux$;
    increment τ;
  else
    decrement τ;
  end if
end while



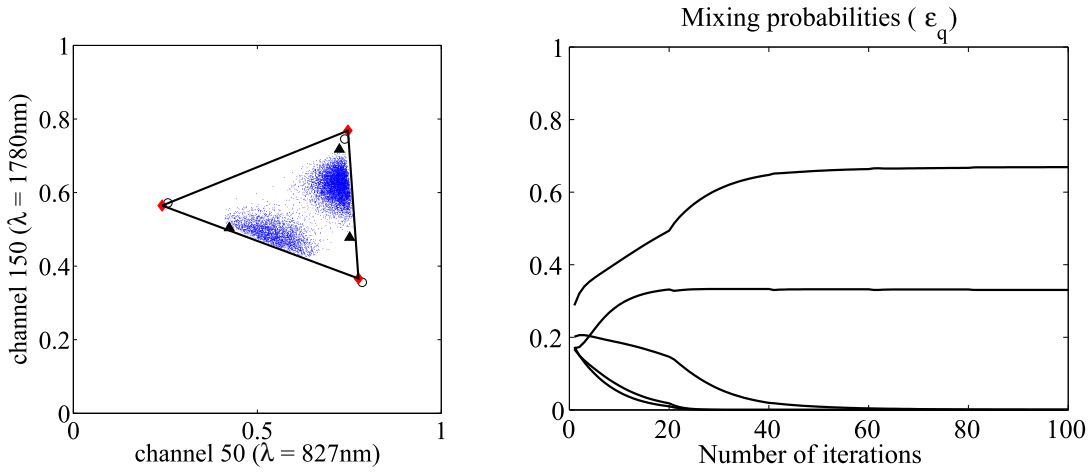
(a) (b)

Figure 5.2: (a) Endmembers signatures used in the simulated experiments; (b) image regions.

5.4 Evaluation with Simulated Data

In this section DECA is tested in simulated scenes. The data is generated according to expression (5.16). Three spectral signatures were selected from the USGS digital spectral library [40] (see Fig. 5.2(a)). The scene is composed by 10^5 pixels partitioned into two regions, region A has the half size of the region B, as schematized in Fig. 5.2(b). The abundance fractions follows a Dirichlet distribution with $\boldsymbol{\theta}_a = [9,\ 2,\ 9]$ and $\boldsymbol{\theta}_b = [2,\ 15,\ 7]$ for region A and region B of the scene, respectively. Pure pixels were removed from the data set in order to illustrate the robustness of DECA in the absence of pure pixels.

In this experiment the number of modes is set to $K = 5$, the Dirichlet parameters are randomly initialized, and the mixing probabilities are set to $\epsilon_q = 1/K$, for $q = 1, \ldots, K$. This setting reflects a situation in which no knowledge of the size and the number of regions in the scene exists. Fig. 5.3(a) presents a scatterplot (bands $\lambda = 827nm$ and $\lambda = 1780nm$) of the simulated scene, where dots represent the pixels. It can be seen the two different clouds corresponding to the region A and B, respectively. It is also presented in the same figure, the true endmembers (circles), the endmembers estimation (diamonds), and for comparison purposes the endmembers estimation by VCA (triangles). Estimates provided by the DECA algorithm are close to the true endmembers.



(a) (b)

Figure 5.3: (a) Scatterplot (bands $\lambda = 827nm$ and $\lambda = 1780nm$) of the three endmembers mixture: true endmembers (circles); VCA estimate (triangles); DECA estimate (diamonds); (b) Dirichlet mixing probabilities.

The algorithm searches for the smallest simplex that contains all data. Whereas, VCA finds the most pure pixels in data (see triangles in Fig. 5.3(a)). Since there is no pure pixels in data, VCA performs worse than DECA.

Fig 5.3(b), presents the evolution of the Dirichlet mixing probabilities (ϵ_q, for $q = 1, \ldots, K$) as function of the number of iterations of the algorithm. Note that three modes tend to zero and the remaining modes have the values of 0.65 and 0.33, corresponding to the weight of the region B and region A respectively. Table 5.1 presents the Dirichlet parameters and their estimates of the two modes. Although the estimated values are near from the true parameter values, we note that this does not have to happen necessarily, since the same distribution can be modeled with different MODs. Note that the main purpose of the DECA algorithm is the estimation of the unmixing matrix $\mathbf{W}$ and not the MOD parameters.

The result of the separation process is illustrated trough the product of the unmixing matrix $\mathbf{W}$ and square mixing matrix $\mathbf{A}$ which is, in an ideal scenario, the identity matrix $\mathbf{I}_p$, apart from

Table 5.1: Estimated Dirichlet parameters.

region A	region B
9	2
2	15
9	7

region A	region B
9.0	2.5
2.2	14.8
10.0	9.7

a permutation. In this experiment the obtained product is

$$\mathbf{WA} = \begin{bmatrix} \mathbf{0.97} & 0.02 & -0.02 \\ 0.03 & \mathbf{0.93} & -0.02 \\ 0.00 & 0.04 & \mathbf{1.03} \end{bmatrix}. \tag{5.26}$$

Fig. 5.4 presents the estimated signatures with the proposed method which shows a good agreement with the true data.

5.5 Experiments with Real Hyperspectral Data

In this section, the proposed method, DECA, is applied to real hyperspectral data collected by the AVIRIS sensor over Cuprite, Nevada (see Section 2.5 for more details). Fig. 5.5 presents the subimage (50×90 pixels and 224 bands) for this experiment. Due to several degradation mechanisms normally found in hyperspectral applications (namely, signature variability, topography modulation, and noise), the observed data is not on a simplex. To obtain a simplex, a projective projection of data onto a hyperplane $\mathbf{y}^T\mathbf{u} = 1$ is implemented as a pre-processing step (see Section 3.2 for more details). The choice of $\mathbf{u}$ ensures that there is no observed vectors orthogonal to the hyperplane.

A visual comparison between the abundance fractions estimates on the cuprite data set and the ground truth presented in [162] shows that first, second, and third extracted endmembers are predominantly Alunite, Kaolinite, and Montmorillonite, respectively (see Fig. 5.6). Note that this results are in accordance with the VCA estimates (see Fig.s 3.12(a), 3.12(g), and 3.12(d),

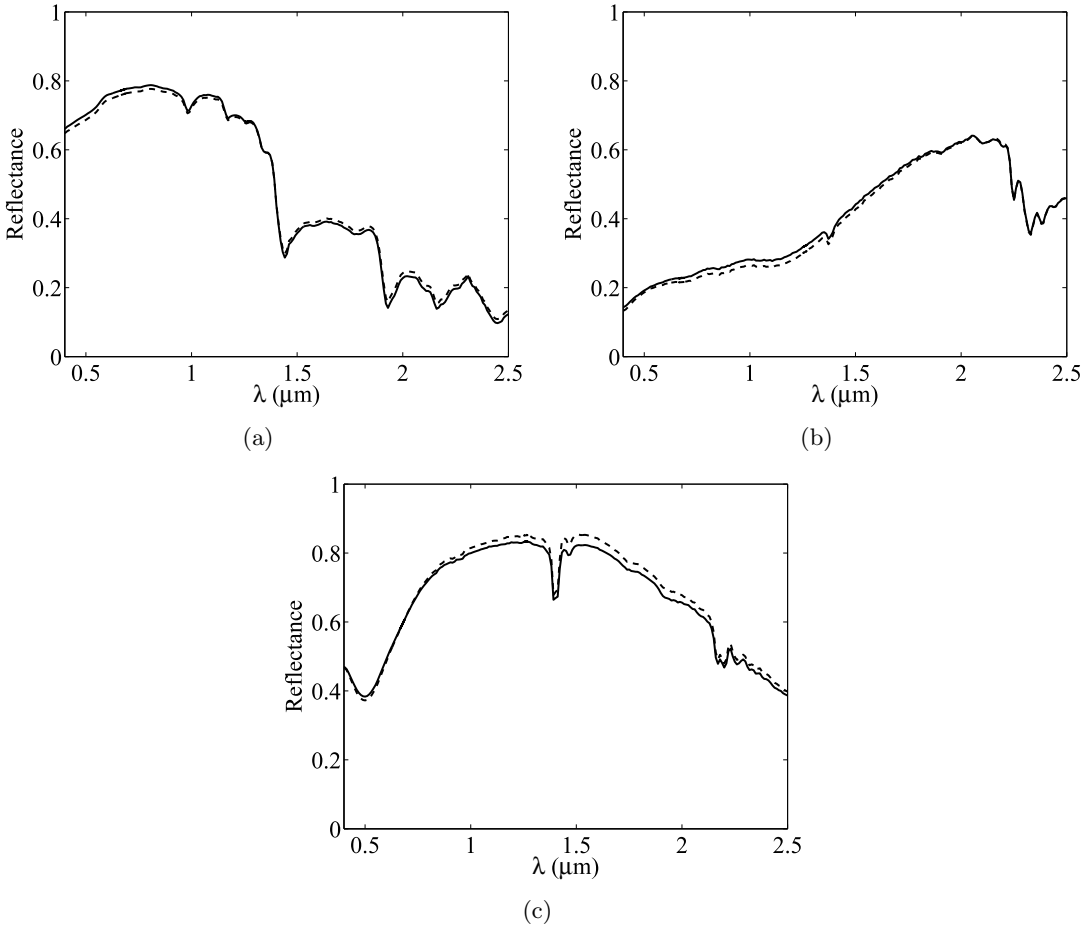


Figure 5.4: Comparison of the endmember signature (solid line) with the DECA estimated signatures (dashed line): (a) Rivadavite; (b) Chlorite; (c) Dumortierite.

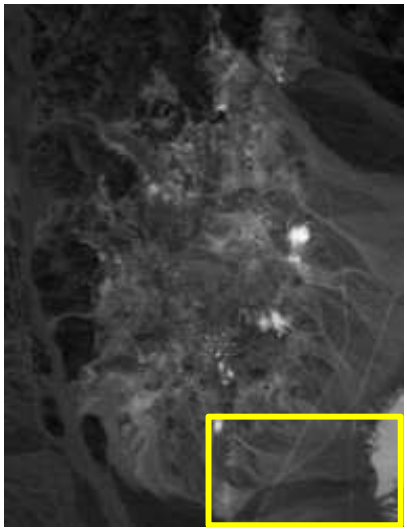


Figure 5.5: Band 30 (wavelength $\lambda = 667.3nm$) of the subimage of AVIRIS cuprite Nevada data set (rectangle denotes the image fraction used in the experiment).

(a) (b) (c)

Figure 5.6: Abundance fractions of three endmembers extracted from cuprite Nevada data set: (a) Alunite; (b) Kaolinite; (c) Montmorillonite.

respectively).

A comparison of the estimated endmember signatures with laboratory spectrum is presented in Fig. 5.7. The signatures provided by DECA are scaled in order to minimize the mean square error between them and the respective library spectra. The estimated signatures are close to the laboratory spectra reflectances. This results are in agreement with the VCA estimates (see Fig.s 3.13(a), 3.13(g), and 3.13(d)).

Fig. 5.8 presents a 3-dimensional visualization of the Cuprite data set in the reduce dimension. It shows that DECA algorithm estimates the smallest simplex (triangle) that contains all data points (dots).

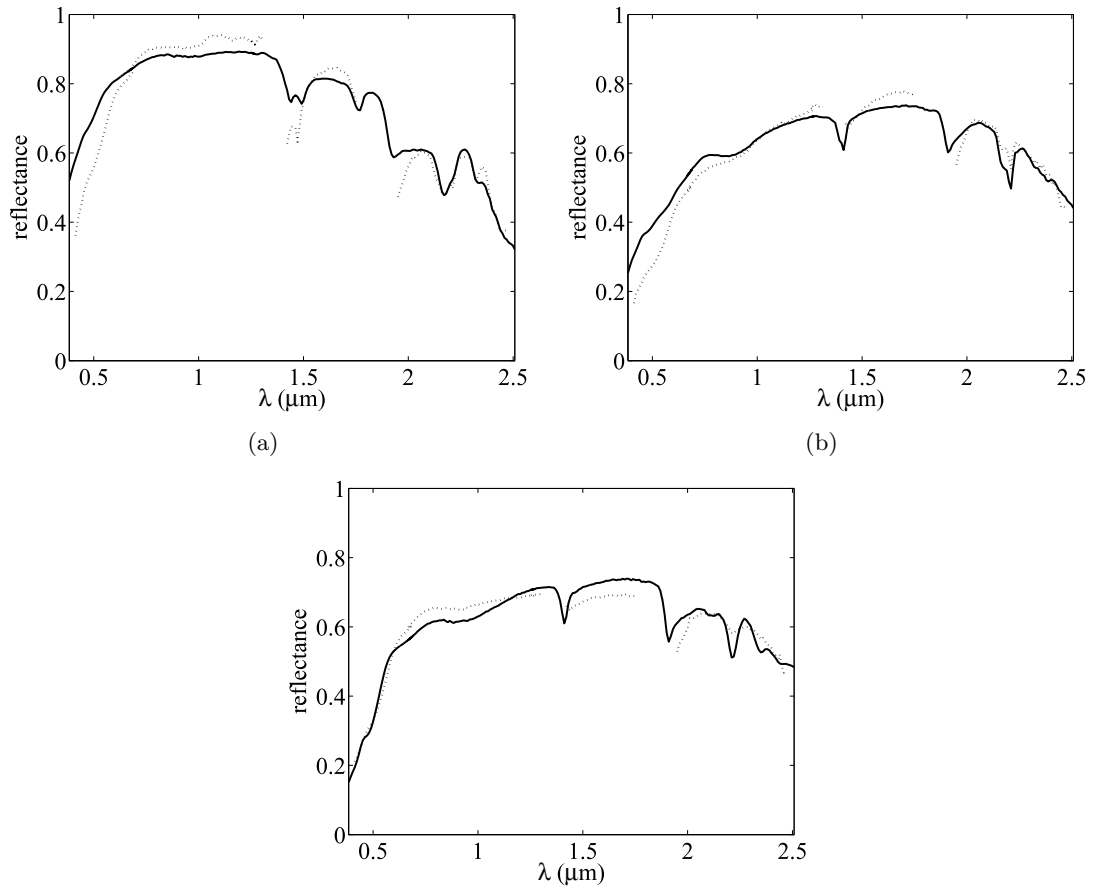


Figure 5.7: Comparison of the DECA estimated signatures (dotted line) with the U.S.G.S spectral library (solid line): (a) Alunite; (b) Kaolinite; (c) Montmorillonite.

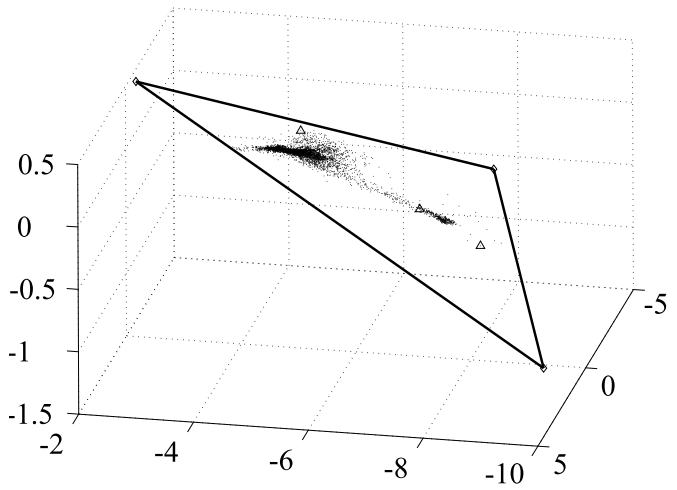


Figure 5.8: 3-dimensional visualization: Cuprite data set (dots); diamonds: DECA estimates; triangles: VCA estimate.

5.6 Conclusions

Blind hyperspectral linear unmixing aims at estimating the number of endmembers, their spectral signatures, and their abundance fractions at each pixel, using only the observed data (mixed pixels). Geometric approaches have been used whenever pure pixels are present in data. In most cases, however, pure pixels can not be found in data. In such cases, unmixing procedures become a difficult task. ICA has been proposed has a tool to unmix hyperspectral data, however, the source dependence present in hyperspectral data compromises the unmixing results. We show in Section 4.3 that the most well known ICA and IFA algorithms do not correctly unmix hyperspectral data.

In chapter paper, a new method is proposed to blindly unmix hyperspectral data, where abundance fractions are modeled as Dirichlet sources. This model forces abundance fractions to be nonnegative and to have constant sum on each pixel. The mixing matrix is inferred by a EM type algorithm. The main advantage of this model is that there is no need to have pure pixels in the observations.

The performance of the proposed model is illustrated with simulated and real hyperspectral data. Comparisons with pure pixel estimation methods are conducted. The results achieved shows the effectiveness of DECA on hyperspectral data unmixing. In future work, the proposed algorithm shall be improved in order to account for sensor noise.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

Unmixing hyperspectral data is the decomposition of the hyperspectral data into a collection of endmembers spectra and their corresponding abundance fractions, thus indicating the proportion of each endmember present in the scene. This thesis introduces new methodologies to perform the different steps involved in the blind unmixing of hyperspectral imagery.

- We adopt a linear mixture model, *i.e.*, each pixel is a linear combination of the endmembers spectra present in the scene. The thesis starts by describing a linear model which accounts for the degradations mechanisms normally found in hyperspectral data, namely, signature variability, topographic modulation, and sensor noise.

- Many hyperspectral data processing and analysis procedures, in particular the unmixing ones, have a dimensional reduction preprocessing step. This operation has a great impact since it reduces the amount of data, implying computational savings in the unmixing step and improvements in the signal-to-noise ratio (SNR).

  This thesis introduces the *hyperspectral signal identification by minimum error* (HySime) method, which is a new mean squared error based approach to infer the signal subspace of hyperspectral data sets. The method first estimates the signal and noise correlations matrices

and then it selects the subset of eigenvalues that best represents the signal subspace in the least square sense.

An alternative method is also proposed, exploiting the structure of hyperspectral mixtures, namely, the fact that spectral vectors are nonnegative.

A set of experiments with simulated and real data leads to the conclusion that the Hysime method is an useful tool in hyperspectral data analysis, yielding comparable or better results than the current state-of-the-art methods.

- Under the linear mixing model, hyperspectral pixels are in a simplex whose vertices correspond to the endmembers present in the scene. Endmember abundance fractions are determined by the position of the mixed pixel in the simplex.

  This thesis introduces the *vertex component analysis* (VCA), which is a new fast method to unmix hyperspectral data exploiting this geometric feature of hyperspectral mixtures. VCA is an unsupervised method that works with project and unprojected data. VCA exploits two facts: i) the endmembers are the vertices of a simplex and ii) the affine transformation of a simplex is also a simplex. Briefly, the VCA algorithm iteratively projects data onto a direction orthogonal to the subspace spanned by the endmembers already determined. The new endmember signature corresponds to the extreme of the projection. The algorithm iterates until all endmembers are exhausted.

  VCA algorithm competes with current state-of-the-art methods, while having a computational complexity between one and two orders of magnitude lower than the best available method.

- In the recent past, several works applied ICA as a tool to unmix hyperspectral data. IFA extends ICA concepts when noise is present. These works are based on the fact that hyperspectral data are, in given circumstances, linear mixtures of endmember signatures weighted by the correspondent abundance fractions. However, these abundance fractions are not inde-

pendent due to physical constraints on the acquisition process.

This thesis addresses the impact of the abundance fraction (sources) dependence on unmixing hyperspectral data with *independent component analysis* (ICA) and with *independent factor analysis* (IFA). The study is based on the behavior of the mutual information of the unmixed sources in the neighborhood of the true unmixed data. The main findings are the following:

1. In hyperspectral linear unmixing, the unmixing matrix minimizing the mutual information might be far from the true one;
2. ICA/IFA performance increases with the SNR;
3. ICA/IFA performance tends to increase with the signature variability and/or with the number of endmembers. The underlying reason is that by increasing the signature variability and/or the number of endmembers the statistical dependence among endmembers is attenuated;
4. There are always endmembers incorrectly unmixed, regardless of the unmixing scenario.

A method based on the source entropy to sort the output of ICA or IFA algorithms according to the likelihood of being correctly separated is proposed.

- Finally, a new direction to blindly unmix hyperspectral data, termed *dependent component analysis* (DECA), using *mixtures of Dirichlet* (MOD) to model abundance fractions is proposed. The method uses a *expectation maximization* (EM) algorithm to estimate the unmixing matrix and to estimate the Dirichlet parameters. Compared with the geometric based approaches, its advantage is that there is no need to have pure pixels in the observations.

6.2 Future Work

This thesis addressed hyperspectral linear unmixing. Among the different methods proposed, DECA is perhaps the one offering more extension possibilites, namely:

- Noise modeling: DECA is developed to unmix noiseless hyperspectral mixtures. An extension of this approach can be derived to account for additive Gaussian noise. This direction is in the vein of Attias work [6].

- Model selection: An important issue is the selection of the number of MOD modes: too many modes lead to overfitting densities; too few modes may not be flexible enough to approximate the source densities. The extension of DECA in order to estimate the number of modes would lead to performance improvements and possible computational savings.

- Parameter initialization: The initialization of the parameters in DECA is by itself an optimization problem that is important to be addressed. The EM framework used in DECA leads to a nonconvex optimization problem, thus sensitive to the initialization. DECA is initialized by an unmixing matrix based on the VCA solution. A better initialization would certainly lead to better results.

Another direction is the exploitation of the nonlinear unmixing. The linear mixture model is simple and well suited to model the many hyperspectral scenarios. However, there are some hyperspectral images where the effects of multiple scattering is not negligible. Thus, it would be useful to derive extensions of DECA to nonlinear mixture. This extensions somehow parallels the extensions of the ICA method to nonlinear ICA.

Appendix A

Atmosphere Correction

Hyperspectral remote sensing exploits the fact that all the materials reflect electromagnetic energy, at specific wavelengths, in distinctive patterns related to their physical and chemical composition [66]. Hyperspectral sensors acquire spectra from the visible region through the near-infrared and mid-infrared ($0.4\mu m - 2.5\mu m$), in hundreds of narrow contiguous bands ($10nm$). Thus, hyperspectral imagery makes possible the remote identification of substances based on their spectral signatures, *i.e.*, their reflectances.

However, when an Earth scene image is acquired by a satellite or an aircraft hyperspectral sensor, the reflectance of the surface substances is only one of the factors that contributes to the radiance acquired by the sensor.

Due to the effects of the illumination source and of the atmosphere, the radiance spectra obtained by an hyperspectral sensor cannot be directly compared to either laboratory spectra or spectra collected at other times or places. To overcome this obstacle, many hyperspectral imagery applications work with the reflectance spectra, which indicates the portion of incident energy which is reflected as a function of wavelength. Thus, one of the most critical steps in most hyperspectral remote sensing imagery is the atmospheric correction, *i.e.*, the method to retrieve the surface reflectance spectra from the observed radiance. Note that if no comparison with laboratory reflectances whatsoever are needed then this step is not required.

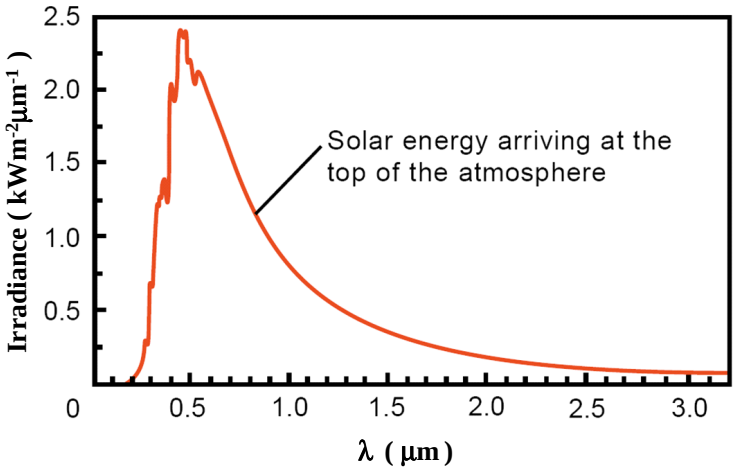


Figure A.1: Example of a typical solar spectrum at the top of the atmosphere. Courtesy of Microimages Inc. (TNTmips® product) available on www.microimages.com

A.1 Atmosphere and Illumination Factors

In a passive remote sensing system, the primary source of illumination is the sun. Solar spectrum is defined as the distribution of the suns emitted energy as a function of wavelength. Fig. A.1 shows a typical solar spectra at the top of the atmosphere. In order to convert sensor measurements to reflectance values, the solar spectrum must be known or estimated. Thus, when a remote image is acquired, the sun angle, the time, the season, and the location of the surveyed area must be known.

As solar energy propagates through the atmosphere, downward and upward (after reflected by the ground), this energy is differently absorbed and scattered depending on the wavelength [63]. Fig. A.2 shows an example of the atmosphere transmittance. Herein it is also shown some gases that contribute to reduce the transmittance in certain wavelengths. Absorption by atmospheric gases is dominated by water vapor with smaller contributions from aerosols, carbon dioxide (CO_2), ozone (O_3), carbon monoxide (CO), methane (CH_4), oxygen (O_2), nitrous oxide (N_2O) and other gases [53, 58]. Strong atmospheric water absorption bands make the atmosphere opaque in many

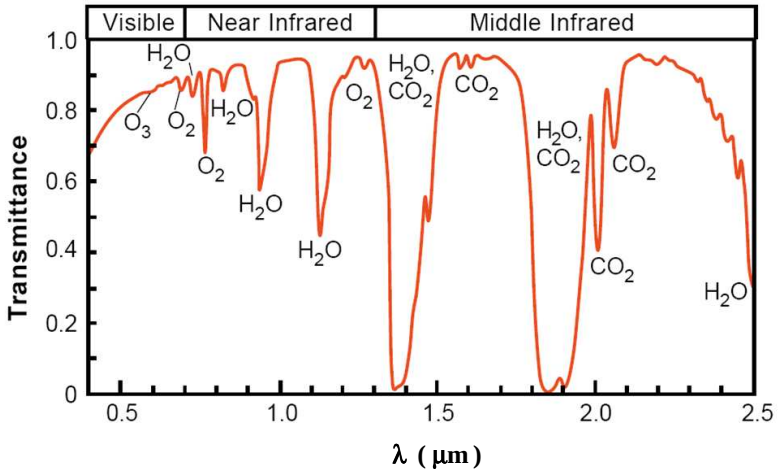


Figure A.2: Example of a typical atmosphere transmittance. Courtesy of Microimages Inc. (TNTmips® product) available on www.microimages.com

regions (for example in the 1.4 and 1.9 μm regions). Note that, as the solar illumination angle, and the viewing angle varies, the path also vary and the atmospheric transmittance as well. The location, the time, and season must be known to derive the atmospheric conditions (water vapor, aerosols, etc).

In addition to atmospheric absorption and scattering, several other important effects have influence on hyperspectral imaging, these effects include

- the angle of the sun and solar spectrum which varies with location, time, and season;
- the viewing angle of the sensor;
- path radiance (upwelling solar radiance scattered by the atmosphere, which never reached the ground);

- skylight (solar radiation scattered by the atmosphere that reach the ground, acting as a secondary illumination);
- adjacency effect (adjacent objects that reflect or scatter the sunlight onto the scene;
- shadowing effects, due to clouds or surface topography.

A.2 Atmospheric Compensation Methods

Over the recent past, several methods and models have been developed to correct atmospheric and illumination effects on hyperspectral imagery. Most of these methods can be classified either as scene-based statistical methods or physics-based modeling methods.

The simplest method to compensate the atmospheric effects is to place a calibration panel, with known reflectance in the scene, in an open area, and use the observed radiance spectrum from the panel to develop statistical relationships between the sensor observations and the known surface reflectance (gain and offset corrections for each waveband of interest). Other methods have been developed, in order to estimate the gain and offset correction terms when calibration panels can not be pre-positioned in the scene. These methods use certain naturally occurring objects in the scene to estimate the gain and offset terms.

The empirical line method (ELM) is a statistics-based model to compensate atmospheric effects [42, 90, 146]. The ELM creates a linear regression equation for each band that provides the gain and the offset of equation

$$\rho(\lambda) = \frac{\mathcal{L}_s(\lambda) - b(\lambda)}{a(\lambda)} \tag{A.1}$$

This method use at least to large and uniform ground target areas with widely different brightness. These selected objects are carefully selected or constructed to provide relatively constant reflectance over the spectral bands of interest. Fig. A.3 shows the principle of this method, where the two lines respect to different wavelengths. This process is equivalent to remove the solar spectral irradiance

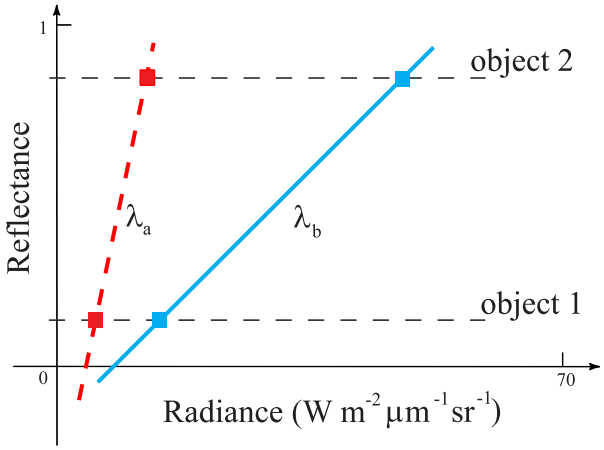


Figure A.3: Schematic of the empirical line method.

modulated by the two-way atmospheric transmittance and the atmospheric path radiance. ELM usually provides accurate estimates, but it does not take into account the topography effects in the scene.

Internal average relative reflectance (IARR) [91, 94] and flat field correction (FFC) [59, 147] are image-based techniques that only take into account the multiplicative gain, *i.e.*, they assume that the atmospheric path radiance is minimal. FFC requires that the image include an uniform area that has a relatively flat and bright spectral reflectance curve. The mean spectrum of that area would be mostly affected by the solar irradiance and atmospheric scattering and absorption. The reflectance image is obtained by dividing the radiance image by the flat field mean spectrum at each band. IARR method also normalizes the radiance image by a mean spectrum. This method derives the mean spectrum from the entire image. The radiance image is first normalized to remove topographic shadows.

However, the dependence upon the presence of suitable reference objects in a scene to perform atmospheric compensation is often limiting. Even if such objects can be found, the implied assumption of atmospheric homogeneity across the scene is often violated. For these reasons, physics based models have been developed to provide atmospheric compensation.

Atmospheric removal (ATREM) [52] and fast line of sight atmospheric analysis of spectral hy-

percubes (FLAASH) [3] are two commonly used methods to model atmospheric effects which apply the radiative transfer theory. The atmospheric compensation used on these two methods are different but both use the three band ratio techniques to account for the effects of the water vapor [51, 53]. The Second Simulation of the Satellite Signal in the Solar Spectrum (6S) [169, 170] is a recent simulator of the atmospheric effects. It enables to simulate plane observations, to account for elevated targets, non Lambertian surface boundary conditions, and Rayleigh and aerosol scattering effects. Moderate Resolution Transmittance (MODTRAN) is a complete package of transmittance and scattering methods, including a full accounting for adjacency effects (the scattering from adjacent pixels into the current pixel sensor line of sight) associated with atmospheric scattering [14, 15]. A detailed discussion, comparison and evaluation of these physics based atmospheric compensation techniques and their implementation differences are described in more detail in [63, 64, 111].

Appendix B

Inversion of the Psi Function

This appendix describes a Newton procedure [116] to compute $x = \mathrm{psi}^{-1}(y)$, *i.e.*, the inverse of the psi function, also know as the digamma function, and given by the logarithmic derivative of the gamma function:

$$y = \mathrm{psi}(x) \equiv \frac{\mathrm{d}\left(\log \Gamma(x)\right)}{\mathrm{d}x}, \tag{B.1}$$

where x must be real and nonnegative. The Newton's method can be used to find the solution of $\mathrm{psi}(x) - y = 0$. The update rule is

$$x^{(t+1)} = x^{(t)} - \frac{\mathrm{psi}(x) - y}{\mathrm{psi}'(x)}, \tag{B.2}$$

where $\mathrm{psi}'(x)$ is the trigamma function, *i.e.*, the first derivative of the psi function, and t denotes the current Newton's iteration. The starting value of x is set to

$$x^{(0)} \simeq \begin{cases} e^y + 0.5 & , y \geq 2.22 \\ -\frac{1}{y+\gamma} & , y < 2.22 \end{cases}, \tag{B.3}$$

based on the following asymptotic expressions of $\mathrm{psi}(x)$:

$$\mathrm{psi}(x) \simeq \begin{cases} \log(x - 0.5) & , x \geq 0.6 \\ -\frac{1}{x} - \gamma & , x < 0.6 \end{cases}, \tag{B.4}$$

where γ is the Euler constant ($\gamma = \text{psi}(1) \simeq 0.577$). An accuracy of fourteen digits is achieved with just only five iterations of the Newton's method.

References

[1] M.J. Abrams, R.P. Ashley, L.C. Rowan, A.F.H. Goetz, and A.B. Kahle. Mapping of hydrothermal alteration in the cuprite mining district, nevada, using aircraft scanner images for the spectral region 0.46 to 2.36 μm. *Geology*, 5:713–718, 1977.

[2] J. B. Adams and M. O. Smith. A new analysis of rock and soil types at the viking lander 1 site. *J. of Geophysical Research*, 91(B8):8098–8112, 1986.

[3] S.M. Adler-Golden, A. Berk, L.S. Bernstein, S. Richtsmeier, P.K. Acharya, M.W. Matthew, G.P. Anderson, C. Allred, L. Jeong, and J. Chetwynd. FLAASH, a MODTRAN4 atmospheric correction package for hyperspectral data retrievals and simulations. In *Proc. 7th Ann. JPL Airborne Earth Science Workshop*, pages 9–14, Pasadena, Calif., 1998. JPL Publications 99-21.

[4] H. Akaike. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.*, 19(6):716–723, 1974.

[5] R.P. Ashley and M.J. Abrams. Alteration mapping using multispectral images - cuprite mining district, esmeralda county,. Open file report 80-367, U.S. Geological Survey, 1980.

[6] H. Attias. Independent factor analysis. *Neural Computation*, 11(4):803–851, 1999.

[7] S. De Backer, P. Kempeneers, W. Debruyn, and P. Scheunders. A band selection technique for spectral classification. *IEEE Geosci. Remote Sensing Let.*, 2(3):319–323, 2005.

[8] C. A. Bateson and B. Curtiss. A tool for manual endmember selection and spectral unmixing. In *Summaries of the V JPL Airborne Earth Science Workshop*, 1993.

[9] C. A. Bateson and B. Curtiss. A method for manual endmember selection and spectral unmixing. *Rem. Sens. of the Environ.*, 55:229–243, 1996.

[10] C.A. Bateson, G.P. Asner, and C.A. Wessman. Endmember bundles: A new approach to incorporating endmember variability into spectral mixture analysis. *IEEE Trans. Geosci. Remote Sensing*, 38:1083–1094, 2000.

[11] Jessica D. Bayliss, J. Anthony Gualtieri, and Robert F. Cromp. Analysing hyperspectral data with independent component analysis. In *Proc. of the SPIE conference 26th AIPR Workshop: Exploiting New Image Sources and Sensors*, volume 3240, pages 133–143, 1997.

[12] A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 10:215–234, 1995.

[13] J.A. Benediktsson, J.A. Palmason, and J.R. Sveinsson. Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Trans. Geosci. Remote Sensing*, 43(3):480–491, 2005.

[14] A. Berk, L. Bernstein, G. Anderson, P. Acharya, D. Robertson, J. Chetwynd, and S. Adler-Golden. MODTRAN cloud and multiple scattering upgrades with application to AVIRIS. *Rem. Sens. of the Environ.*, 65:367–375, 1998.

[15] Alexander Berk, Lawrence S. Bernstein, and David C. Robertson. MODTRAN: A moderate resolution model for LOWTRAN 7. Technical Report GL-TR-89-0122, Spectral Sciences, Burlington, MA, 1989.

[16] José M. Bioucas-Dias and José M. P. Nascimento. Estimation of signal subspace on hyperspectral data. In Lorenzo Bruzzone, editor, *Proc. of SPIE conference on Image and Signal Processing for Remote Sensing XI*, volume 5982, pages 191–198, 2005.

[17] José M. Bioucas-Dias and José M. P. Nascimento. Mixture of dirichlet sources applied to hyperspectral unmixing. In *Proc. of 25th European Association of Remote Sensing Laboratories Symposium*, 2005.

[18] José M. Bioucas-Dias and José M. P. Nascimento. *Signal and Image Processing for Remote Sensing*, chapter Vertex Component Analysis: A Geometric-Based Approach to Unmix Hyperspectral Data, pages 413–437. CRC Press, 2006. in press.

[19] J.W. Boardman. Automating spectral unmixing of AVIRIS data using convex geometry concepts. In *Summaries of the Fourth Annual JPL Airborne Geoscience Workshop, JPL Pub. 93-26, AVIRIS Workshop.*, volume 1, pages 11–14, 1993.

[20] J.W. Boardman. Geometric mixture analysis of imaging spectrometry data. In *Proc. of the IEEE Int. Geosci. and Remote Sensing Symp.*, volume 4, pages 2369–2371, 1994.

[21] J.W. Boardman, F. A. Kruse, and R. O. Green. Mapping target signatures via partial unmixing of AVIRIS data. In *Summaries of the V JPL Airborne Earth Science Workshop*, volume 1, pages 23–26, 1995.

[22] J.W. Boardman and F.A. Kruse. Automated spectral analysis: a geological example using AVIRIS data, northern grapevine mountains, Nevada. In *Proc. of the 10th Thematic Conference, Geologic Remote Sensing*, 1994.

[23] C. C. Borel and S. A. Gerstl. Nonlinear spectral mixing models for vegetative and soils surface. *Rem. Sens. of the Environ.*, 47(2):403–416, 1994.

[24] V. Botchko, E. Berina, Z. Korotkaya, J. Parkkinen, and T. Jaaskelainen. Independent component analisys in spectral images. In *Proc. of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation*, pages 203–207, 2003.

[25] Jeffrey H. Bowles, John A. Antoniades, Mark M. Baumback, John M. Grossmann, Daniel Haas, Peter J. Palmadesso, and John Stracka. Real-time analysis of hyperspectral data sets using NRL's ORASIS algorithm. In *Proc. of the SPIE Conference on Imaging Spectrometry III*, volume 3118, pages 38–45, 1997.

[26] Jeffrey H. Bowles, Mark Daniel, John M. Grossmann, John A. Antoniades, Mark M. Baumback, and Peter J. Palmadesso. Comparison of output from orasis and pixel purity calculations. In *Proc. of the SPIE Conference on Imaging Spectrometry IV*, volume 3438, pages 148–156, 1998.

[27] Jeffrey H. Bowles, Peter J. Palmadesso, John A. Antoniades, Mark M. Baumback, and Lee J. Rickard. Use of filter vectors in hyperspectral data analysis. In *Proc. of the SPIE Conference on Infrared Spaceborne Remote Sensing III*, volume 2553, pages 148–157, 1995.

[28] J. Bruske and G. Sommer. Intrinsic dimensionality estimation with optimaly topologic preserving maps. *IEEE Trans. Pattern Anal. Machine Intell.*, 20(5):572–575, 1998.

[29] J.F. Cardoso. Infomax and maximum likelihood of source separation. *IEEE Signal Processing Lett.*, 4(4):112–114, 1997.

[30] Chein-I Chang. An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis. *IEEE Trans. Inform. Theory*, 46(5):1927–1932, 2000.

[31] Chein-I Chang. *Hyperspectral Imaging: Techniques for spectral detection and classification.* Kluwer Academic, New York, 2003.

[32] Chein-I Chang and Q. Du. Estimation of number of spectrally distinct signal sources in hyperspectral imagery. *IEEE Trans. Geosci. Remote Sensing*, 42(3):608–619, 2004.

[33] Chein-I Chang and D. Heinz. Subpixel spectral detection for remotely sensed images. *IEEE Trans. Geosci. Remote Sensing*, 38(3):1144–1159, 2000.

[34] Chein-I Chang and A. Plaza. A fast iterative algorithm for implementation of pixel purity index. *IEEE Geosci. Remote Sensing Let.*, 3(1):63–67, 2006.

[35] Chein-I Chang, X. Zhao, M. L. G. Althouse, and J. J. Pan. Least squares subspace projection approach to mixed pixel classification for hyperspectral images. *IEEE Trans. Geosci. Remote Sensing*, 36(3):898–912, 1998.

[36] C.H. Chen and X. Zhang. Independent component analysis for remote sensing study. In *Proc. of the SPIE Symp. on Remote Sensing Conference on Image and Signal Processing for Remote Sensing V*, volume 3871, pages 150–158, 1999.

[37] M. Chevrier, A. Bannari, J.-C Deguise, H. McNairn, and K. Staenz. Hyperspectal narrowwavebands for discriminating crop residue from bare soil. In *Proc. of the IEEE Int. Geosci. and Remote Sensing Symp.*, volume 4, pages 2202 – 2204, 2002.

[38] S.-S. Chiang, Chein-I Chang, and I. W. Ginsberg. Unsupervised hyperspectral image analysis using independent component analysis. In *Proc. of the IEEE Int. Geosci. and Remote Sensing Symp.*, 2000.

[39] R. N. Clark and T. L. Roush. Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *J. of Geophysical Research*, 89(B7):6329–6340, 1984.

[40] R. N. Clark, G. A. Swayze, A. Gallagher, T. V. King, and W. M. Calvin. The U.S. geological survey digital spectral library: Version 1: 0.2 to 3.0 μm. Open file report 93-592, U.S. Geological Survey, 1993.

[41] P. Common. Independent component analysis: A new concept. *Signal Processing*, 36:287–314, 1994.

[42] J.E. Conel, R.O. Green, G. Vane, C.J. Bruegge, R.E. Alley, and B.J. Curtiss. Airborne imaging spectrometer-2: Radiometric spectral characteristics and comparison of ways to compensate for the atmosphere. In *proc. of SPIE*, volume 834, pages 140–157, 1987.

[43] T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., New York, 1991.

[44] M. D. Craig. Minimum-volume transforms for remotely sensed data. *IEEE Trans. Geosci. Remote Sensing*, 32:99–109, 1994.

[45] P. Demartines and J. Hérault. Curvilinear component analysis : A self-organizing neural network for nonlinear mapping of data sets. *IEEE Trans. Neural Networks*, 8(1):148–154, 1997.

[46] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39(B):1–38, 1977.

[47] José M. Bioucas Dias. An EM algorithm for the estimation of dirichlet parameters. Technical report, Instituto de Telecomunicações, http://www.lx.it.pt/˜bioucas/, 2005.

[48] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. John Wiley & Sons, Inc., second edition, 2001.

[49] Mario A. T. Figueiredo and Anil K. Jain. Unsupervised learning of finite mixture models. *IEEE Trans. Pattern Anal. Machine Intell.*, 44(3):381–396, 2002.

[50] Bernard R. Foy and James Theiler. Scene analysis and detection in thermal infrared remote sensing using independent component analysis. In *Proc. of SPIE, conference on Independent Component Analyses, Wavelets, Unsupervised Smart Sensors, and Neural Networks II*, volume 5439, pages 131–139, 2004.

[51] B.-C. Gao and A.F.H. Goetz. Column atmospheric water vapor and vegetation liquid water retrievals from airborne imaging spectrometer data. *J. of Geophysical Research*, 95(D4):3549–3564, 1990.

[52] B.-C. Gao, K.B. Heidebrecht, and A.F.H. Goetz. Derivation of scaled surface reflectances from AVIRIS data. *Rem. Sens. of the Environ.*, 44((2/3)):165–178, 1993.

[53] B.-C. Gao and YJ Kaufman. The MODIS near-IR water vapor algorithm, algorithm technical background doc. id: Mod05: Total precipitable water. Technical report, NASA Goddard Space Flight Center, Greenbelt, Md., 1997.

[54] Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman and Hall, second edition, 2004.

[55] S. Geman and D. Geman. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.*, 6(6):721–741, 1984.

[56] A. R. Gillespie, M. O. Smith, J. B. Adams, S. C. Willis, A. F. Fisher, and D. E. Sabol. Interpretation of residual images: Spectral mixture analysis of AVIRIS images, Owens Valley, California. In *Proc. of the 2nd AVIRIS Workshop, R. O. Green, Ed., Jpl Publ. 90-54*, pages 243–270, 1990.

[57] David Gillis, Jeffrey Bowles, Gia M. Lamela, William J. Rhea, Charles M. Bachmann, Marcos Montes, and Tom Ainsworth. Manifold learning techniques for the analysis of hyperspectral ocean data. In Sylvia S. Shen and Paul E. Lewis, editors, *Proc. of the SPIE conferece on Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XI*, volume 5806, pages 342–351, 2005.

[58] A. F. H. Goetz, G. Vane, J. E. Solomon, and B. N. Rock. Imaging spectrometry for earth remote sensing. *Science*, 228(4704):1147–1153, 1985.

[59] A.F.H. Goetz and V. Strivastava. Mineralogical mapping in the cuprite mining district. In *Proc. of the Airborne Imaging Spectrometer Data Analysis Workshop, JPL Publication 85-41*, pages 22–29, 1985.

[60] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. Mathematical Sciences. John Hopkins University Press, third edition, 1996.

[61] A. Green, M. Berman, P. Switzer, and M. D. Craig. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Trans. Geosci. Remote Sensing*, 26(1):65–74, 1988.

[62] Robert O. Green, Michael L. Eastwood, Charles M. Sarture, Thomas G. Chrien, Mikael Aronsson, Bruce J. Chippendale, Jessica A. Faust, Betina E. Pavri, Christopher J. Chovit, Manuel Solis, Martin R. Olah, and Orlesa Williams. Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS). *Rem. Sens. of the Environ.*, 65(3):227–248, 1998.

[63] Michael K. Griffin and Hsiao hua K. Burke. Compensation of hyperspectral data for atmospheric effects. *Lincoln Laboratory Journal*, 14(1):29–54, 2003.

[64] M.K. Griffin, H. K. Burke, J. Vail, S.M. Adler-Golden, and M. Matthew. Sensitivity of atmospheric compensation model retrievals to input parameter specification. In *Proc. AVIRIS Earth Science and Applications Workshop*, pages –, Pasadena, California., 1999. JPL Publications 99-17.

[65] B. Hapke. Bidirection reflectance spectroscopy. I. theory. *J. of Geophysical Research*, 86:3039–3054, 1981.

[66] B. Hapke. *Theory of Reflectance and Emmittance Spectroscopy*. Cambridge Univ. Press, Cambridge, U. K., 1993.

[67] J. Harsanyi, W. Farrand, and Chein-I Chang. Determining the number and identity of spectral endmembers: An integrated approach using neyman-pearson eigenthresholding and iterative constrained rms error minimization. In *Proc. 9th Thematic Conf. Geologic Remote Sensing*, 1993.

[68] J. C. Harsanyi and Chein-I Chang. Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach. *IEEE Trans. Geosci. Remote Sensing*, 32(4):779–785, 1994.

[69] G. Healey and D. Slater. Models and methods for automated material identification in hyperspectral imagery acquired under unknown illumination and atmospheric conditions. *IEEE Trans. Geosci. Remote Sensing*, 37(6):2706–2717, 1999.

[70] Pi-Fuei Hsieh and David Landgrebe. Classification of high dimensional data. Phd thesis and school of electrical & computer engineering technical report tr-ece 98-4, Purdue University, 1998.

[71] Y. H. Hu, H. B. Lee, and F. L. Scarpace. Optimal linear spectral unmixing. *IEEE Trans. Geosci. Remote Sensing*, 37:639–644, 1999.

[72] Rui Huang and Mingyi He. Band selection based on feature weighting for classification of hyperspectral data. *IEEE Geosci. Remote Sensing Let.*, 2(2):156–159, 2005.

[73] A. Hyvärinen. Survey on independent component analysis. *Neural Computing Surveys*, 2:94–128, 1999.

[74] Aapo Hyvarinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley & Sons, Inc., 2001.

[75] Aapo Hyvarinen and Erkki Oja. Independent component analysis: Algorithms and applications. *Neural Networks*, 13(4-5):411–430, 2000.

[76] A. Ifarraguerri and Chein-I Chang. Multispectral and hyperspectral image analysis with convex cones. *IEEE Trans. Geosci. Remote Sensing*, 37(2):756–770, 1999.

[77] A. Ifarraguerri and Chein-I Chang. Unsupervised hyperspectral image analysis with projection pursuit. *IEEE Trans. Geosci. Remote Sensing*, 38(6):127–143, 2000.

[78] A. Ifarraguerri and M. W. Prairie. Visual method for spectral band selection. *IEEE Geosci. Remote Sensing Let.*, 1(2):101–106, 2004.

[79] A. K. Jain. *Fundamentals of Digital Image Processing*. Prentice Hall, 1989.

[80] A. K. Jain and R. C. Dubes. *Algorithms for clustering data*. Prentice Hall, N. J., 1988.

[81] L. O. Jimenez and D. A. Landgrebe. Hyperspectral data analysis and supervised feature reduction via projection pursuit. *IEEE Trans. Geosci. Remote Sensing*, 37(6):2653–2664, 1999.

[82] Ian T. Jolliffe. *Principal Component Analysis*. Spriger Verlag, New York, 1986.

[83] S. Kaewpijit, J. Le Moigne, and T. El-Ghazawi. Automatic reduction of hyperspectral imagery using wavelet spectral analysis. *IEEE Trans. Geosci. Remote Sensing*, 41(4):863– 871, 2003.

[84] E. Kaltofen and G. Villard. On the complexity of computing determinants. In Kiyoshi Shirayanagi and Kazuhiro Yokoyama, editors, *Proc. of the Fifth Asian Symposium on Computer Mathematics*, volume 9 of *Lecture Notes Series on Computing*, pages 13–27, Singapore, 2001.

[85] John P. Kerekes and Jerrold E. Baum. Spectral imaging system analytical model for subpixel object detection. *IEEE Trans. Geosci. Remote Sensing*, 40(5):1088–1101, 2002.

[86] Nirmal Keshava, Jonh Kerekes, Dimitris Manolakis, and Gary Shaw. An algorithm taxonomy for hyperspectral unmixing. In *Proc. of the SPIE AeroSense Conference on Algorithms for Multispectral and Hyperspectral Imagery VI*, volume 4049, pages 42–63, 2000.

[87] Nirmal Keshava and John Mustard. Spectral unmixing. *IEEE Signal Processing Mag.*, 19(1):44–57, 2002.

[88] Naoko Kosaka and Yukio Kosugi. ICA aided linear spectral mixture analysis of agricultural remote sensing images. In *Proc. of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation*, pages 221–226, 2003.

[89] Naoko Kosaka, Kuniaki Uto, and Yukio Kosugi. ICA-aided mixed-pixel analysis of hyperspectral data in agricultural land. *IEEE Geosci. Remote Sensing Let.*, 2(2):220–224, 2005.

[90] F. A. Kruse, K. S. Kierein-Young, and J. W. Boardman. Mineral mapping at cuprite, nevada with a 63 channel imaging spectrometer. *Photogrammetric Engineering and Remote Sensing*, 56(1):83–92, 1990.

[91] F.A. Kruse. Use of airborne imaging spectrometer data to map minerals associated with hydrothermally altered rocks in the northern grapevine mountains, nevada, and california. *Rem. Sens. of the Environ.*, 24(1):31–51, 1988.

[92] F.A. Kruse. Spectral identification of image endmembers determined from AVIRIS data. In *Summaries of the VII JPL Airborne Earth Science Workshop*, 1998.

[93] F.A. Kruse, J.W. Boardman, and J.F. Huntington. Comparison of airborne and satellite hyperspectral data for geologic mapping. In *Proc. of the SPIE Aerospace Conference*, volume 4725, pages 128–139, 2002.

[94] F.A. Kruse, G.L. Raines, and K. Watson. Analytical techniques for extracting geologic information from multichannel airborne spectroradiometer and airborne imaging spectrometer data. In *Proc. of the 4th Thematic Conference on Remote Sensing for Exploration Geology*, pages 309–324, Ann Arbor, Michigan, 1985. Environmental Research Institute of Michigan.

[95] S. Kullback. *Information Theory and Statistics*. Peter Smith, 1978.

[96] D. Landgrebe. Multispectral data analysis: A signal theory perspective. Technical report, Purdue University, 1998.

[97] D. Landgrebe. Hyperspectral image data analysis. *IEEE Signal Processing Mag.*, 19(1):17–28, 2002.

[98] D. Lavenier, J. Theiler, J. Szymanski, Maya Gokhale, and Janet Frigo. Fpga implementation of the pixel purity index algorithm. In *Proc. of the SPIE Photonics East, Workshop on Reconfigurable Architectures*, 2000.

[99] J. B. Lee, S. Woodyatt, and M. Berman. Enhancement of high spectral resolution remote-sensing data by noise-adjusted principal components transform. *IEEE Trans. Geosci. Remote Sensing*, 28(3):295–304, 1990.

[100] T.-W. Lee, M. Girolami, A. Bell, and T. Sejnowski. A unifying information-theoretic framework for independent component analysis. *Int. Journal. on Mathematical and Computer Modeling*, 31:1–21, 2000.

[101] M. Lennon, G. Mercier, M. Mouchot, and L. Hubert-Moy. Curvilinear component analysis for nonlinear dimensionality reduction of hyperspectral images. In *Proc. of the SPIE Symp. on Remote Sensing Conference on Image and Signal Processing for Remote Sensing VII*, volume 4541, pages 157–169, 2001.

[102] M. Lennon, M. Mouchot, G. Mercier, and L. Hubert-Moy. Independent component analysis as a tool for the dimensionality reduction and the representation of hyperspectral images. In *Proc. of the IEEE Int. Geosci. and Remote Sensing Symp.*, 2001.

[103] M. Lennon, M. Mouchot, G. Mercier, and L. Hubert-Moy. Spectral unmixing of hyperspectral images with the independent component analysis and wavelet packets. In *Proc. of the IEEE Int. Geosci. and Remote Sensing Symp.*, 2001.

[104] J. Levesque and K. Staenz. A method for monitoring mine tailings revegetation using hyperspectral remote sensing. In *Proc. of the IEEE Int. Geosci. and Remote Sensing Symp.*, volume 1, 2004.

[105] S. Liangrocapart and Maria Petrou. Mixed pixels classification. In *Proc. of the SPIE Conference on Image and Signal Processing for Remote Sensing IV*, volume 3500, pages 72–83, 1998.

[106] Thomas M. Lillesand, Ralph W. Kiefer, and Jonathan W. Chipman. *Remote Sensing and Image Interpretation*. John Wiley & Sons, Inc., fifth edition, 2004.

[107] K. Liou. *An Introduction to Atmospheric Radiation*. Academic Press, 2nd edition, 2002.

[108] Dimitris Manolakis, Christina Siracusa, and Gary Shaw. Hyperspectral subpixel target detection using linear mixing model. *IEEE Trans. Geosci. Remote Sensing*, 39(7):1392–1409, 2001.

[109] Paolo Mantero, Gabriele Moser, and Sebastiano B. Serpico. SVM-based density estimation for supervised classification of remotely sensed images with unknown classes. In *Proc. of the SPIE Conference on Image and Signal Processing for Remote Sensing IX*, volume 5238, pages 386–397, 2004.

[110] P. Mather. *Land Cover Classification Revisited*, chapter 2, pages 7–16. John Wiley & Sons, Inc., U.K., July 1999.

[111] Michael W. Matthew, Steven M. Adler-Golden, Alexander Berk, Gerald W. Felde, Gail P. Anderson, David Gorodetzky, Scott E. Paswaters, and Margaret Shippert. Atmospheric correction of spectral imagery: Evaluation of the FLAASH algorithm with AVIRIS data. In *proc. of SPIE*, volume 5093, pages 474–482, 2003.

[112] A. S. Mazer, M. Martin, et al. Image processing software for imaging spectrometry data analysis. *Rem. Sens. of the Environ.*, 24(1):201–210, 1988.

[113] Geoffrey McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. John Wiley & Sons, Inc., 1996.

[114] Geoffrey McLachlan and David Peel. *Finite Mixture Models*. John Wiley & Sons, Inc., 2000.

[115] F. Melgani and L. Bruzzone. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sensing*, 42(8):1778–1790, 2004.

[116] Thomas Minka. Estimating a dirichlet distribution. Technical report, M.I.T., 2000.

[117] Eric Moulines, Jean-François Cardoso, and Elisabeth Gassiat. Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models. In *Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, volume 5, pages 3617–3620, 1997.

[118] José M. P. Nascimento and José M. Bioucas-Dias. Classificação não-supervisionada de dados hiperespectrais usando análise em componentes independentes. In *Proc. of the 1st Conferência Científica e Tecnológica em Engenharia*, 2002. in portuguese on CD-ROM.

[119] José M. P. Nascimento and José M. Bioucas-Dias. Does independent component analysis play a role in unmixing hyperspectral data? In F. j. Perales, A.J.C. Campilho, and N. P. Blanca A. Sanfeliu, editors, *Pattern Recognition and Image Analysis*, volume 2652 of *Lecture Notes in Computer Science*, pages 616–625. Springer-Verlag, 2003.

[120] José M. P. Nascimento and José M. Bioucas-Dias. Fast unsupervised extraction of endmembers spectra from hyperspectral data. In *Proc. of SPIE 10th International Symposium on Remote Sensing. Remote Sensing for Environmental Monitoring, GIS Applications, and Geology III*, volume 5239, pages 314–321, 2003.

[121] José M. P. Nascimento and José M. Bioucas-Dias. Fast unsupervised technique for extraction of endmembers spectra from hyperspectral data. In *Proc. of the Fourth Conference on Telelcommunications*, pages 537–540, 2003.

[122] José M. P. Nascimento and José M. Bioucas-Dias. Independent component analysis applied to unmixing hyperspectral data. In *Proc. of SPIE 10th International Symposium on Remote Sensing. Image and Signal Processing for Remote Sensing IX*, volume 5238, pages 306–315, 2003.

[123] José M. P. Nascimento and José M. Bioucas-Dias. Vertex component analysis: A fast algorithm to extract endmembers spectra from hyperspectral data. In F. j. Perales, A.J.C. Campilho, and N. P. Blanca A. Sanfeliu, editors, *Pattern Recognition and Image Analysis*, volume 2652 of *Lecture Notes in Computer Science*, pages 626–635. Springer-Verlag, 2003.

[124] José M. P. Nascimento and José M. Bioucas-Dias. Does independent component analysis play a role in unmixing hyperspectral data? *IEEE Trans. Geosci. Remote Sensing*, 43(1):175–187, 2005.

[125] José M. P. Nascimento and José M. Bioucas-Dias. Estimação do subespaço de sinal em dados hiperespectrais. In *Proc. of the 3rd Jornadas de Engenharia de Electrónica e Telecomunicações e de Computadores*, 2005. in portuguese on CD-ROM.

[126] José M. P. Nascimento and José M. Bioucas-Dias. Signal subspace identification in hyperspectral linear mixtures. In Jorge S. Marques, Nicolás Pérez de la Blanca, and Pedro Pina, editors, *Pattern Recognition and Image Analysis*, volume 3523 of *Lecture Notes in Computer Science*, pages 207–214. Springer-Verlag, 2005.

[127] José M. P. Nascimento and José M. Bioucas-Dias. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE Trans. Geosci. Remote Sensing*, 43(4):898–910, 2005.

[128] José M. P. Nascimento and José M. Bioucas-Dias. *Hyperspectral Data Exploitation: Theory and Applications*, chapter Unmixing Hyperspectral Data: Independent and Dependent Component Analysis. John Wiley & Sons, Inc., 2006. in press.

[129] B. Nash and J. Conel. Spectral reflectance systematics for mixtures of powdered hypersthene, labradoride, and ilmenite. *J. of Geophysical Research*, 79:1615–1621, 1974.

[130] R. Neville, K. Staenz, J. Lefebvre, and P Hauff. Automatic endmember extraction from hyperspectral data for mineral exploitation. In *Proc. of 21st Canadian Symposium on Remote Sensing*, pages 21–24, 1999.

[131] Optimization. *Kenneth Lange*. Springer, first edition, 2004.

[132] Hisham Othman and Shen-En Qian. Noise reduction of hyperspectral imagery using hybrid spatial-spectral derivative-domain wavelet shrinkage. *IEEE Trans. Geosci. Remote Sensing*, 44(2):397–408, 2006.

[133] Lucas Parra, Klaus-Robert Mueller, Clay Spence, Andreas Ziehe, and Paul Sajda. Unmixing hyperspectral data. *Advances in Neural Information Processing Systems*, 12:942–948, 2000.

[134] J. Pearlman, C. Segal, L. Liao, S. Carman, M. Folkman, B. Browne, L. Ong, and S. Ungar. The HYMAP airborne hyperspectral sensor: The system, calibration and performance. In *Proc. of the 1st EARSEL Workshop on Imaging Spectroscopy*, 1998.

[135] J. Pearlman, C. Segal, L. Liao, S. Carman, M. Folkman, B. Browne, L. Ong, and S. Ungar. Development and operations of the eo-1 hyperion imaging spectrometer. In *Proc. of SPIE*, volume 4135, pages 243–254, 2000.

[136] J.S. Pearlman, P.S. Barry, C.C. Segal, J. Shepanski, D. Beiso, and S.L. Carman. Hyperion, a space-based imaging spectrometer. *IEEE Trans. Geosci. Remote Sensing*, 41(6-1):1160–1173, 2003.

[137] M. Petrou and P. G. Foschi. Confidence in linear spectral unmixing of single pixels. *IEEE Trans. Geosci. Remote Sensing*, 37:624–626, 1999.

[138] A. Plaza, P. Martinez, R. Perez, and J. Plaza. Spatial/spectral endmember extraction by multidimensional morphological operations. *IEEE Trans. Geosci. Remote Sensing*, 40(9):2025–2041, 2002.

[139] A. Plaza, P. Martinez, R. Perez, and J. Plaza. A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data. *IEEE Trans. Geosci. Remote Sensing*, 42(3):650–663, 2004.

[140] Antonio Plaza and Chein-I Chang. Fast implementation of pixel purity index algorithm. In Sylvia S. Shen and Paul E. Lewis, editors, *Proc. of the SPIE conference on Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XI*, volume 5806, pages 307–317, 2005.

[141] Antonio Plaza and Chein-I Chang. An improved N-FINDR algorithm in implementation. In Sylvia S. Shen and Paul E. Lewis, editors, *Proc. of the SPIE conference on Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XI*, volume 5806, pages 298–306, 2005.

[142] John A. Richards and Xiuping Jia. *Remote Sensing Digital Image Analysis: An Introduction.* Springer, fourth edition, 2005.

[143] L.J. Rickard, R. Basedow, E. Zalewski, P. Silvergate, and M. Landers. HYDICE: An airborne system for hyperspectral imaging. In *Proc. of SPIE*, volume 1937, pages 173–179, 1993.

[144] J. Rissanen. Modeling by shortest data description. *Automatica*, 14:465–471, 1978.

[145] D. Roberts, M. Gardener, J. Regelbrugge, D. Pedreros, and S. Ustin. Mapping the distribution of wildfire fuels using AVIRIS in the santa monica mountains. In *Summaries of the VIII JPL Airborne Earth Science Workshop*, 1998.

[146] D. Roberts, Y. Yamaguchi, and R.P. Lyon. Calibration of airborne imaging spectrometer data to percent reflectance using field spectral measurements. In *Proceeding of the Nineteenth International Symposium on Remote Sensing of Environment 2*, pages 679–688, Ann Arbor, Michigan, 1985.

[147] D. A. Roberts, Y. Yamaguchi, and R. J. P. Lyon. Comparison of various techniques for calibration of AIS data. In *Proc. of the 2nd AIS workshop*, pages 21–30, Pasadena, CA, 1986. JPL Publication 86-35.

[148] R.E. Roger and J.F. Arnold. Reliably estimating the noise in aviris hyperspectral imagers. *Int. J. Rem. Sens.*, 17(10):1951–1962, 1996.

[149] Louis L. Scharf. *Statistical Signal Processing, Detection Estimation and Time Series Analysis.* Addison-Wesley Pub. Comp., 1991.

[150] G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6:461–464, 1978.

[151] J. J. Settle. On the relationship between spectral unmixing and subspace projection. *IEEE Trans. Geosci. Remote Sensing*, 34:1045–1046, 1996.

[152] G. Shaw and H. Burke. Spectral imaging for remote sensing. *Lincoln Laboratory Journal*, 14(1):3–28, 2003.

[153] G. Shaw and D. Manolakis. Signal processing for hyperspectral image exploitation. *IEEE Signal Processing Mag.*, 19(1):12–16, 2002.

[154] Sylvia S. Shen and Edward M. Bassett. Information-theory-based band selection and utility evaluation for reflective spectral systems. In *Proc. of the SPIE Conference on Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery VIII*, volume 4725, pages 18–29, 2002.

[155] R. Singer. Near-infrared spectral reflectance of mineral mixtures: Systematic combinations of pyroxenes, olivine, and iron oxides. *J. of Geophysical Research*, 86:7967–7982, 1981.

[156] R. B. Singer and T. B. McCord. Mars: Large scale mixing of bright and dark surface materials and implications for analysis of spectral reflectance. In *Proc. of the 10th Lunar and Planetary Sci. Conf.*, pages 1835–1848, 1979.

[157] M. O. Smith, J. B. Adams, and D. E. Sabol. *Spectral mixture analysis-New strategies for the analysis of multispectral data*, pages 125–143. Kluwer Academic Publishers, Brussels and Luxemburg, Belgium, 1994.

[158] K. Staenz, T. Szeredi, and J. Schwarz. ISDAS - a system for processing/analysing hyperspectral data. *Can. J. of Remote Sensing*, 24:99–113, 1998.

[159] G. W. Stewart and J.-G. Sun. *Matrix Perturbation Theory*. Academic Press Inc., 1990.

[160] Lixin Sun and Wen Gao. Method of selecting the best classification bands from hyperspectral images based on genetic algorithm and rough set. In Robert O. Green and Qingxi Tong, editors, *Proc. of the SPIE conference on Hyperspectral Remote Sensing and Application*, volume 3502, pages 179–184, 1998.

[161] G.A. Swayze. *The hydrothermal and structural history of the Cuprite Mining District, Southwestern Nevada: An integrated geological and geophysical approach*. PhD thesis, Purdue University, University of Colorado, 1997.

[162] G.A. Swayze, R.N. Clark, S. Sutley, and A Gallagher. Ground-truthing aviris mineral mapping at cuprite, nevada,. In *Summaries of the Third Annual JPL Airborne Geosciences Workshop*, pages 47–49, 1992.

[163] D. Tanre, , M. Herman, P. Deschamps, and A. de Leffe. Atmospheric modeling for space measurements of ground reflectances, including bidirectional properties. *Applied Optics*, 18:3587–3594, 1979.

[164] J. Theiler, D. Lavenier, N. Harvey, S. Perkins, and J. Szymanski. Using blocks of skewers for faster computation of pixel purity index. In *Proc. of the SPIE International Conference on Optical Science and Technology*, volume 4132, pages 61–71, 2000.

[165] Te Ming Tu. Unsupervised signature extraction and separation in hyperspectral images: A noise-adjusted fast independent component analysis approach. *Optical Engineering of SPIE*, 39(4):897–906, 2000.

[166] J. Scott Tyo, J. Robertson, J. Wollenbecker, and R. C. Olsen. Statistics of target spectra in hsi scenes. In *Proc. of the SPIE conference on Imaging Spectrometry VI*, volume 4132, pages 306–314, 2000.

[167] S.G. Ungar, J.S. Pearlman, J.A. Mendenhall, and D. Reuter. Overview of the earth observing one (eo-1) mission. *IEEE Trans. Geosci. Remote Sensing*, 41(6-1):1149–1159, 2003.

[168] G. Vane, R. Green, T. Chrien, H. Enmark, E. Hansen, and W. Porter. The airborne visible/infrared imaging spectrometer (AVIRIS). *Rem. Sens. of the Environ.*, 44:127–143, 1993.

[169] E. Vermote, D. Tanré, J. Deuzé, M. Herman, and J. Morcette. Second simulation of the satellite signal in the solar spectrum 6S: An overview. *IEEE Trans. Geosci. Remote Sensing*, 35(3):675–686, 1997.

[170] E. Vermote, D. Tanre, J.L. Deuze, M. Herman, and J.J. Morcrette. *Second Simulation of the Satellite Signal in the Solar Spectrum (6S): 6S Users Guide Version 2*. NASA Goddard Space Flight Center, Greenbelt, MD, 1997.

[171] Larry Wasserman. *All of Statistics : A Concise Course in Statistical Inference*, chapter 5. Springer, 2004.

[172] M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *IEEE Trans. Acoust., Speech, Signal Processing*, 33(2):387–392, 1985.

[173] M. E. Winter. N-findr: an algorithm for fast autonomous spectral end-member determination in hyperspectral data. In *Proc. of the SPIE conference on Imaging Spectrometry V*, volume 3753, pages 266–275, 1999.

[174] M. E. Winter. Comparison of approaches for determining end-members in hyperspectral data. In *Proc. of the IEEE Aerospace Conference*, volume 3, pages 305–313, 2000.

[175] H.-H.P. Wu and R. A. Schowengerdt. Improved estimation of fraction images using partial image restoration. *IEEE Trans. Geosci. Remote Sensing*, 31(4):771–778, 1993.

[176] R. H. Yuhas, A. F. H. Goetz, and J. W. Boardman. Discrimination among semi-arid landscape endmembres using the spectral angle mapper (SAM) algorithm. In *Summaries of the 3rd annu. JPL Airborne Geosci. Workshop, R. O. Green, Ed. Publ., 92-14*, volume 1, pages 147–149, 1992.