1-1-2011

Hyperspectral image analysis using a simultaneous denoising and intrinsic order selection (DIOS) approach

Masoud Farzam
Ryerson University

Follow this and additional works at: <http://digitalcommons.ryerson.ca/dissertations>

Part of the [Electrical and Computer Engineering Commons](#)

Recommended Citation

Farzam, Masoud, "Hyperspectral image analysis using a simultaneous denoising and intrinsic order selection (DIOS) approach" (2011). *Theses and dissertations*. Paper 872.

This Dissertation is brought to you for free and open access by Digital Commons @ Ryerson. It has been accepted for inclusion in Theses and dissertations by an authorized administrator of Digital Commons @ Ryerson. For more information, please contact bcameron@ryerson.ca.

Hyperspectral Image Analysis Using A Simultaneous Denoising and Intrinsic Order Selection (DIOS) Approach

by

Masoud Farzam

M. Sc. in Engineering Physics, University of Amirkabir, Tehran, July 1998

B. Sc. in Electrical Engineering, Isfahan University of Technology, Isfahan, June 1989

A dissertation

presented to Ryerson University

in partial fulfillment of the
requirement for the degree of

PhD

in the Program of
Electrical and Computer Engineering.

Toronto, Ontario, Canada, 2011

© Masoud Farzam, 2011

Author's Declaration

I hereby declare that I am the sole author of this thesis.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Signature: _____

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Signature: _____

Instructions on Borrowers

Ryerson University requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

Hyperspectral Image Analysis Using A Simultaneous Denoising and Intrinsic Order Selection (DIOS) Approach

PhD in the Program of Electrical and Computer Engineering, 2011

Masoud Farzam, Ryerson University

Abstract

Recent hyperspectral applications demand for higher accuracy and speed. This thesis develops a hyperspectral analysis solution to address challenges in the different steps of denoising, order selection and unmixing of hyperspectral data. Currently, all these steps process the data in cascade to achieve the optimum results. While in existing approaches the desired criterion is different in these steps, the proposed simultaneous Denoising and Intrinsic Order Selection (DIOS) method unifies these criteria. This property not only makes more sense for the desired optimization problem, but also leads to a faster processing algorithm. Consequently, DIOS avoids possible error propagation from the denoising stage to the dimension estimation stage, leading to more accurate results. The proposed method is based on minimizing the estimated Mean Square Error (MSE). The success rate of existing dimension estimation methods declines with the increase of image dimension and the decrease of Signal-to-Noise Ratio (SNR). The most competitive method fails to detect the correct dimension in 30% of cases around 2dB. However, in simulation results DIOS is shown to be successful with a failure rate of about 5%. The proposed unmixing method, based on a simple least square estimation, improves the speed performance least 10 times for an average-sized data cube of 2MB. Compared to some well known existing approaches, the unmixing method improves the estimated MSE up to 60% for $\text{SNR} < 10\text{dB}$. A new whitening process for hyperspectral applications with coloured noise is also proposed. Since the proposed method avoids the inversion of large matrices, computational complexity is substantially decreased. In the presence of coloured noise, simulation results show that the proposed whitening method lowers the MSE of unmixing and outperforms the existing whitening methods particularly when the noise correlation factors increase.

Acknowledgments

I wish to dedicate this thesis to my family. To my mother, whom I wish was alive to see this success, my father, whom I get my scientific curiosity and persistence from, my lovely wife who has been always supporting me in my life and my lovely kids who were patient with me to finish this work.

To my graduate supervisory committee for the time you invested as well as the guidance offered to improve my graduate research.

To my graduate supervisor, Dr. Soosan Beheshti, whose great vision and support made the most contribution to the success of this research.

Contents

1	Introduction	1
1.1	Hyperspectral Imaging Problem Formulation	2
1.1.1	Terminology	2
1.2	Challenges of Hyperspectral Applications	4
1.3	Thesis Contribution and Outline	5
1.3.1	Thesis Outline	7
2	Background	8
2.1	Mixing Models	8
2.1.1	Linear Mixture Model (LMM)	8
2.1.2	Non-linear Mixing Model	10
2.1.3	Stochastic Model	11
2.1.4	Mixing Model Used in Hyperspectral Applications	12
2.2	Hyperspectral Analysis Methods	12
2.2.1	Advancement in Hyperspectral Imaging Applications	12
2.2.2	Dimension Reduction Methods	14
2.2.3	Denoising Methods	19
2.2.4	Order Selection Methods	23
2.2.5	Hyperspectral Unmixing Methods	25
3	Simultaneous Denoising and Intrinsic Order Selection (DIOS) Method	32
3.1	Introduction	32
3.2	Ultra-Fast Unmixing (UFU) Method	33
3.2.1	Step 1: Least Square Error (LSE) Based Extraction Method	34
3.2.2	Step 2: Iterative Selection of Endmembers ($m > 2$)	35
3.2.3	Spectral Library Matching Approach	37
3.2.4	Simulation Results for UFU	39

3.2.5	A Note on Speed Efficiency of UFU	39
3.2.6	A Note on the Computational Complexity of UFU	44
3.3	Hyperspectral Order Selection Method	45
3.3.1	Data Error versus Reconstruction Error	45
3.3.2	MRE Calculation	47
3.3.3	MRE Calculation for Hyperspectral Images	49
3.3.4	Noise Variance Estimation	52
3.3.5	Non-orthogonality and the Estimation Error	52
3.3.6	Simulation Results for MRE	55
3.4	Simultaneous Denoising and Order Selection Method	60
3.4.1	Simulation Results for DIOS	61
3.5	Hyperspectral Correlation Extractor (HYCE) Method: Colored Noise Correlation Calculation	76
3.5.1	Correlation Between Bands	78
3.5.2	Mathematical Noise Model	79
3.5.3	Noise Estimation	79
3.5.4	Simulation Results for HYCE	81
3.5.5	Simulation Results of HYCE with MRE and UFU	83
3.6	Summary	84
4	Concluding Remarks and Future Work	90
4.1	Some Potential Future Work	91
A	Upper Bound and Lower Bound on Δ_{S_m}	92
B	GA-LSE Algorithm	94
B.0.1	Step 1: Estimation of Virtual Dimensionality	95
B.0.2	Step 2: LSE-Based Endmember Extraction	96
B.0.3	Step 3: GA estimation of abundance vectors	96
B.1	Simulation Results	97
C	Regression Method and RAP for Noise Variance Estimation	102
C.1	Regression Method	102
C.2	Residual Autocorrelation Power	103

D Hyperspectral Robust Analysis Package (HyRap)	107
D.1 ENVI Software	107
D.2 HyRap: Hyperspectral Analysis Package	110
D.3 ENVI versus HyRap	112
E List of Publications	117
Bibliography	119

List of Figures

1.1	An example of a hyperspectral data cube and pixel radiance	3
1.2	Endmembers and spectral signatures for some minerals (USGS library [1])	3
1.3	New applications of hyperspectral imaging in science and industry (Adopted from Cytoviva Inc. [2])	4
1.4	HSI analysis diagram	6
2.1	Mixing and unmixing diagram	8
2.2	Pure pixel versus mixed pixel	9
2.3	Left: Pixel constitution in deterministic, Right: Pixel constitution in stochastic approach	11
2.4	Left: Random correlated data set, Right: Zero-mean data set	15
2.5	SNR versus the number of eigenimages for the Cuprite site data	17
2.6	First, fourth, sixth and ninth PC	18
2.7	The relation between eigenvalues of correlation and covariance matrix for SNR=10 and $c=6$	25
2.8	Linear classifier	31
3.1	A 3-D model of a hyperspectral data vector. $S = [s_{11} \ s_{12} \ s_{13}]$	34
3.2	UFU estimation of endmembers and abundances (SNR=10dB). MSE data is presented in Table 3.1 and Fig. 3.3)	40
3.3	Line of true and estimated abundances for 32 pixels in a row of Fig. 3.2(SNR=10dB)	41
3.4	A comparison of the processing time (using MATLAB implementation with dimension $c = 10$) for different unmixing methods	42
3.5	Averaged processing time and 2σ intervals versus data cube size (using MATLAB implementation with $c = 5$, SNR=5dB) for 100 runs	43

3.6	Comparison of processing time versus data cube size (using MATLAB implementation with $c = 5$, SNR=5dB) for UFU, VCA, GDME and FCLS . .	43
3.7	Averaged processing time versus number of endmembers with 2σ intervals (using MATLAB implementation with datacube size=2MB, SNR=5dB) for UFU	44
3.8	Comparison of processing time versus number of endmembers(using MATLAB implementation with datacube size=2MB, SNR=5dB) for UFU, VCA, GDME and FCLS	45
3.9	Reconstruction error in the presence or absence of noise as a function of m , number of endmembers	47
3.10	Data error in presence or absence of noise as a function of the number of endmembers	47
3.11	Error versus number of endmembers for SNR=10 and applying orthogonalization	53
3.12	Geometric representation of error due to the orthogonalization	54
3.13	Error graph for SNR=10 and applying SVD and orthogonalization	55
3.14	Rank estimation for SNR=5, c = number of constituent endmembers=7 . .	56
3.15	Rank estimation for SNR=5, 3, 2, 0	57
3.16	Rank estimation for different numbers of constituent endmembers	58
3.17	DIOS analysis diagram	62
3.18	Error behavior as a function of choice of dimension when SNR is 15dB and for various values for the true dimensions $c=12, 13, 15$, and 16 respectively.	63
3.19	Error behavior as a function of choice of dimension for a fixed value of $c = 12$ and variable SNRs (SNR=2, 5, 8, and 10dB)	64
3.20	Comparing different methods of dimension estimation at SNR=10dB. Statistical data is presented in Fig. 3.22.	65
3.21	Comparing different methods of dimension estimation at SNR=5dB. Statistical data is presented in Fig. 3.22.	66
3.22	Averaged dimension estimation at SNR=5dB with 2σ line intervals. . . .	66
3.23	Histogram of estimated dimensions for HFC, EIF and DIOS with SNR=2dB and $c=15$ for 100 runs. Statistical success rate is presented in Table 3.4. . .	69
3.24	a) Original synthetic noiseless data in band 100, b) Noisy data in the same band (SNR=10dB), c) Denoised data using DIOS	69
3.25	Dimension estimation with different maximum abundance values (exact dimension=10)	72

3.26 Spectral Information Divergence (SID) comparison of VCA, GDME and DIOS	73
3.27 Averaged MSE of endmember estimation error for 100 runs in VCA, GDME, FCLS and DIOS	74
3.28 Abundance fractions estimated with the DIOS for Cuprite site data. SAD data is presented in Table 3.8.	77
3.29 Extracted signatures from Cuprite site data using the DIOS: Library endmembers (dashed line), Estimated endmembers (Solid line). SAD data is presented in Table 3.8.	85
3.30 Partial correlation coefficient plot for Indian Pine image (area in the square is magnified on the right side)	86
3.31 Variance versus mean squared scatterplot	86
3.32 Unity-lag covariance and variance scatterplot	87
3.33 Two-lag covariance and variance scatterplot	87
3.34 Estimation of different variances and correlation factors	88
3.35 Sensitivity of HYCE to the variation of correlation factors	88
3.36 HYCE comparison to NWHFC for different ρ_1 and ρ_2	89
3.37 a)MRE estimation when correlated noise is ignored (SNR=5 dB, c=12) with no minimum point to detect the correct dimension, b)MRE estimation for a correlated noise (SNR=5 dB, c=12) after HYCE is applied. Minimum point is representing the correct dimension.	89
B.1 LSE error versus the population size	98
B.2 The GA-LSE estimation of endmembers and abundances (SNR=20)	99
B.3 Gaussian noise effect on unmixing: SNR=10	100
B.4 Noise scattered plot for SNR=7	100
C.1 Residuals autocorrelation power (RAP) for a range of standard deviation candidates when the true stddev is 15 for the Lenna image.	105
C.2 Difference of RAP, D_κ in (C.8), for the RAPs in Figure C.1.	105
D.1 Hyperspectral processing flow implemented in ENVI and HyRap	108
D.2 Snapshot of the MNF eigenvalue plot used in ENVI (adopted from [3]) . .	109
D.3 Snapshot of an ENVI mapping output (adopted from [3])	109
D.4 Snapshot from Hyrap datacube manipulation features	110
D.5 Hyrap dimension estimation feature	111

D.6 Hyrap unmixing feature	112
D.7 Hyrap classification feature	113
D.8 Peanut butter infected by the salmonella bacteria	114
D.9 ENVI mapping of the salmonella bacteria	114
D.10 Hyrap classification results	114
D.11 Salmonella and PB separate signatures	115
D.12 Salmonella and PB mixed signatures	115

List of Tables

2.1	Eigenvalues of the Cuprite site in AVIRIS data along with the proportions and accumulations	16
3.1	Averaged MSE error of abundance estimates for UFU and VCA	40
3.2	Comparison of rank estimation for SNR=10 and $c=16$	59
3.3	Comparison of the Multiple Regression method and RAP noise estimation	60
3.4	Success rate statistical results of DIOS, HFC and EIF for 100 runs. Histogram has been presented in Fig. 3.23.	67
3.5	Comparison of dimension estimation methods with different SNRs and different constituent endmembers	70
3.6	Statistical results of DIOS and VCA estimation for 100 runs and $C = 10$.	71
3.7	Comparison of dimension estimation for Indian Pines test site with number of end members $c=16$	75
3.8	Spectral angle distance between library spectra and extracted endmembers by DIOS, VCA and GDME	76
B.1	Spectral angle distance between USGS spectra and estimated spectra by UFU, VCA and GA-LSE	101

List of Appendices

A.	Upper Bound and Lower Bound on Δ_{S_m}	92
B.	GA-LSE Algorithm	94
C.	Regression Method and RAP for Noise Variance Estimation	102
D.	Hyperspectral Robust Analysis Package (HyRap)	107
E.	List of Publications	117

ABBREVIATIONS

HSI	HyperSpectral Imaging
EM	EndMember
MSE	Mean Squared Error
IFOV	Instantaneous Field of View
LMM	Linear Mixture Model
NLMM	NonLinear Mixture Model
LSMA	Linear Spectral Mixture Analysis
NLSMA	Non-Linear Spectral Mixture Analysis
SVD	Singular Value Decomposition
USGS	United States Geological Survey
PCA	Principal Component Analysis
MNF	Maximum Noise Fraction
ICA	Independent Component Analysis
SNR	Signal to Noise Ratio
VCA	Vertex Component Analysis
GDME	Gradient Descent Maximum Entropy
NCLS	Nonnegative Constrained Least Squares
SVM	Support Vector Machine
GA	Genetic Algorithm
UFU	Ultra Fast Unmixing
DIOS	Denoising and Intrinsic Order Selection
MRE	Minimum Reconstruction Error
RMS	Root Mean Squared
FCLS	Fully Constrained Least Squares
SIP	Signal and Information Processing
SID	Spectral Information Divergence
ANC	Abundance Non-negativity Constraint
ASC	Abundance Sum-to-one Constraint
MDL	Minimum Description Length
VD	Virtual Dimensionality

ID	Intrinsic Dimensionality
RAP	Residual Autocorrelation Power
HRDSDC	Homogeneous Regions Division and Spectral De-Correlation
GGD	General Gaussian Distribution
HFC	Harsanyi-Farrand-Chang
NWHFC	Noise Whitenes Harsanyi-Farrand-Chang
HYSIME	Hyperspectral Subspace Identification
MVC-NMF	Minimum Volume Constrained Nonnegative Matrix Factorization
SAD	Spectral Angle Mapper
PCC	Partial Correlation Coefficient
JPL	Jet Propulsion Laboratory
VIS	Visual Information Solutions
RSI	Research Systems Incorporation
PPI	Pure Pixel Index
SAM	Spectral Angle Mapper

NOMENCLATURE

List of lowercase symbols

$\tilde{y}_{i,j,k}$	Noiseless hyperspectral signal of (i, j) th pixel on k th band
$y_{i,j,k}$	Noisy hyperspectral signal of (i, j) th pixel on k th band
$w_{i,j,k}$	Noise of (i, j) th pixel on k th band
$\tilde{y}_{i,j}$	Noiseless pixel vector of (i, j) th pixel
$y_{i,j}$	Noisy pixel vector of (i, j) th pixel
$w_{i,j}$	Noise vector of (i, j) th pixel
$\tilde{y}_i$	i th Concatenated noiseless pixel vector
y_i	i th Concatenated noisy pixel vector
w_i	i th Concatenated noise vector
a_i	i th Endmember vector
s_i	i th abundance in the abundance vector
Var	Variance
b	Bias of the best fit line
$h(s)$	Objective function for abundances
p	Number of free parameters
$f_0(s)$	Entropy of abundances
t_b	Bayesshrink threshold
t_v	Visuhrink threshold
t_s	Sureshrink threshold
x_m	Data error sample
z_m	Reconstruction error sample

List of uppercase symbols

A_c	Endmember matrix
S	Abundance Vector
W	Noise random variable
C	Number of endmembers
C_{max}	Total number of endmembers in the library
A_{NL}	Non-linear endmember matrix
L	Number of eigenvalues
T_i	Transformation matrix in the i th step
K_1	First pixel index
K_2	Second pixel index

$\hat{A}$	Estimation of the endmember matrix
N	Total number of pixels
M	Margin in SVM method
P_s	Signal power
P_t	Total power
Z_m	Reconstruction error random variable
X_m	Data error random variable
B_m	Endmembers not in the subset m
R_{ww}	Auto correlation matrix of random variable w
D_k	k th difference

List of Greek symbols

Σ	Covariance matrix
ℓ	Number of spectral bands
μ_k	Mean of the k th cluster
$\sum_k$	Covariance of the k th cluster
α_{NL}	Non-linear abundances
φ	Eigenvector
λ	Eigenvalue
Ψ_k	k th left-hand eigenvector
Σ_w	Noise covariance matrix
ξ	Lagrange multiplier
v	Left singular value
α	Abundance subset
η	Signal size or number of samples
ω	Slope of the best fit line
ϵ	Sureshrink error
θ	Coefficient of the signal
φ	Pseudo inverse matrix
ε_w	White Gaussian noise
Ω	Standard deviation set
Δ_{sm}	Abundances in the subset m
ρ	Correlation coefficient
$\theta^*(ij)$	The j th coefficient of the data $\bar{y}_i$
ζ	Diagonal elements of the covariance matrix

δ	Value that minimizes the SURE estimator
ς	A general region in the image
ϱ	Modeling error
Φ	MNF transform matrix

Notational conventions

$\ \cdot\ $	Euclidean norm for vectors
$ \cdot $	Absolute value

Chapter 1

Introduction

Hyperspectral Imaging (HSI) analysis is the study of the interaction between matter and radiated energy. Hyperspectral studies are central to the development of quantum mechanics and include Max Planck's explanation of blackbody radiation, Albert Einstein's explanation of the photoelectric effect and Niels Bohr's explanation of atomic structure and spectra. Spectral analysis is used in physical and analytical chemistry because atoms and molecules have unique spectra. Spectroscopy, the first generation of spectral analysis techniques, was first used in chemistry to obtain information about atomic and molecular energy levels, molecular geometries, chemical bonds, interactions of molecules, and related processes. Historically, HSI originated from the study of visible light dispersed according to its wavelength, e.g., by a prism. Later, the concept was expanded greatly to include any interaction with radiative energy as a function of its wavelength or frequency. Spectroscopic data are often represented by a spectrum, a plot of the response of interest as a function of wavelength or frequency in multiple bands. With emerging of multispectral sensors which collect between 3 to 10 different band measurements in each image pixel, the application of spectral analysis was extended to other fields of science. The first multispectral photography from space occurred during the famous 1968 Apollo 9 mission. Scientists mounted four Hasselblad cameras in a holder, all aimed at the same target point when an astronaut triggered their shutters simultaneously. Astronauts took three filtered black and white photos of southern California around San Diego, in the green and red photo IR bands, and a (false) colour IR picture. Between May 1973 and February 1974, a six-camera multispectral system was mounted on the Skylab spacecraft. Photos taken showed the value of multispectral photography to scientists, particularly geologists, hydrologists, agronomists and those concerned with environmental monitoring and land use and cover assessment. The Landsat-D Earth-Observing satellite, Ikonos, was

launched on July 16, 1982, into a sun-synchronous orbit at a nominal altitude of 704 km. Ikonos was the world's first commercial satellite providing very high resolution (up to 1 m) imagery of the earth. It was followed by Quickbird and Spot satellites which were also using multispectral sensors. Multispectral sensors typically include the visible green, visible red and near infrared spectra. Landsat, Quickbird, and Spot satellites are well-known satellite sensors that use multispectral sensors. Advances in optics led to the production of hyperspectral sensors which could measure energy in narrower and more numerous bands than multispectral sensors. NASA's Hyperion was the first hyperspectral imager, launched in 2001. Hyperion was able to see the Earth in 220 spectral bands from visible to shortwave infrared with a 30-meter spatial resolution. The numerous narrow bands of hyperspectral sensors provide a continuous spectral measurement across the entire electromagnetic spectrum and therefore are more sensitive to subtle variations in reflected energy. Images produced from hyperspectral sensors contain much more data than images from multispectral sensors and have a greater potential to detect differences among the scanned objects. For example, multispectral imagery can be used to map forested areas, while hyperspectral imagery can map tree species within the forest. Thus, hyperspectral imagers started to be used by researchers for a variety of complex applications. Yet, many theoretical issues are left to be challenged to use hyperspectral imaging effectively in real applications. Existing algorithms are not clearly reflecting the improvements made in photonics and optics technology. New algorithms are needed to address this issue.

1.1 Hyperspectral Imaging Problem Formulation

1.1.1 Terminology

The following are the terms used in hyperspectral imaging technology:

Data Cube: The spatially and spectrally sample information of a hyperspectral image can be described by a three dimensional structure, referred to as a data cube as shown in Figure 1.1.

Spectral Signature: Characteristic of the reflected electromagnetic radiation as a function of the wavelength which can uniquely define an object is referred to as spectral signature. Figure 1.2 shows some sample spectral signatures related to mineral endmembers.

Endmember: The output data from hyperspectral sensor at a given spatial resolution and spectral band is a mixture of components originating from the constituent substances,

termed endmembers.

Abundance: Proportion of each endmember present in a pixel is defined by an abundance value.

Mixed Pixel: A mixed pixel is formed by combination of endmembers. The spectral signature of mixed pixels is the combined spectral signatures of the forming endmembers.

Pure Pixel: A pixel which includes only one endmember in its spectral signature.

Hyperspectral Unmixing: is the decomposition of the pixel spectrum into a collection of constituent spectra, or spectral signatures, and their corresponding abundances.

Figure 1.1: An example of a hyperspectral data cube and pixel radiance

Figure 1.2: Endmembers and spectral signatures for some minerals (USGS library [1])

1.2 Challenges of Hyperspectral Applications

Figure 1.3 (adopted from the Cytoviva Inc. [2]) shows how hyperspectral imaging has been extensively used in different fields of science and industry. In geology hyperspectral imaging is used for site exploration and mapping. In medical studies, HSI has been used to investigate the dynamics of nanoparticles in a cancerous tumor. In medical diagnosis applications, HSI can potentially eliminate the need for the first stage biopsy (second stage biopsy would be still needed for more details) due to its ability in detecting the decomposition of tumors. In pharmaceutical industry, HSI technology can be used to detect defective pills. In a medical diagnosis application accuracy is vital to correctly

Pharmaceutical
 •Blending Quality Control
 •Drug Discovery
 •Polymorph Analysis
 •Spray-Dry Dispersion

Life Science & Biotechnology
 •High Throughput Screening
 •Multi-Channel Spectroscopy
 •Nanobead & Quantum Dot detection

Forensic
 Crime Scene Investigation
 Counterfeit Detection
 Materials Identification
 Non-Invasive Latent Print Analysis

Space Research
 •Atmospheric Sciences
 •Environmental Monitoring
 •Remote Sensing

Process Manufacturing
 •LCD Quality Control
 •Semiconductor Operations
 •Pharmaceuticals

Military & Defense
 •Border Protection
 •Reconnaissance & Surveillance
 •Spectral Tagging

Figure 1.3: New applications of hyperspectral imaging in science and industry (Adopted from Cytoviva Inc. [2])

diagnose the disease before it reaches the critical stage. The quality control application of hyperspectral imaging in a production line also requires a very fast algorithm to separate defective products before they join the next batch. In all these application there is more demand for accuracy. For example, a biochemistry study demands an algorithm which

is robust to noise in order to eliminate the effect of illumination variations and noises in the microscope. As more and more data becomes available, processing time will also be of the higher importance.

The following are some existing challenges in hyperspectral imaging which are the focus of this thesis:

- **Denoising Efficiency** Existing hyperspectral imaging analysis methods utilize a denoising method to eliminate the noise effect on the signal. For example, in a microscopic application (e.g. detection of cancerous cells in tissue) the contribution of the desired object to be detected in the hyperspectral signal may be too small. In such a scenario, a preliminary denoising will eliminate not only the noise but also the desired object from the signal, leading to misdiagnosing of the endmember.
- **Dimension Estimation Accuracy** In some hyperspectral applications (e.g. polymorph analysis or nanobead quantum dot detection) dimension estimation plays a very important role in the process of unmixing. In the existing approaches, denoising step processes the data before the order selection step so the possible errors of denoising step propagates into the order selection process leading to an inaccurate dimension estimation.
- **Processing Speed Demand** Many applications of hyperspectral imaging (e.g. quality control in a pharmaceutical product line) demand a fast processing speed to detect the impurities or defects in a fairly short time. In another notes, some other applications (e.g. geological studies) even though may not require an online processing but since they deal with a very large volume of data, they still demand for a fast algorithm to process high volume data. The complexity of the existing methods has made them ineffective to address such demands.
- **Noise Structure Accuracy** In many hyperspectral studies, noise is considered to be additive Gaussian. Accurate estimation of the spectrum of colored noise in these cases is a challenge and requires huge matrix inversion.

1.3 Thesis Contribution and Outline

This thesis focuses on main components of HSI analysis: denoising, order selection and unmixing steps as shown in Figure 1.4. The following is a summary of thesis contributions.

Figure 1.4: HSI analysis diagram

- **A Hyperspectral Simultaneous Denoising and Intrinsic Order Selection Method (Chapter 3)** This chapter presents the main contribution of the research including the following important sections to explain the proposed methods. **UFU: A Fast Unmixing Method**

In this section a new least square based unmixing method is proposed. Simulations with synthetic data and real data are presented to evaluate and compare the accuracy of the method. It is shown that the method is faster than the existing approaches by the order of at least ten times.

MRE: An Order Selection Method This section proposes an order selection method based on the minimization of the reconstruction error defined by the MSE between the unavailable noiseless image and the available noisy image. The proposed method chooses the optimum order based on the estimate of the reconstruction error for different possible orders. The order selection in this method is intertwined with

the denoising goal simultaneously.

HYCE: A Hyperspectral Correlation Extractor Method In this section, the problem of the correlated noise in the hypercube has been addressed. A model to extract the noise correlation between neighboring bands is proposed and the mathematical calculations to extract the noise autocorrelation matrix is presented. Several simulations show the effectiveness of the proposed method.

Chapter 3 integrates the proposed order selection approach with the proposed unmixing method into a structure denoted by DIOS. Simulation results in this chapter compare the proposed method with the existing state of the art approaches to show the advantages in sense of accuracy and speed.

1.3.1 Thesis Outline

Chapter 1 is the introduction including HSI terminologies, thesis objectives and outlines. Chapter 2 is the background of HSI technology including the most commonly used hyperspectral models and methods, with their related drawbacks. Chapter 3 presents an order selection and denoising method along with a fast hyperspectral unmixing method to overcome the accuracy and speed issue in hyperspectral imaging applications. This chapter also addresses a proposed method for extracting the correlated noise structure in a hypercube. Chapter 4 concludes this thesis with suggestions for future work. A genetic algorithm based unmixing method is presented as a side study to this research in Appendix B. Appendix C briefly describes HyRap as a hyperspectral analysis software which has been developed based on this research.

Chapter 2

Background

Each pixel in the HSI image (hypercube) is considered to be a function of set of parameters denoted by endmembers and abundances through the mixing process. HSI analysis deals with restoring these parameters through denoising and unmixing process (Fig. 2.1).

Figure 2.1: Mixing and unmixing diagram

Figure 2.2 shows pure pixel and mixed pixel. In a pure pixel only one constituent material (endmember) is contributing while in a mixed pixel, combination of constituent materials are contributing to form that pixel.

2.1 Mixing Models

Mixing models in hyperspectral imaging analysis attempt to present the physics of the mixing phenomena. Based on the optical situation of the scanned object, different models have been suggested. We briefly overview all important models in next section.

2.1.1 Linear Mixture Model (LMM)

The basic assumption of linear mixing is that within the Instantaneous Field Of View (IFOV) of a single pixel, the surface is dominated by a few number of endmembers with

Figure 2.2: Pure pixel versus mixed pixel

relatively constant spectral signatures. Assuming that the hyperspectral image has ℓ spectral bands, the noiseless data in this model for the pixel (i, j) , is formulated as

$$\tilde{y}_{ij} = A_c S_{ij} = \sum_{\iota=1}^c a_{\iota} s_{ij\iota} \quad (2.1)$$

where the elements of $\tilde{y}_{ij} \in R^{\ell}$ are the original noiseless pixel vector. Total of c endmembers are considered as the constituent materials in the scene. Therefore, $A_c = [a_1 \ a_2 \ \dots \ a_c]$ is an $\ell \times c$ source matrix, with each column a_{ι} being the spectral signature of ι th endmember. LMM assumes that $\tilde{y}_{i,j}$ has been originally formed by combination of " c " endmembers which belong to a spectral library including c_{max} spectral signatures ($c < c_{max}$). The abundance vector for the pixel i, j , $S_{ij} = [s_{ij1} \ s_{ij2}, \dots, s_{ijc}]^T \in R^c$, consists of the mixing coefficients for the pixel (i, j) . The noisy measured hyperspectral signal, y_{ij} , for pixel (i, j) is given as

$$y_{ij} = \tilde{y}_{ij} + w_{ij} \quad (2.2)$$

The last term w_{ij} takes into account possible errors and sensor noises. Noise is considered to be additive with covariance matrix, $\Sigma_w = \sigma_w^2 I$, where I is the identity matrix. For simplicity, we concatenate the columns of datacube pixel-by-pixel to form a matrix of hyperdata. We use the notations, $y, \tilde{y}, w$ all $\in R^{\ell \times N}$ as the concatenated noise, noiseless

and noisy spectral data of all N pixels respectively defined as,

$$w = [w_1, w_2, \dots, w_N] \quad (2.3)$$

$$\tilde{y} = [\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_N] \quad (2.4)$$

$$y = [y_1, y_2, \dots, y_N] \quad (2.5)$$

It should be pointed out that together with the non-negative and sum-to-one constraints,

$$s_{ij\iota} \geq 0 \quad (2.6)$$

$$\sum_{\iota=1}^c s_{ij\iota} = 1 \quad (2.7)$$

the $\tilde{y}$ is a convex combination of endmember signatures; that is, in the $c - 1$ hyperspace, all the mixtures are within a simplex, whose vertices correspond to the endmembers.

2.1.2 Non-linear Mixing Model

In a non-linear mixing model the constituent materials are not organized proportionably on the surface of the scanned object. In this scenario, each endmember is randomly distributed in a homogenous way. As a result the incident light will experience scattering from different substances and therefore the aggregate spectrum of the reflected light may no longer be the linear proportion of the endmembers. A naive Non-Linear Mixture Model (NLMM) can be constructed by endmembers and their multiplications where it is assumed that the product of two or more endmembers can represent the multiple scattering effect [4]. The products after array multiplication are considered as new endmembers set. If only the scattering between two endmembers are considered, then the nonlinear endmember matrix A_{NL} becomes $[a_1, a_2, \dots, a_c, a_1a_2, \dots, a_{c-1}a_c]$. Hence, a pixel vector y_{ij} can be expressed as

$$y_{ij} = A_{NL}\alpha_{NL} + w_{ij} \quad (2.8)$$

where A_{NL} is the nonlinear form of the endmember matrix, α_{NL} is the related abundance vector and w is the noise term. For a large value of c , $2^c - 1$ nonlinear terms are added as endmembers to construct a pixel vector. The exponentially growth of columns of α_{NL} causes impractical complexity in hyperspectral applications.

2.1.3 Stochastic Model

The linear model in deterministic model considers the noise to be contributed in the signal as presented in (2.2). This implies that deterministic model considers the noise to be added to each component of image signal. The stochastic approach models each observation vector as a constrained linear combination of normally distributed random variables. Let c be the number of classes, and let $N(\mu_k, \Sigma_k), 1 \leq k \leq c$ denote the normal distribution with mean μ_k and covariance Σ_k . The model indicates that:

$$y_{ij} = \sum_{\iota=1}^c a_{\iota} s_{ij\iota} \quad (2.9)$$

such that:

$$a_{\iota} \sim N(\mu_{\iota}, \Sigma_{\iota}) \quad (2.10)$$

Non-negativity and sum-to-one constraint in (2.7) are applied. Figures 2.3 illustrates both deterministic and stochastic model. In deterministic model, endmembers are considered to be constant and noiseless over the scene while the noise is added to the linear combination of the endmembers. In stochastic model, each endmember belongs to a stochastic process with specific mean and covariance.

Figure 2.3: Left: Pixel constitution in deterministic, Right: Pixel constitution in stochastic approach

2.1.4 Mixing Model Used in Hyperspectral Applications

Size of the endmember matrix in a non-linear model causes the resulting unmixing process to be computationally very expensive and consequently unsuitable for practical application. Multiple scattering from three or more endmembers are usually negligible as the corresponding abundances are very small by the time the reflected light reaches the detector. Specifically, in cases where the area under study is large enough to damp the multiple scattering the assumption of linear mixing model is well supported. Geological surveys are examples of such scenarios. In many cases, even if the surface area is not smooth, some preparation methods is used to provide a proper sample that matches to the linear model (e.g. preparation of tissue sample for lab analysis). In pharmaceutical applications, the surface of the pills are smooth and therefore indirect scattering is eliminated. On the other hand, a stochastic mixing model requires either calculation of large covariance matrices (generally in the order of $N\ell \times N\ell$) or acquiring sample training data for classification purposes. In either case, the requirement are computationally very expensive or in some instances impossible. To our knowledge, although there are some developments of non-linear HSI modeling [5], [6], [7], no commercialized HSI analysis for this model is available. Therefore, primary focus of this thesis is on LMM as both it's research and application is well supported.

2.2 Hyperspectral Analysis Methods

2.2.1 Advancement in Hyperspectral Imaging Applications

An overview of some of research examples is presented here to show the advancements made through the years in HSI methods starting 2001. Research in 2001 [8] used hyperspectral techniques to estimate the green leaf area index and canopy chlorophyll density. Research published in Earth Observation magazine (2001) [9] searched for oil seeps and oil-impacted soil with hyperspectral imagery. In 2002, a study at McGill university [10] investigated the use of hyperspectral imaging in the impact of nitrogen and environmental conditions on corn. In the same year, at university of Idaho, hyperspectral imaging was used to detect spotted knapweed in forest sites [11]. Hyperspectral applications in 2003 continued to be mostly limited to environmental and geological studies. For example, in 2003 hyperspectral imaging was used for estimating estuarine and coastal water quality using data from Hyperion in Australia [12] and the Department of Geography at the University of Oregon studied high spatial resolution hyperspectral mapping of in-stream

habitats, depths, and woody debris in mountain streams [13]. Maximum likelihood supervised classification using principal component analysis was used in this study to achieve the mapping. In 2004, support vector machine was used to classify Indian Pines site (in Indiana, US) data acquired by the AVIRIS sensor [14]. In the same year, the potential of radiative transfer modelling and inversion techniques for operational uses was investigated to retrieve the leaf area index in a poplar plantation [15]. In 2005, more advanced techniques such as Kernel based analysis (e.g. [16]) were used for classification of hyperspectral data. Advanced optics technology led to increasing the number of spectral bands in hyperspectral sensors. This increase in the volume of hyperspectral data called for some more effective dimension reduction methods to facilitate some hyperspectral applications. Some morphological analysis [17] and projection based methods (e.g. [18]) were proposed to overcome the dimension reduction problem. In 2006, independent component analysis [19] was also proposed as a dimension reduction technique. Application of hyperspectral imaging in agricultural studies attracted attention in this year [20]. The year 2007 could be considered a starting point for hyperspectral imaging to be used in a new variety of applications beyond geological and environmental studies, urban mapping and agricultural studies. In [21], hyperspectral imaging of convective CO_2 ice clouds in the equatorial mesosphere of Mars was investigated. With the advent of hyperspectral fast applications in 2007, fast algorithms in hyperspectral analysis became essential. In [22], maximum entropy analysis was proposed as a fast algorithm to unmix hyperspectral data. Paper presented in 2008 [23] marked the advent of biological and medical applications of hyperspectral imaging. In this paper, an in-vivo hyperspectral confocal fluorescence imaging technique was presented to determine pigment localization and distribution in cyanobacterial cells. More advanced theories such as convex analysis [24] and a composite support vector machine [25] were exploited to support more complex applications such as the detection of citrus canker using hyperspectral image processing in [26]. Up to 2010, hyperspectral imaging was mainly focused on algorithms and commercial software. However, many practical issues related to hyperspectral sensors in photonics and optics were resolved by 2010. As a result, new fast hardware improvements have been developed. The last two years have seen the advent of high end hyperspectral equipment, such as the hyperspectral endoscope [27] and hyperspectral two-photon microscope, [28] to overcome the challenges in medical and biological applications. The imperative role of hyperspectral imaging in science and technology comes into view more and more as researchers explore this technique to solve variety of problems. In [28] a 3-D visualization of intrinsic contrast in neoplastic colon tissue was explored. In a 2010 survey [29],

the potential role of hyperspectral imaging in nanotechnology for cancer treatment was discussed.

The process of hyperspectral analysis is a concatenation of three distinct processes, denoising, order selection, unmixing and an optional dimension reduction stage. Next sections, briefly explain the existing methods for each step and their drawbacks.

2.2.2 Dimension Reduction Methods

Since hyperspectral scenes can possess extremely large volumes of data (e.g., 640 scan lines, 480 samples per line, 400 spectral bands), some unmixing algorithms first reduce the dimension of the data to minimize the corresponding computation. Not surprisingly, the familiar trade-off for representing data in a reduced dimension is a decrease in the accuracy of the intended application product. Dimension reduction step is optional and applies mostly to the applications where lose of information due to the dimension reduction is not a serious concern (e.g. site exploration and urban mapping). Principal Component Analysis (PCA) [30] is one of the most commonly used methods in dimension reduction of a hyperspectral signal. PCA is based on squared error, identifies orthogonal axes for dimension reduction by performing an eigen-decomposition of a covariance estimate of the data. Another statistical technique that optimizes SNR, is the Maximum Noise Fraction (MNF) [31]. MNF minimizes the ratio of noise energy to the received signal energy and requires estimation of the sample covariance matrix. The Noise-Adjusted Principal Components (NAPC) [32] transform formulates the problem, but achieves the mathematically equivalent answer as MNF. The rest of this sections explains the most commonly used dimension reduction methods.

Principal Component Analysis (PCA)

PCA is a standard method for deriving a new set of images with reduced spectral redundancy. Principles of PCA dimension reduction has been illustrated in Figures 2.4. This figure shows a random set of a 2D data on a scatter plot. Correlation factor between x1-data and x2-data is 0.9¹. In Fig. 2.4 orthogonal eigenvectors have been plotted on the zero-mean scatter plot. It can be noticed that one of the eigenvectors is inline with the

¹Correlation between two random variables V and W is defined as:

$$\rho_{v,w} = \frac{E(V - \mu_V)(W - \mu_W)}{\sigma_V \sigma_W}$$

where σ_V, σ_W denote the standard deviation and μ_V, μ_W denote expected values of V, W .

middle points and simulates the best fit line of the points on the graph. This eigenvector represents the correlation of data along this line. The second eigenvector provides the less important pattern in the data. Ignoring less dominant eigenvectors in the process of signal reconstruction usually won't cause a major loss of information. The process of calculating the eigenvectors from the sample covariance matrix is the underlying principle for PCA to characterize the data.

Figure 2.4: Left: Random correlated data set, Right: Zero-mean data set

The PCA process in hyperspectral imaging is a linear transformation that projects each image cells spectrum to a new set of orthogonal coordinate axes (axes formed by the set of eigenvectors). Since image data is projected onto the orthogonal basis, the transformed images are uncorrelated and can be ordered by decreasing variance, with the first principal component axis corresponding to the direction of maximum variance in spectral space. Image information is generally concentrated in the low-order components, while noise increases with increasing component number. Use of low-order PC rasters in place of original image bands can speed visual analysis and classification of the hyperspectral image. Using the notation of y in (2.5), the sample covariance matrix $\Sigma \in R^{\ell \times \ell}$ is given by

$$\Sigma = yy^T \quad (2.11)$$

The eigenvectors φ and eigenvalues λ of $\Sigma \in R_{\ell \times \ell}$ can be calculated as follow:

$$\Sigma\varphi = \lambda\varphi \quad (2.12)$$

Table 2.1 shows the eigenvalues, proportions and accumulations for an AVIRIS data [33]. Accumulation value in this table on each row is the sum of the proportions on previous rows. As we go toward the bottom of the table, accumulation values approach to 100%. It can be seen that most variations occur along a few eigenvectors. In fact, the first four eigenvalues can cover more than 90% of the signal's energy. Figure 2.5 shows how SNR changes as more eigenvectors(or eigenimages in case of a hyperspectral data cube) are used to reconstruct the signal. It can be seen that after 20 eigenimages the SNR rather saturates. Principal components analysis uses the above mentioned facts to ignore as many as possible principal components to achieve the dimension reduction goal.

Table 2.1: Eigenvalues of the Cuprite site in AVIRIS data along with the proportions and accumulations

	Eigenvalue	Proportion(%)	accumulation(%)
1	6.21×10^9	68	68
2	1.01×10^9	13	81
3	3.11×10^8	8	89
4	2.46×10^7	0.5	89.5
5	1.76×10^7	0.4	89.9
6	1.04×10^7	0.2	90.1
7	5.17×10^6	0.1	90.2
8	2.71×10^5	0.05	90.25

Wavelet PCA

Wavelets are an efficient and practical way to represent image information at multiple spatial scales. Image features at a given scale, such as houses or roads, can be directly enhanced by filtering the wavelet coefficients. For many tasks, wavelets may be a more useful image representation than pixels. Hence, PCA dimensionality reduction of wavelet coefficients can be considered in order to maximize information in the reduced dimensionality set of images [35]. Note that the wavelet transform will take place spatially over each image band, while the PCA transform will take place spectrally over the set of images. Thus, the two transforms operate over different domains. Still, PCA over a complete set

Figure 2.5: SNR versus the number of eigenimages for the Cuprite site data

of wavelet and approximation coefficients will result in exactly the same eigenspectra as PCA over the pixels. However, PCA over a subset of wavelet coefficients can be used to find eigenspectra that maximize the energy of that subset of wavelet coefficients. For example, PCA on only the vertical wavelet subbands will result in eigenspectra that maximize vertical wavelet energy. More generally, the term Wavelet PCA is used to refer to computing principal components for a masked or modified set of wavelet coefficients to find Wavelet PCA eigenspectra, and then projecting the original image onto the Wavelet PCA eigenspectra basis. In this way, features at a particular scale are indirectly emphasized by the computed projection basis, enhancing the reduced dimensionality images without filtering artifacts.

Drawbacks:

- PCA and Wavelet-PCA methods are sorting the principal components based on the variance. However, in applications where image quality is a concern, the image quality does not necessarily increase by having more principal components. Principal components with lower index may have better SNR ratio or quality compared to the principal components with higher indexes.

- To obtain the PCA transform, eigenvalue decomposition of sample covariance matrix is required. In many practical situations, these covariance matrices are unknown and need to be estimated. In another note, calculation of large covariance matrices (order of $\ell \times \ell$) is very time consuming. This is a limiting factor to use PCA for speed-demanding applications.

Maximum Noise Fraction Transform

PCA sorts the principal components based on the variance. However, in applications where image quality is a concern, the image quality does not necessarily increase by having more principal components. Principal components with lower index may have better SNR ratio or quality compared to the principal components with higher indexes. Figure 2.6 shows the first, fourth, sixth and ninth PC of a data acquired by AVIRIS. It can be seen that the ninth PC actually has a better quality compared to the lower order PCs. If such a data goes through a PC dimension reduction or PC denoising with a threshold to eliminate all the PCs lower than 9, the ninth PC will be eliminated during such a process leading to a lose of image quality and probably lose of useful information in the image. The

Figure 2.6: First, fourth, sixth and ninth PC

MNF is a modified version of the PCA that orders the output components by decreasing SNR. This method produces a component set in which noise levels increase uniformly with increasing component number. The low-order components should contain most of the image information and little image noise. Considering the notations of concatenated noisy and noiseless data in (2.5) and the sample covariance matrix in (2.11) and assuming an additive noise,

$$\Sigma = \Sigma_{\tilde{y}} + \Sigma_w \quad (2.13)$$

$$(2.14)$$

where $\Sigma_{\tilde{y}}$ and Σ_w are the covariance matrices of $\tilde{y}$ and w , respectively. The noise fraction of the k th band is defined as

$$\text{Var}\{W_k\}/\text{Var}\{Y_k\} \quad (2.15)$$

the ratio of the noise variance to the total variance for that band. Noise and signal random variables are noted by W and Y respectively. The Maximum Noise Fraction (MNF) transform chooses linear transformations

$$D_k = \Phi_k^T y, \quad k = 1, \dots, \ell \quad (2.16)$$

such that the noise fraction for D_k is maximum among all linear transformations orthogonal to $D_j, j = 1..k$. It can be shown that the vectors Φ_k are the left-hand eigenvectors of $\Sigma_w \Sigma^{-1}$, and that, μ_k , the eigenvalue corresponding to Φ_k , equals the noise fraction in D_k . Hence, from the definition of the MNF transform, it can be seen that $\mu_1 \geq \mu_2 \geq \dots \geq \mu_\ell$, and so the MNF components will show steadily increasing image quality (unlike the usual ordering of principal components). An important property of the MNF transform (not shared by principal components) is that, because it depends on SNR, it is invariant under scale changes to any band. Another useful property is that it orthogonalizes D and w as well as y .

Drawbacks:

To obtain the MNF transform, both Σ, Σ_w are needed. In many practical situations, these covariance matrices are unknown and need to be estimated. Usually, Σ is estimated using the sample covariance matrix of Y . The use of sample covariance matrix in hyperspectral signals is also questionable knowing that hyperspectral data is not necessarily stationary with spatial variation. Another drawback is that MNF accuracy is limited to the accuracy of noise variance estimation. Inaccurate estimation of the noise variance substantially degrades the validity of the calculated results.

2.2.3 Denoising Methods

Several techniques exist to denoise hyperspectral data on a band-by-band basis. Since each band is a two dimensional image, a Wiener filter might be considered to provide the minimum MSE estimate of an image from its noisy observation. However, the band-by-band denoising ignores the correlation between bands in hyperspectral application. A more feasible and useful denoising technique is wavelet thresholding or shrinkage, pi-

oneered by Donoho and Johnstone [36], [37]. In this method, a 2-D Discrete Wavelet Transform (DWT) is performed on the image followed by soft thresholding of the coefficients in the detail subbands. Atkinson et al. created a denoising technique that uses Discrete Fourier Transforms (DFTs), 2-D Discrete Wavelet Transforms (DWTs), and soft thresholding of wavelet coefficients to denoise hyperspectral imagery [38]. Singular Value Decomposition (SVD) denoising [39] and Wavelet thresholding have been commonly used in hyperspectral denoising where calculation of the sample correlation matrix or wavelet transform is feasible and computationally effective.

SVD Denoising

SVD denoising is implemented based on the following decomposition:

$$\hat{y} = \vartheta \vartheta^T y. \quad (2.17)$$

Matrix $\hat{y}$ is the estimated denoised signal and ϑ is the left singular vectors of $y y^T$. Using the unitary property of ϑ and considering all singular values in ϑ , (2.17) is just estimating the signal $\hat{y}$ by it's noisy version y . Denoising process starts with sorting the singular values in ϑ . Considering a threshold value can remove the small singular values of ϑ associated with noise. New ϑ and (2.17) can be used to accomplish the denoising scheme. Since in a hyperspectral image, noise elements contribute to the singular values $\sigma_k : k > c$, a partial SVD by ignoring all singular values $\sigma_k : k > c$ is considered a denoising scheme.

Wavelet Thresholding

The term wavelet thresholding is explained as decomposition of the data or the image into wavelet coefficients, comparing the detail coefficients with a given threshold value, and shrinking these coefficients close to zero to take away the effect of noise in the data. The image is reconstructed from the modified coefficients. This process is also known as the inverse discrete wavelet transform. During thresholding, a wavelet coefficient is compared to a given threshold and is set to zero if its magnitude is less than the threshold; otherwise, it is retained or modified depending on the threshold rule. Thresholding distinguishes between the coefficients due to noise and the ones consisting of important signal information. The choice of a threshold is an important point of interest. There exist various methods for wavelet thresholding, with different choice of threshold values. Some commonly used methods for image noise removal include VisuShrink [40], SureShrink [41] and BayesShrink [42]. This section, briefly explains these three common methods.

VisuShrink

VisuShrink uses a threshold value t_v that is proportional to the standard deviation of the noise. It follows the hard thresholding rule. It is also referred to as universal threshold and is defined as

$$t_v = \sigma_w \sqrt{2 \log(\eta)}. \quad (2.18)$$

σ_w^2 is the noise variance present in the signal and η represents the signal size or number of samples. An estimate of the noise level $\hat{\sigma}_w$ is usually by the Median Absolute Deviation (MAD),

$$\hat{\sigma}(\text{MAD}) = \frac{\text{Median}(|\theta|)}{0.6745}, \theta \in \text{subband HH}_1. \quad (2.19)$$

VisuShrink is known to yield recovered images that are overly smoothed as VisuShrink removes too many coefficients. Another disadvantage is that it cannot remove speckle noise available in some hyperspectral data and that it can only deal with an additive noise.

SureShrink

A threshold based on Steins Unbiased Risk Estimator (SURE) was proposed in [41] and is called as SureShrink. This method specifies a threshold value t_s for each resolution level ϵ in the wavelet transform which is referred to as level dependent thresholding. The goal of SureShrink is to minimize the mean squared error, ϵ , defined as

$$\epsilon = \frac{1}{\eta^2} \sum_{i,j=1}^{\eta} (\hat{y}_{ij} - \tilde{y}_{ij})^2, \quad (2.20)$$

where $\hat{y}_{ij}$ is the estimate of the signal while $\tilde{y}_{ij}$ is the original noiseless signal. SureShrink suppresses noise by thresholding the empirical wavelet coefficients. The SureShrink threshold t_s is

$$t_s = \min\{\delta, \sigma_w \sqrt{2 \log(\eta)}\}, \quad (2.21)$$

where δ denotes the value that minimizes the SURE estimator as shown in [41]. SureShrink follows the soft thresholding rule. The thresholding employed is adaptive, i.e., a threshold level is assigned to each dyadic resolution level by the principle of minimizing the Steins Unbiased Risk Estimator for threshold estimates. The smoothness is adaptive, meaning that if the unknown function contains abrupt changes or boundaries in the image, so does

the reconstructed image.

BayesShrink

This method is called BayesShrink as it minimizes the Bayesian risk. BayesShrink uses soft thresholding and is subband-dependent, which means that thresholding is done at each band of resolution in the wavelet decomposition. Like the SureShrink procedure, it is smoothness adaptive. The Bayes threshold, t_B , is defined as

$$t_b = \frac{\sigma_w^2}{\sigma_{\tilde{y}}}, \quad (2.22)$$

where σ_w^2 is the noise variance and $\sigma_{\tilde{y}}$ is the noiseless signal variance. The noise variance σ_w^2 is estimated from the subband HH1 by the median estimator in (2.19). From the definition of additive noise and assuming the independency between signal and noise

$$\sigma_y^2 = \sigma_{\tilde{y}}^2 + \sigma_w^2. \quad (2.23)$$

σ_y^2 can be computed as

$$\sigma_y^2 = \frac{1}{\eta^2} \sum_{i,j=1}^{\eta} y_{ij}^2. \quad (2.24)$$

The variance of the noiseless signal, $\sigma_{\tilde{y}}^2$, is computed as

$$\sigma_{\tilde{y}} = \sqrt{\max(\sigma_y^2 - \sigma_w^2, 0)}. \quad (2.25)$$

With $\sigma_{\tilde{y}}$ and σ_w , the Bayes threshold is computed from (2.22). Using this threshold, the wavelet coefficients are thresholded at each band to achieve the denoising scheme.

Drawbacks:

- SVD is based on defining a threshold to remove the components associated with the lower singular values. The accuracy of the method highly depends on the accuracy of the selected threshold. Yet, an optimum threshold for the case of hyperspectral images has not been proposed.
- Calculation of large sample correlation matrix to calculate the singular values is questionable in a hyperspectral application with large volume of data.
- Wavelet based methods are based on defining a threshold. Sensitivity of the methods to this threshold for hyperspectral applications is high. In hyperspectral applications where accurate estimation of the order is critical (e.g. impurity detection) utilizing such

methods leads to an inaccurate results.

2.2.4 Order Selection Methods

Estimation of the number of signal sources in a Hyperdata cube is very challenging. According to the definition, the effective dimensionality, is the minimum number of parameters required to account for the observed properties of the data. It is difficult to determine the effective dimensionality of hyperspectral data in practice. This is mainly because effective dimensionality cannot be simply determined by the dimensionality of a data sample vector, referred to as component dimensionality(which is defined by the number of components in a data vector). For hyperspectral data, the effective dimensionality is in general much smaller than the component dimensionality due to the high-dimensional structure of data cube. Several methods have been proposed such as principal components analysis (PCA) and factor analysis [43] which make use of the eigenvalue distribution to determine the effective dimensionality. These approaches were basically developed for multispectral imagery with small limited number of bands where component dimensionality is comparable to the effective dimensionality. Also the application of sample correlation matrix is questionable knowing that hyperspectral data is not necessarily stationary with spatial variation. Some other SVD based dimension estimation methods are also in some cases inefficient since 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 c (c is the number of endmembers in the signal) singular vectors nor by any other set of eigenvalues. A NeymanPearson detection theory-based eigen-threshold method, referred to as the HFC method [44], was previously developed to determine the number of endmembers in hyperspectral data. Consequently, method of Virtual Dimensionality (VD) [45] was proposed that used HFC as the base. VD is defined as the minimum number of spectrally distinct signal sources that characterizes the hyperspectral data from the perspective view of target detection and classification rather than the image endmembers, which are idealized pure signatures. In [46], an eigen-decomposition based, unsupervised order selection method has been proposed. The method first estimates the signal and noise correlation matrices and then selects the subset of eigenvalues that best represents the signal subspace in the least squared error sense. Most of the hyperspectral order selection methods use an eigenvalue analysis of the sample correlation matrix to define the order of the signal. Next section, explains VD as the most commonly used order selection method in hyperspectral imaging analysis.

Virtual Dimensionality

VD first calculates the sample correlation matrix $R_{L \times L}$ and sample covariance matrix $K_{\ell \times \ell}$ and then finds the difference between their corresponding eigenvalues. Let $\hat{\lambda}_1 > \hat{\lambda}_2 > \dots > \hat{\lambda}_\ell$ and $\lambda_1 > \lambda_2 > \dots > \lambda_\ell$ be two sets of eigenvalues generated by $R_{\ell \times \ell}$ and $K_{\ell \times \ell}$, called correlation eigenvalues and covariance eigenvalues, respectively. By assuming that signal sources are non-random unknown positive constants and noise is white with zero mean, we can expect that

$$\hat{\lambda}_\ell > \lambda_\ell \quad \text{for } l = 1, \dots, VD \quad (2.26)$$

$$\hat{\lambda}_\ell = \lambda_\ell \quad \text{for } l = VD + 1, \dots, \ell \quad (2.27)$$

More specifically, the eigenvalues in the l th spectral channel can be related by

$$\hat{\lambda}_\ell > \lambda_\ell > \sigma_{nw}^2 \quad \text{for } l = 1, \dots, VD \quad (2.28)$$

$$\hat{\lambda}_\ell = \lambda_\ell = \sigma_{nw}^2 \quad \text{for } l = VD + 1, \dots, \ell \quad (2.29)$$

where σ_{wl}^2 is the noise variance in the l th spectral band. The relation between eigenvalues of correlation and covariance matrix is shown in Figure 2.7 for $SNR = 10$ where number of endmembers was identical to 6. It can be seen from the figure that both eigenvalues from correlation matrix and covariance matrix suddenly drop to a fixed level on the point matched to the original number of endmembers. This dropped level is identical to the noise variance. This behavior can be explained by considering spectral values of each pixel as a random variable vector that is resulted from random linear combination of endmembers. This assumption implies that the image matrix is consisting of different columns each representing an observation of a sample random variable. As a result, the eigenvalues of sample correlation matrix is an indication of the contribution of endmembers and noise. In order to determine the VD , [44] formulated the problem of determination of VD as a binary hypothesis problem:

$$H_0 : \hat{\lambda}_l - \lambda_l = 0 \quad (2.30)$$

$$H_1 : \hat{\lambda}_l - \lambda_l > 0, \quad \text{for } l = 1, 2, \dots, \ell \quad (2.31)$$

where the null hypothesis H_0 and the alternative hypothesis H_1 represent the case that the correlation eigenvalue is equal to its corresponding covariance eigenvalue and the case that

the correlation eigenvalue is greater than its corresponding covariance eigenvalue, respectively. In other words, when H_1 is true implies that there is an endmember contributing to the correlation eigenvalue in addition to noise, since the noise energy represented by the eigenvalue of R in that particular component is the same as the one represented by the eigenvalue of K in its corresponding component.

Figure 2.7: The relation between eigenvalues of correlation and covariance matrix for SNR=10 and $c=6$

2.2.5 Hyperspectral Unmixing Methods

Based on the mixing models explained in the Section 2.1, hyperspectral unmixing process is defined as the reversed mixing process to find the spectral signatures of endmembers and the related abundances. Most well known hyperspectral unmixing methods are explained in next sections.

Independent Component Analysis

Independent Component Analysis (ICA) [19] has received considerable interest in signal processing due to its versatile applications ranging from source separation, channel equalization to speech recognition and functional magnetic resonance imaging. The key idea of the ICA assumes that data are linearly mixed by a set of separate independent sources and demix these signal sources according to their statistical independency measured by

mutual information. In order to validate this approach, an underlying assumption is that at most one source in the mixture model can be allowed to be a Gaussian source. This is due to the fact that a linear mixture of Gaussian sources is still a Gaussian source. The criteria used by the PCA and the MNF are data variance and SNR which are designed to measure data second-order statistics. ICA uses mutual information as a criterion to measure data statistical independency that exceeds second-order statistics. As a result, the ICA can capture information that cannot be retained or preserved by second-order statistics-based techniques. If the original source signals have been mixed linearly, and the mixed signals are available, then ICA finds, in a blind manner, a linear combination of the mixed signals which recovers the original source signals, possibly rescaled. The starting point for ICA is the assumptions that the components are statistically independent, and that they must have non-Gaussian distributions. ICA estimates the endmember matrix A in (2.2) and its inverse, A^{-1} and further calculates the independent components S (abundance vectors):

$$S = A^{-1}y \quad (2.32)$$

But, it is not possible to determine neither the covariances, the signs nor the order of the independent components, because both of A and S are unknown. Many ICA algorithms are available. A computationally efficient ICA algorithm, called the FastICA [47], is usually used in hyperspectral imaging. The FastICA algorithm requires that the data must be of zero-mean and whitened, to be able to function correctly.

Drawbacks:

- ICA assumes the independency of the sources in the signal. Such an assumption considering the linear combination of the endmembers to create the hyperspectral signal is questionable.
- Very large number of the hyperspectral pixels in the image, generates a giant covariance matrix, which leads to a very slow processing unsuitable for fast hyperspectral applications.
- The sum-to-one constraint of abundance fractions implies the statistical dependency among them. This dependency, compromises ICA applicability to the hyperspectral images.

Vertex Component Analysis

The Vertex Component Analysis (VCA) algorithm [18] is an unsupervised method that exploits two facts: 1) the endmembers are the vertices of a simplex and 2) the affine transformation of a simplex is also a simplex. It works with unprojected and with projected data. VCA also assumes the presence of pure pixels in the data. The 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 performs better than PCA or ICA as it has a computational complexity between one and two orders of magnitude lower than these methods.

Drawbacks:

The use of image-derived spectral endmembers in VCA cannot accurately characterize intimate spectral mixtures due to the effect of mixed pixels. VCA may provide accurate results for a low mixing grade application, however for a high degree of mixing level arising from the spatial resolution limitation or the natural mixing of endmembers, VCA is lacking accuracy to address many hyperspectral applications.

Maximum Entropy Method

In maximum entropy method, an unsupervised decomposition method has been proposed based on the classic maximum entropy principle, termed as the Gradient Descent Maximum Entropy (GDME). The method addresses the importance of the maximum entropy principle for mixed-pixel decomposition from a geometric point of view. This method considers the LMM model (2.2) for the mixing of endmembers. Given the model in (2.2) and assuming that the signature matrix is known a priori, the problem of spectral unmixing becomes a constrained linear regression problem. One popular solution to such a problem is the fully constrained least squares problem [48]. The algorithm is developed from the Nonnegative Constrained Least Squares (NCLS) method [48] in conjunction with the sum-to-one constraint. The objective is to minimize the least-squares error. The maximum entropy principle is applied to estimate abundance fractions. The algorithm maximizes the formulated relative entropy using the quadratic penalty function method. The GDME method proposes to minimize the following objective function:

$$f_0(s) = \sum_{j=1}^c s_j \ln s_j \quad (2.33)$$

$$\text{subject to : } h_0(s) = 1^T s - 1 = 0, \quad (2.34)$$

$$h_i(s) = \sum_{j=1}^c a_{ij}s_{ij} - y_i = 0, i = 1, \dots, N \quad (2.35)$$

Using the method of Lagrange multipliers, GDME transforms the original constrained minimization problem to an unconstrained problem with the following objective function:

$$\mathcal{L}(s, \xi, \xi_0) = f_0(s) + \xi^T h(s) + (\xi_0 - 1)h_0(s) \quad (2.36)$$

where ξ, ξ_0 are the Lagrange multipliers and $h(s)$ is a column vector with the i th element being $h_i(s)$. The Lagrange function in 2.36 is minimized with respect to s, ξ, ξ_0 . However, as derived in the following, these three parameters are not independent. Both the abundance vector s and the multiplier ξ_0 can be expressed as a function of ξ . Thus, the minimization problem can be tailored to the finding of the optimal ξ . The connection between the abundance vector s and the multiplier ξ is given by the formula given in [22]:

$$s_j = \frac{1}{1 + \sum_{k=1, k \neq j}^c \exp(a_j^T \xi - a_k^T \xi)} \quad (2.37)$$

where a_j is the j th column of A . With known A , the abundance solution is uniquely determined by the Lagrange multiplier ξ , whose learning strategy thus becomes the key to the GDME algorithm. Traditional gradient descent method used in GDME leads to the following learning rule:

$$\xi^{k+1} = \xi^k + \eta^k (A s^k - y) \quad (2.38)$$

where η^k is a small step size, and denotes the iteration index. The GDME algorithm iteratively finds all endmembers and abundances.

Drawbacks:

The main drawbacks to the GDME method are the high sensitivity of the algorithm to existence of the pure pixel as well as the computational complexity. GDME can not be

used for any fast application of the hyperspectral imaging. The simulation results also can prove that GDME is not applicable to the noisy applications.

Stochastic Unmixing

In the stochastic model [49], initial estimates of the class means are obtained by applying deterministic linear unmixing techniques to determine a set of endmembers. Several methods of estimating endmembers are available. Clusters containing a prescribed number of points near each endmember are defined, and an initial estimate of each class covariance is obtained as the sample covariance of the corresponding cluster and then updating process is performed as follows:

Updating Abundance Estimates For given parameters $(\mu_k, \Sigma_k), 1 \leq k \leq d$ and given abundances $\alpha = (a_1, \dots, a_d)$ let (dropping the pixel indices):

$$\mu(\alpha) = \sum_{k=1}^d a_k \mu_k \quad (2.39)$$

$$\Sigma(\alpha) = \sum_{k=1}^d a_k^2 \Sigma_k \quad (2.40)$$

Then, $y_i \sim N(\mu(\alpha^i), \Sigma(\alpha^i))$ Maximum likelihood abundance estimates are obtained by solving :

$$\hat{\alpha}^i = \arg\{ \max \left(\frac{1}{|\Sigma(\alpha^i)|^{0.5} (2\pi)^{n/2}} \exp[-0.5(y_i - \mu(\alpha^i))' \Sigma(\alpha^i)^{-1} (y_i - \mu(\alpha^i))] \right) \} \quad (2.41)$$

subject to the sum-to-one and non-negativity constraints for abundances.

Updating Class Parameters

It has been shown in [49] that for given abundance estimates, the class parameters, $\Theta(\mu_k, \Sigma_k), 1 \leq k \leq d$ may be estimated by applying the expectation-maximization algorithm [49].

Support Vector Machine

One problem often noted in the classification of hyperspectral data is the Hughes effect [50]. The Hughes phenomenon has been observed in many remote sensing studies. The key characteristics of the phenomenon may be illustrated for a typical scenario in which features are incrementally added to a classification analysis. Initially, classification ac-

curacy increases with the addition of new features. The rate of increase in accuracy, however, declines, and eventually, accuracy will begin to decrease as more features are included. Thus, the addition of features may lead to a reduction in classification accuracy. For example, a parametric technique, such as the maximum likelihood classifier, may not be able to classify a data set accurately if the ratio of sample size to number of features is small, as it will not be able to correctly estimate the first and second-order statistics that are fundamental to the analysis. Given that training data acquisition may be difficult and costly, some means to accommodate the negative issues associated with high-dimensional data sets are required. Support Vector Machine (SVM) method has been proposed as a method which is insensitive to the Hughes effect. SVM covers the linear and non-linear classification categories. An example of SVM is presented here. Suppose that our goal is to classify the scattered points shown in Fig. 2.8. We have

$$\omega x_i + b \geq 1 \text{ if } x_i = 1 \quad (2.42)$$

$$\omega x_i + b \leq -1 \text{ if } x_i = -1 \quad (2.43)$$

where ω, b are the slope and bias of the classifier line. As a result

$$x_i(\omega x_i + b) \geq 1 \quad (2.44)$$

Margin is defined as : $M = 2 / |w|$. It is required to maximize the margin which is equal to minimizing the $\frac{1}{2}w \cdot w^*$. This will lead to a quadratic optimization problem as

$$\phi(\omega) = \frac{1}{2} \omega \cdot \omega^* \quad (2.45)$$

$$\text{Subject to : } x_i(\omega x_i + b) \geq 1 \quad (2.46)$$

Following two steps can extend the SVM to a multi-class option:

Step 1: In m -dimension, learn m SVMs:

SVM 1: learns output==1 vs output != 1

SVM 2: learns output==2 vs output != 2

..

SVM m : learns output== m vs output != m

Step 2: To predict the output for a new input, just predict with each SVM and find out which one puts the prediction the furthest into the positive region.

Drawbacks:

- SVM is very sensitive to noise. In fact, a relatively small number of mislabeled examples

Figure 2.8: Linear classifier

can dramatically decrease the performance.

- The accuracy of the classification by SVM does vary as a function of the number of features used. This highlights the dependency of the accuracy of classification by SVM on the dimensionality of the data and, therefore, the potential value of undertaking a feature-selection analysis prior to classification.
- SVM in general, is limited to a two-class classification problem explained in this section. The extension of SVM to a multi-class version even though is possible but demands a high computational power.

Chapter 3

Simultaneous Denoising and Intrinsic Order Selection (DIOS) Method

3.1 Introduction

Order selection in a noisy hyperspectral application has been always a concern in existing hyperspectral methods. Conventional hyperspectral methods divide unmixing process into two separate processes of order selection and unmixing. Order selection methods generally use a denoising approach at the beginning stage. The data in this case pass through three stages: denoising, order selection and unmixing. Each of these steps optimize a different criterion independently. In addition, any error created in the denoising process will be propagated not only to the order selection stage but also consequently to the unmixing results. Note that both the denoising step and the dimension estimation step aim to provide the optimum estimate of the same noiseless data. Consequently, adopting a simultaneous denoising and dimension estimation method with the goal to provide the optimum estimate of the desired noiseless data is rational. This process not only avoids possible error propagations from the denoising stage to the dimension estimation stage, but also unifies the optimization criteria that were used in each of these steps. In this chapter, a simultaneous denoising and dimension estimation method is introduced. The approach is based on minimizing the estimated MSE. Minimization is done by comparing the estimated data in a range of subsets dictated by a simultaneous process. Minimizing the error at once, the proposed method denoises the data and provides the optimum dimension simultaneously. This chapter organizes as follows:

Section 3.2 presents a fast unmixing process to estimate the abundance vectors and end-members. Section 3.3 explains the proposed order selection method to estimate the hy-

perspectival dimension. Section 3.5 includes the proposed denoising method to consider the existence of colored noise in some hyperspectral applications.

3.2 Ultra-Fast Unmixing (UFU) Method

The UFU method is a projection-based unmixing method to find the data estimate in subsets of different order. Based on the LMM model, the UFU uses the fact that the noiseless part of the sensor measurement y_i is actually a convex combination of endmember signatures. This indicates that in the $c-1$ hyperspace (c points form a $c-1$ dimension space), all the mixtures are within a simplex, whose vertices correspond to the endmembers given in the LMM model. To clarify the concept of endmember transformation, Fig. 3.1 has been provided which considers a noiseless signal $\tilde{y}_1$ obtained from (2.1) with $c = 3$. The LSE solution for abundance vectors is

$$\begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} = T_3 y_1$$

, where

$$T_3 = (A_3' A_3)^{-1} A_3' \quad , \quad A_3 = [a_1 \ a_2 \ a_3]. \quad (3.1)$$

In real applications, the noisy version of y_1 (a couple of samples are shown in the figure by stars in Fig. 3.1) is projected into the space spanned by the endmembers and the LSE estimate of the data is used. In UFU, the process of finding a transformation matrix is performed iteratively. Note that A_3 is a subset of the spectrum library, A_c , which includes 3 endmembers of the spectrum library. The product of the transform matrix and the i th pixel data y_i would result in the estimated abundance vector for that pixel. Retrieving the abundance factors from UFU needs the applicability of the “condition of identifiability.” The condition basically means that the number of linear independent endmembers cannot exceed the number of linear independent bands. Multispectral imaging, as the earlier version of hyperspectral imaging, couldn’t satisfy this condition due to lack of enough spectral bands. Considering the improvements recently made in hyperspectral imaging systems, the number of spectral bands is usually greater than the number of components to be identified. Therefore, the condition of identifiability is mostly met and a transformation matrix always exists regardless of the structure of the hyperspectral sensor. Initialization

plays a very important role in any unmixing method. In [51] the impact of initialization on design of endmember extraction algorithms has been discussed. In this section, UFU is adopted as the initializing and unmixing approach. UFU follows an iterative approach in two steps to unmix data. The first step initiates the process by calculating the first two endmembers and abundances. The second step is an iterative step to calculate the remained endmembers and abundances up to the number of endmembers estimated by the order selection method (which will be proposed in later sections).

Figure 3.1: A 3-D model of a hyperspectral data vector. $S = [s_{11} \ s_{12} \ s_{13}]$

3.2.1 Step 1: Least Square Error (LSE) Based Extraction Method

In order to unmix the given hyperspectral data cube, this section proposes an iterative LSE-based method to maximize the fitness between the sensor measurement and the estimation. The algorithm starts with certain initial endmembers, then finds the most possible abundance distributions of all image pixels based on the constrained least squares. A new endmember is selected as the pixel that generates the largest residual. The method starts from initialization of the first two endmembers ($m = 1, 2$). As part of the initialization step, the first endmember is selected as the pixel with the largest magnitude; that

is

$$k_1 = \arg \max_i \|y_i\|_2 \quad \hat{a}_1 = y_{k_1}. \quad (3.2)$$

where the operator $\|\cdot\|$ is the Euclidean norm and $\hat{a}_1$ is the first estimated endmember. The conventional reason for choosing this pixel as the first endmember is that no convex combination can yield a vector that is longer than the individual components; thus it must correspond to one of the purest pixels [22]. The second endmember is chosen as the pixel that is the most distinct from $\hat{a}_1$ based on the Euclidean distance measure:

$$k_2 = \arg \max_i \|\hat{a}_1 - y_i\|_2 \quad \hat{a}_2 = y_{k_2}. \quad (3.3)$$

This will provide the estimate of the endmember matrix with the first two elements, $\hat{A}_2$:

$$\hat{A}_2 = [\hat{a}_1 \ \hat{a}_2]. \quad (3.4)$$

The abundance vector estimate associated to this estimate of the endmembers is denoted by $S_{i,2}$, where i , is the i th pixel and 2 indicates that the first two endmember estimates are used. Using the LSE solution of the LMM model, we can define the transformation matrix T_2 for the subset that is spanned by the first two endmember estimates:

$$S_{i,2} = T_2 y_i \quad (3.5)$$

$$T_2 = (\hat{A}_2' \hat{A}_2)^{-1} \hat{A}_2'. \quad (3.6)$$

Transformation matrix T_m for the m th subset has the property of finding the abundance vectors for each pixel by multiplying with the same pixel vector such that

$$S_{i,m} = T_m y_i. \quad (3.7)$$

Therefore, the abundance vectors for the first subset ($m=2$) is

$$S_{i,2} = T_2 y_i. \quad (3.8)$$

3.2.2 Step 2: Iterative Selection of Endmembers ($m > 2$)

Given the two initial endmembers, the UFU algorithm is applied to all pixels to find the abundances related to each pixel using a double-LSE based method. In each subset m , the

estimated endmember matrix A_m is calculated by augmenting the previous endmember matrix. The estimate of the endmember matrix in the m th subset is in form of

$$\hat{A}_m = [\hat{A}_{m-1} \hat{a}_m], \quad (3.9)$$

where the new endmember $\hat{a}_m$ is selected from the pixels yielding the largest residual (first LSE solution). The maximum residual based endmember extraction is implemented as

$$i^* = \arg \max_{i \in 1 \dots N} \| y_i - \hat{A}_{m-1} S_{i,m-1} \|_2 \quad (3.10)$$

$$\hat{a}_m = y_{i^*}. \quad (3.11)$$

Next the associated abundance vector is calculated by using the projection procedure. In this step, the transformation matrix, T_m , is extracted from the second LSE solution:

$$S_{i,m} = \arg \min_x \| x - \hat{A}_m \|_2. \quad (3.12)$$

where

$$S_{i,m} = T_m y_i. \quad (3.13)$$

$$T_m = (\hat{A}_m' \hat{A}_m)^{-1} \hat{A}_m'. \quad (3.14)$$

Calculation of the endmembers and abundance vectors is continued iteratively for the considered maximum number of endmembers. The optimum choice of a subset of these selected endmembers will be provided in the next section.

Suppressing the Noise Effect: In the presence of a strong noise, selection of a single pixel with the largest residual as the next endmember in (3.10), will lead to an inaccurate estimation. In this case, using the averaging scheme to suppress the severe effect of noise is essential. The averaging scheme calculates an average of the pixels with largest residual and then the pixel matching the average is selected as the next endmember. To detect this situation, the UFU approach first estimates the SNR ¹. If SNR is higher than a predetermined threshold, the procedure in (3.10) is followed ². Otherwise, the method

¹ The SNR estimate is calculated by

$$SNR = 10 \log_{10} \frac{\sum_{i=1}^N y_i^T y_i}{N \hat{\sigma}_w^2},$$

where $\hat{\sigma}_w^2$ is the estimate of the noise variance

²The threshold needs to match the SNR of acceptable noise level in hyperspectral imaging applications,

takes the average of a set of pixels with the largest LSEs as the new endmember. The last 10 pixels are averaged with the largest residual obtained by (3.10). Consequently, the noise effect is successfully suppressed when the SNR is low and the smoothness effect is avoided when the SNR is high.

3.2.3 Spectral Library Matching Approach

UFU provides an estimate of the endmembers. In case the endmember reference library set is available, it is possible to find the exact endmember match. This final step further improves the endmember estimation. In this case, in each subset an estimated endmember is replaced with an improved version from the spectral library:

$$\hat{a}_i \rightarrow \text{improved}(\hat{a}_i). \quad (3.15)$$

It is proposed to use a new spectral library matching approach based on the spectral matching of the endmembers to improve the error performance (root mean squared error between the actual noiseless data and the noisy data). The process of finding the exact matched endmembers is performed by comparing the identified endmembers with the spectral library data, based on their spectral correlations. Using the estimation of the spectral correlation in [14], the library spectrum with the highest correlation is selected as the improved version of the estimated endmember:

$$\rho_{a_i, \hat{a}_j} = \frac{(\hat{a}_j - \hat{\varpi})^T (a_i - \varpi)}{\|\hat{a}_j - \hat{\varpi}\|_2 \|a_i - \varpi\|_2} \quad (3.16)$$

$$\text{improved}(\hat{a}_i) = \arg \max_{a_j, 1 \leq j \leq c_{max}} \rho_{a_j, \hat{a}_i}. \quad (3.17)$$

where $\hat{a}_j, a_i$ correspond to the estimated spectrum and the library spectrum of each endmember respectively. The mean value of all estimated endmember spectra and all library spectra, denoted by $\hat{\varpi} = \frac{1}{c_{max}} \sum_{i=1}^{c_{max}} \hat{a}_i$ and $\varpi = \frac{1}{c_{max}} \sum_{i=1}^{c_{max}} a_i$, is subtracted from the original spectra to give more accurate correlation coefficients. UFU approach is summarized in Algorithm 1.

for example in the range of 20dB.

Algorithm 1: UFU Algorithm

Data: Mixture data $y \in \Re^{\ell \times N}$ (N is the number of pixels, ℓ is the number of spectral bands).

Result: The source matrix $A \in \Re^{\ell \times c}$ and the abundance matrix $S \in \Re^{c \times N}$ (c is the number of endmembers)

//Initialization

$p = 2$; $A^{(p)} \leftarrow (a_1 \ a_2)$, where a_1 and a_2 are chosen based on (3.2)

//Main loop

while $p \leq c$ do

 Set $S^{(p)}$ as a $p \times N$ zero matrix

 Calculate transformation matrix $T^{(p)}$ from (3.14)

 //Calculation of residuals

for every pixel y_i in the image **do**

 Calculate $s^{(i)} = T^{(p)} y_i$

 Calculate the residual of pixel y_i by

 calculating the $\|y_i - A^{(p)} s^{(i)}\|_2$

 Save $s^{(i)}$ as the i th column of matrix $S^{(p)}$

end

 //LSE-based endmember selection

 Identify a new endmember $a^{(p+1)}$ using the

 LSE-based method and include it into $A^{(p+1)}$

 Increase p by 1

end

for $k \leq c$ **do**

 Calculate the spectral correlation between a_k

 and library spectral signatures from (3.17)

 save the highest correlated signature in the

 library as a_k in A

end

Output: New abundances $S^{(c)} = T^{(c)} y$

$A \leftarrow A^{(c)}$, $S \leftarrow S^{(c)}$

3.2.4 Simulation Results for UFU

The spectral reflectance used in the subsequent experiments are selected from the USGS digital spectral library which contains 224 spectral bands covering wavelengths ranging from 0.38 μ m to 2.5 μ m. A set of four spectral profiles are selected as the endmembers to create the mixture as shown in Figure 3.2 (indicated as the true EM). To generate linear mixtures, positive abundance vectors were randomly selected and were multiplied to the spectral endmembers followed by addition of a Gaussian noise. The resulting image is then degraded by a spatial $k \times k$ average filter to produce mixed pixels (k controls the degree of mixing). With a small k , only the pixels close to the block boundary are mixed, so the mixture data are very likely to contain pure pixels. The first simulation was performed by selecting a SNR value of 10dB. The simulation results are shown in 3.2 for SNR=10dB. As it can be seen, all four spectral endmembers were successfully extracted from the image. To show the abundance values, a gray scale mapping was selected. The brightness of the pixels is proportional to the abundances values. The gray value comparison of the true abundances for each endmember and the estimated ones showed a high accuracy in the estimation process.

To get an idea of the error value in UFU method, MSE error was calculated for all N pixels between the noiseless data and the estimated data as $\|y_i - \hat{A}\hat{s}_i\|$, where $\hat{s}_i$ is the i th estimated abundance vector. The error was equal to 0.012. In next step, the noise performance was tested for improvement. Calculations showed that the proposed improvement approach lowered the MSE error down to 0.010 which indicates %15 improvement in the total error.

Figure 3.3 is showing the line of true and estimated abundances for all four endmembers in 32 different pixels existed in a row for SNR=10dB. Table 3.1 shows the averaged MSE error for this estimation. Even though, VCA is claimed to be one of the most accurate existing methods, a comparison of the abundance estimation errors showed that UFU was an improvement not only in speed but also in accuracy.

3.2.5 A Note on Speed Efficiency of UFU

Assuming that the number of endmembers is known, the optimum value of c is available. The UFU method iteratively repeats the unmixing steps until all c endmembers are extracted. Since the transformation matrix is calculated only once for all pixels in each iteration and since UFU is not a convergence-based algorithm, the projection-based unmixing process is much faster than the existing unmixing methods. To demonstrate

Figure 3.2: UFU estimation of endmembers and abundances (SNR=10dB). MSE data is presented in Table 3.1 and Fig. 3.3)

Table 3.1: Averaged MSE error of abundance estimates for UFU and VCA

Method	VCA	UFU
Endmember 1	0.024	0.019
Endmember 2	0.042	0.024
Endmember 3	0.016	0.009
Endmember 4	0.034	0.022

Figure 3.3: Line of true and estimated abundances for 32 pixels in a row of Fig. 3.2(SNR=10dB)

the speed efficiency of the UFU approach, UFU is compared to the VCA and Gradient Descend Maximum Entropy [22] approach as well as the Fully Constrained Least Squares (FCLS) method in [48]. These methods are among the fastest unmixing methods used in hyperspectral imaging. In each case 10 endmembers were selected from the USGS library to create the synthetic data. For each run different SNR values of 5 to 30 dB were selected. Each method was used to unmix the created synthetic data and the processing speed in MATLAB was used as a comparison. It can be seen in Fig. 3.4 that processing time

has been improved using the projection approach in the UFU method. It is also evident from the figure that processing speed is almost constant as the SNR values varies from 5 to 30 dB. The UFU processing time, compared to that of the VCA, FCLS, and GDME, is on the scale of 5, 30 and 60 times lower. To analyze the computational complexity of

Figure 3.4: A comparison of the processing time (using MATLAB implementation with dimension $c = 10$) for different unmixing methods

UFU, different size of the data cube was used as the input to UFU when the number of endmembers were 5 and $\text{SNR} = 5\text{dB}$. In each scenario UFU was repeated hundred times to show the consistency of the experiment. The variance of processing times for each data size was calculated and the 2σ intervals were plotted. It can be seen from Fig. 3.5 that UFU processing speed is changing quasi-linearly with the size of data cube in the range of 0.2-2MB. The processing speed changed by 0.2 second when the size of data cube multiplied 10 times from 0.2MB to 2MB. Figure 3.6 shows the comparison of the processing speed for UFU, VCA, GDME and FCLS when the size of data cube is changing. It can be seen that only UFU and VCA are showing a linearity in the processing speed profile. GDME seems to follow an exponential trend. The fastest method used in comparison is VCA. For a data cube size of 2MB VCA can be as high as 4 seconds compared to the value for UFU which is 0.3 second. UFU shows an improvement of 10 times better than the fastest method.

Figure 3.7 shows the averaged processing time of the UFU method for different number of sources (endmembers). SNR was fixed at 5dB with a data cube size of 2MB as number of endmembers used to generate the synthetic data, was changed from 3 to 7. Lines for

Figure 3.5: Averaged processing time and 2σ intervals versus data cube size (using MATLAB implementation with $c=5$, SNR=5dB) for 100 runs

Figure 3.6: Comparison of processing time versus data cube size (using MATLAB implementation with $c=5$, SNR=5dB) for UFU, VCA, GDME and FCLS

2σ intervals have been plotted. It can be seen that UFU is exhibiting a linear trend with variation of the endmembers when number of endmembers is higher than 4. This is mainly because the algorithm follows a different process to extract the first two endmembers. As the number of endmembers increases, the time to extract the first two endmembers

will be shadowed by the time spent to extract other endmembers leading to a similar processing time for extracting each endmember. Figure 3.8 compares the processing

Figure 3.7: Averaged processing time versus number of endmembers with 2σ intervals (using MATLAB implementation with datacube size=2MB, SNR=5dB) for UFU

speed versus number of endmembers for UFU, VCA, GDME and FCLS. All compared methods are more or less exhibiting a linear behavior in this graph. However, the most competitive method, VCA, seems to be following a decreasing speed trend as the number of endmembers increases. For $c = 3$, VCA is about 10 times slower than UFU while for $c = 7$ it tends to be about 15 times slower. UFU speed is also decreasing linearly with number of endmembers, however since other methods are much slower than UFU the graph for this method is very close to the horizontal axis to follow the graph scale.

3.2.6 A Note on the Computational Complexity of UFU

Calculation of the transformation matrix $T^{(p)}$ (p is the loop index) in Algorithm 1 is the most time consuming term. Calculation of this matrix requires an inversion of a $p \times p$ matrix and two multiplications of $p \times \ell$ by $\ell \times p$ and $p \times p$ by $p \times \ell$. Loop parameter p increases from 2 to c . The complexity of inversion part and multiplication part is defined as $O(p^3)$ individually. The total complexity resulted from the calculation of transformation matrix for a complete loop is $O(\sum_{p=2}^c 2p^3)$. Considering the calculation of residuals for N pixels, the complexity of algorithm can be calculated as $O(2c^3N)$. Performing the same calculation for the VCA algorithm listed in [18] results in the complexity of $O(3c^3N + N^3)$ for this method which confirms the superiority of UFU over VCA.

Figure 3.8: Comparison of processing time versus number of endmembers (using MATLAB implementation with datacube size=2MB, SNR=5dB) for UFU, VCA, GDME and FCLS

3.3 Hyperspectral Order Selection Method

This section proposes Minimum Reconstruction Error (MRE) method as an order selection method for hyperspectral applications. The proposed method addresses the problems in noisy hyperspectral applications. MRE is based on estimation of reconstruction error in different subsets to find the subset with minimum error. Next section defines this error and mathematical formulas to find the minimum subset.

3.3.1 Data Error versus Reconstruction Error

Assuming that UFU has already provided the estimate of the endmember matrix, the estimate of noiseless data for the i th pixel in each subset of order m ($1 \leq m \leq C$), denoted by $\hat{y}_{i,m}$ is

$$\hat{A}_m = [\hat{a}_1 \hat{a}_2 \cdots \hat{a}_m] \quad (3.18)$$

$$\hat{y}_{i,m} = \sum_{j=1}^m \hat{a}_j \hat{s}_{ij}, \quad (3.19)$$

where $\hat{a}_i$ and $\hat{s}_{ij}$ are the estimated endmembers and abundances, respectively. The dimension estimation problem is the process of finding a subset Υ_m (and therefore $\hat{y}_m \in \Upsilon_m$)

that best represents the noiseless data. In [52]³ it was suggested to focus on estimating the desired Reconstruction Error (RE) of each subset in form of

$$z_m = \frac{1}{N} \sum_{i=1}^N \|\bar{y}_i - \hat{y}_{i,m}\|_2^2. \quad (3.20)$$

and the concept of data error for the same subset

$$x_m = \frac{1}{N} \sum_{i=1}^N \|y_i - \hat{y}_{i,m}\|_2^2. \quad (3.21)$$

Data error plays an important role in estimating the desired reconstruction error. The (noisy) data error, x_m , is the distance between the noisy observed data and its estimate in subset Υ_m . Since both these components are available, this error is also available for each subset. A typical behavior of this error is shown in Fig. 3.10. While in the absence of noise this error is the same as the reconstruction error, the data error is monotonically decreasing with the variation of the subset even in the presence of noise. It simply indicates that the more endmembers is fitted to the noisy image, the smaller will be the error between the *noisy* image and the estimate. The hidden cost is that the estimate is getting farther from the desired *noiseless* image.

Note that for the data error, since the comparison of the estimate is with the *noisy* data itself, as more and more endmembers are used, the error becomes smaller regardless of the intensity of noise. This performance is known as the Maximum Likelihood problem which does not allow us to use this error for the order selection. The available data error, however, can be used in estimating the desired reconstruction error. In the absence of the noise, the problem of the order selection is straight forward. However, noise can cause the data to be represented in a higher dimension compared to the dimension of the original noiseless data. Consequently, a denoising scheme could affect the result of the order selection stage. Thus, the problem of denoising and order selection seems to be quite correlated and hence a simultaneous denoising and order selection method seems to

³In [52] and [53], the considered subset selection problem is as follows: Given the observed *noisy* data that has been generated linearly by a matrix transformation of a parametric vector, what is the best estimate of that vector? The main challenge in this setting arises after estimating the parametric vector in subsets of different order. The question is then how to find the optimum order, i.e, how many of the estimated parameters best fit the data. The problem is solved for when the transformation matrix is orthogonal in [52] and is generalized for nonorthogonal matrices in [53]. The problem considered in this paper has an identical setting with matrix transformation of each pixel in the form of (2.1) and the noisy available data in the form of (2.2).

Figure 3.9: Reconstruction error in the presence or absence of noise as a function of m , number of endmembers

Figure 3.10: Data error in presence or absence of noise as a function of the number of endmembers

be inevitable.

3.3.2 MRE Calculation

The proposed MRE method considers the reflectance data on the i th pixel, $y_i \in R^\ell$, as a random vector corrupted by an additive white Gaussian noise w_i . The noisy data y_i of

length ℓ for the i th pixel is available,

$$y_i = \tilde{y}_i + w_i, \quad (3.22)$$

where $\tilde{y}_i$ is the noiseless pixel vector $i = 1, 2, \dots, N$. The noiseless data is represented by this basis as follows:

$$\bar{y}_i^\ell = \sum_{j=1}^{\ell} \theta_{ij}^* s_j, \quad (3.23)$$

where $\bar{y}_i^\ell = [\tilde{y}_i(1), \tilde{y}_i(2), \dots, \tilde{y}_i(\ell)]^T$ and $\theta^*(ij)$ is the j th coefficient of the noiseless data $\bar{y}_i^\ell$. Consider S_m , a subset including only m endmembers which is spanned by m elements of the basis. The estimate of noiseless data, $\bar{y}_i^\ell$, in this subset using only the first m coefficients, $\hat{\theta}_{m,i}$, is

$$\hat{y}_{m,i} = \sum_{i=1}^m \hat{\theta}_{m,i} s_i. \quad (3.24)$$

For example, if we plan to use only the first $m = 7$ estimated endmembers, then the estimate of the i th image pixel is $\hat{y}_{i,7}$ which is obtained by setting $m = 7$ in (3.24). This will result in an error between the i th pixel of the desired noiseless data $\bar{y}_i$ and that of the estimated pixel. The average l_2 norm of this error over all the pixels is the RE for setting $m = 7$ in (3.20). If this error for all the subsets is available, the subset with smallest RE is the optimum subset to represent the data. It is important to mention that this optimum order that minimizes the RE is the result of a denoising process. To illustrate this fact, Fig. 3.9 shows the general behavior of the RE in the presence or absence of the additive noise. In the absence of noise, the error will decrease as more endmembers are used to describe the data. In fact the abundances of the endmembers with smaller effect become negligible and do not affect the error. On the other hand, in the presence of noise, the estimates of all the abundances will be noisy. Therefore, the consistent presence of noise in all these values causes forces a tradeoff between choosing more and more endmembers and noise fitting. Adding more endmembers at the beginning of the process causes the RE to decrease. As the absolute value of the abundance members associated with the chosen endmembers decreases, the noise dominant error in these values causes the reconstruction error to start increasing. Consequently, the use of reconstruction error for subset comparison illustrates the inevitable intertwined procedure of denoising and order selection. The main challenge in the simultaneous order selection and denoising

approach is estimating the RE by using *only* the available noisy image. Based on the Parseval's theorem, the coefficient error $z_{m,i} = \|\theta_i^* - \hat{\theta}_{m,i}\|_2^2$ for the m th subset and the i th pixel is the same as the reconstruction error for the same subset and pixel:

$$z_{m,i} = \|\bar{y}_i^\ell - \hat{y}_{m,i}^\ell\|_2^2 = \|\theta_i^* - \hat{\theta}_{m,i}\|_2^2. \quad (3.25)$$

We define the subset coefficient error for all pixels in the m th subset as

$$z_m = \frac{1}{N} \sum_{i=1}^N z_{m,i}. \quad (3.26)$$

For subset S_m , (3.22) can be rewritten as follows

$$y_i = \begin{bmatrix} A_m & B_m \end{bmatrix} \begin{bmatrix} \theta_{m,i}^* \\ \Delta_{S_m,i} \end{bmatrix} + w_i, \quad (3.27)$$

where B_m includes the endmembers not in $A_m = [a_1, a_2, \dots, a_m]$. Coefficients $\theta_{m,i}^*$ are the abundances related to the endmembers in the subset m for the i th pixel and the coefficients $\Delta_{S_m,i}$ are the abundances related to the endmembers in endmember complete set of $A_m = [a_1, a_2, \dots, a_C]$ which are not listed in the S_m . The subset coefficient error z_m can be expressed as a function of the basis vectors, additive noise, and the noiseless coefficients:

$$z_m = \frac{1}{N} \sum_{i=1}^N (\|A_m^H w_i\|_2^2 + \|\Delta_{S_m,i}\|_2^2). \quad (3.28)$$

3.3.3 MRE Calculation for Hyperspectral Images

In a hyperspectral image, noise vector $w_i \in R^\ell$ is a Gaussian random vector. Since z_m is the summation of m Gaussian random vectors over N pixels, random vector Z_m has a Chi-square distribution (as defined in Appendix A) of order mN such that

$$\sum_{i=1}^N \|A_m w_i\|_2^2 \sim \sigma_\omega^2 \chi_{mN}^2, \quad (3.29)$$

where $w_i \in R^\ell$ is the noise vector for the i th pixel and σ_ω^2 is the variance of noise for that pixel. (3.29) can be rewritten as

$$\frac{N}{\sigma_\omega^2}(Z_m - \sum_{i=1}^N \|\Delta_{S_{m,i}}\|_2^2) \sim \chi_{mN}^2. \quad (3.30)$$

Knowing the properties of a Chi-square distribution for the mean and variance

$$\mathbb{E}(Z_m) = \frac{mN\sigma_\omega^2}{N} + \sum_{i=1}^N \|\Delta_{S_{m,i}}\|_2^2 \quad (3.31)$$

$$\text{Var}(Z_m) = \frac{2mN}{N^2}\sigma_\omega^4, \quad (3.32)$$

which is simplified as

$$\mathbb{E}(Z_m) = m\sigma_\omega^2 + \sum_{i=1}^N \|\Delta_{S_{m,i}}\|_2^2 \quad (3.33)$$

$$\text{Var}(Z_m) = \frac{2m}{N}\sigma_\omega^4. \quad (3.34)$$

From the definition of the data error 3.21 and using 3.27

$$x_m = \frac{1}{N} \sum_{i=1}^N \|B_m^H w_i\|_2^2 + \frac{1}{N} \|\Delta_{S_{m,i}}\|_2^2. \quad (3.35)$$

Since B_m has the remaining $C - m$ endmembers included, x_m has a Chi-square distribution of $(C - m)\ell N$ such that

$$\frac{N}{\sigma_\omega^2}(x_m - \sum_{i=1}^N \|\Delta_{S_{m,i}}\|_2^2) \sim \chi_{(C-m)\ell N}^2. \quad (3.36)$$

Taking the expected value of the x_m and using the mean and variance properties of the Chi-square distribution,

$$\mathbb{E}(X_m) = (C - m)\sigma_\omega^2 + \sum_{i=1}^N \|\Delta_{S_{m,i}}\|_2^2 \quad (3.37)$$

$$\text{Var}(X_m) = \frac{2(C - m)}{N}\sigma_\omega^4. \quad (3.38)$$

Note that both Z_{S_m} and X_{S_m} are Chi-square random variables and their structure is such that the available sample of x_m can be used in estimating bounds on the unavailable sample of Z_m through the Chi-square look up table. However, a Chi-square distribution is well estimated by a Gaussian distribution through the Central Limit Theorem, for the sum (order) as low as 10. Therefore, for a wide range of orders we have the advantage of finding the following mathematical expressions for the bounds on Z_m without using the Chi-square lookup table. Two errors z_m and x_m are samples of two random variables Z_m and X_m and the relation between these two random variables is provided as

$$\mathbb{E}(Z_m) = \mathbb{E}(X_m) - (C - 2m)\sigma_w^2 \text{Var}(Z_m) = \frac{2(C - m)}{N}\sigma_w^4, \quad (3.39)$$

and with confidence probability $cp = Q(\beta)$ and validation probability $vp = Q(\alpha)$ the bounds on z_m are

$$\begin{aligned} \overline{z_m(Q(\beta), y, Q(\alpha))} = \\ m\sigma_w^2 + L_m(y, \alpha) - \beta \frac{\sqrt{2(C - m)}}{N}\sigma_w^2 \end{aligned} \quad (3.40)$$

$$\begin{aligned} \underline{z_m(Q(\beta), y, Q(\alpha))} = \\ m\sigma_w^2 + U_m(y, \alpha) + \beta \frac{\sqrt{2(C - m)}}{N}\sigma_w^2, \end{aligned} \quad (3.41)$$

where $L_m(y, \alpha)$ and $U_m(y, \alpha)$ are provided in Appendix A, and $Q(\alpha) = \int_{-\alpha}^{\alpha} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$. The optimum order is then chosen based on the probabilistic worst case behavior of z_m . With a certain probability z_m is bounded with the upper and lower limits. Therefore, “the worst case” probabilistic bound is the upper bound which is the provided probabilistic upper bound in (3.41). Therefore, the optimum order that minimizes this criterion is

$$\hat{c}(\sigma_w) = m^* = \arg \min_m \{ \overline{z_m(Q(\beta), y, Q(\alpha))} \}. \quad (3.42)$$

The range of parameters α and β is a function of data length as provided in [52]. These selection criteria for α and β guarantee that the upper and lower bounds are within reasonable distance of each other. If we consider $\alpha = 0$ and $\beta = 0$, then the two bounds will be overlapping, meaning that the effect of variance change for different subsets has been ignored⁴. Extensive simulation results show that the method is very robust to the variations of these two values of probabilities for $vp = Q(\alpha) > 0.9$ and $cp = Q(\beta) > 0.9$.

⁴Setting α and β to zero the estimate of RE becomes the Akaike Information Criterion(AIC) [54]. Relationship between this criterion and the method used in MRE is provided in [52]

In fact observations indicate not only that the method is robust for choosing this range of probability, but that it also works optimally within this range.

3.3.4 Noise Variance Estimation

Estimation of noise variance is an essential component in MRE method. Noise variance is not known in a hyperspectral application. A variety of methods have been proposed to model and measure variance of the random noise in digital images. Some of them rely on the assumption of an additive white noise, others employ homogeneous image areas for measuring the noise variance by means of statistical estimation [55]. The problem with direct estimation methods is that they must be supervised, requiring prior knowledge about homogeneous regions. In [56] a Homogeneous Regions Division and Spectral De-Correlation method (HRDSDC) is proposed to automatically estimate the SNR in hyperspectral images. This method is based on the internal regularity of earth objects and their strong spectral correlations and can be applied to heterogeneous reflectance and radiance images. To improve the noise estimation in this method, selection of a homogeneous area must be carried out. This, however, is not always possible. The Residual Autocorrelation Power (RAP) [57] method is the focus of MRE to estimate the noise variance. RAP and the Regression method [55] (as a common noise variance estimation approach) are explained in Appendix C.

3.3.5 Non-orthogonality and the Estimation Error

In this subsection, effect of independency and orthogonality of the endmembers on estimation error is explained. Test data for this study was based on the USGS library. Preliminary calculation of the rank of the endmember matrix showed that endmembers were independent. To be in safe side, the singular values of the endmembers matrix were calculated using the SVD method. The lowest singular value indicated a full rank matrix of endmembers. This implied that endmembers in USGS library were independent. Later studies confirmed that in general the endmembers in different spectral libraries are independent. However, further calculations confirmed that endmembers were not orthogonal. If endmembers were to be orthogonal, the minimum point of the reconstruction error would be matching the original number of constructing endmembers. To confirm this claim, in a later study, Gram-Schmidt method was used to orthogonalize the endmembers in the USGS library to synthetically generate the hyperspectral data. The following stabilized GramSchmidt process was used:

Algorithm 2: Gram-Schmidt Algorithm

```

for  $j$  from 1 to  $k$  do
    for  $i$  from 1 to  $j - 1$  do
         $v_j \leftarrow A_j - \langle A_j, A_i \rangle A_i$  (remove component in direction  $A_i$ )
    end for
     $v_j \leftarrow \frac{A_j}{\|A_j\|}$  (normalize)
end for

```

The new orthogonal matrix was used to build the mixed data as an input to the order selection process. Six endmembers were used to produce the mixed data. Further, the final RMS reconstruction error was calculated considering that the number of extracted endmembers was changing from 2 to 10. The simulation result for SNR=10 is shown in Figures 3.11.

Figure 3.11: Error versus number of endmembers for SNR=10 and applying orthogonalization

Figure 3.11 is indicating that orthogonalization has a major effect on the behavior of final error specially after number of extracted EMs becomes greater than the original number of components. This can be a key factor to find out the optimum number of endmembers. The point of reaching to a break point is the point of interest for optimum endmembers. This is coming from the fact that orthogonalization removes the effect of noise fitting that was observed before in error graphs. Figure 3.12 provides a geometrical representation for this behavior. For simplicity, a scenario with two primary orthogonal endmembers was considered. The combinational result, x_1 is supposed to be on the dashed line connecting a_1 and a_2 together but due to the presence of noise it may fall distant to this point. The

noisy vector is shown in a light gray color in Figure 3.12(a). When we underestimate the number of endmembers by one, we will have a considerable error as shown in Figure 3.12(b). When we consider the exact number of endmembers (in this case 2) then the error will be decreased to the level represented only by the error from the noise. This is shown as a component that is orthogonal to the plane formed by two endmembers (as shown in Figure 3.12(c)). If we overestimate the endmembers by one, then the combinational vector will fall on a convex which is covering also the regions filled by the noise. This will make the combination result to be closer to the original pixel. The error difference compared to the case with two endmembers will be equal to $cf - df = cd$ as shown in 3.12(d). This indicates a lower error in case of the exact number of endmembers.

Figure 3.12: Geometric representation of error due to the orthogonalization

Further, SVD analysis was performed on the hyperspectral data for the different cases of orthogonal and non-orthogonal endmembers. When endmembers are orthogonal (in this example, there are six orthogonal vectors), the sixth *SVD* singular value suddenly drops to a very low number to indicate the rank of combinational data. As for the non-orthogonal endmembers this singular value is still considerable as the original endmember

Figure 3.13: Error graph for SNR=10 and applying SVD and orthogonalization

matrix is not full rank (which implies a slight dependency in endmembers). Calculations showed a ratio of 100 : 1 between two singular values for these two scenarios. This explains how noise is contributed to the signal even after applying *SVD* to the noisy signal. As a result, the section of the error graph after the breaking point only represents the noise contribution when the endmembers are orthogonal. This further leads to observing a very sluggish trend after this point in the error graph as shown in Figure 3.13.

3.3.6 Simulation Results for MRE

In this subsection, the RAP algorithm is applied to synthetic and real data. Results are compared to the HFC and Noise-Whitened HFC (NWHFC) detectors [45] and the Hyperspectral Subspace Identification (HYSIME) [46] methods. As concluded in [45], the first two algorithms outperform the information theoretical criteria approaches making them proper candidates for comparison purposes. The HYSIME algorithm, as the most recent rank estimation method, was used for comparison as well. The spectral signatures were selected from the USGS digital spectral library. The input data was a simulated hyperspectral image composed of 10^4 spectral vectors, each one following the linear mixing model. The abundance fractions were generated according to a Dirichlet distribution. The noise was zero-mean, independent with variances along the bands, following a Gaussian shape. The parameter σ^2 was defined by the desired SNR. The results presented here are organized into the following four experiments.

Experiment 1: RAP analysis was performed on synthetic data. Δ_{S_m} and $m\sigma^2$ were cal-

culated and plotted (m is the subset index). As it can be seen in Figure 3.14, the sum of these two components has a minimum at some value of m . This summation represents the reconstruction error. According to (3.42), this minimum defines the index of the subset which is the best representation of the signal. In other words, the minimum on this figure represents the rank estimation of data. In Figure 3.14 the minimum reconstruction error has been marked with an arrow. As expected the arrow is pointing to $\hat{c} = m^* = 7$, which is a correct estimation of the number of original endmembers. SNR in this experiment was equal to 5, which matches a very noisy hyperspectral application. The simulation result implied that RAP was an accurate intrinsic order estimation method for low SNR hyperspectral applications.

Figure 3.14: Rank estimation for SNR=5, c = number of constituent endmembers=7

Experiment 2: To evaluate the robustness of the RAP method to the noise, four different SNRs, 5, 3, 2, 0, were considered. These SNRs matched the ones in any noisy hyperspectral application. As shown in Figure 3.15, in each case RAP estimated the rank of data accurately. Even in the presence of extremely powerful noise (SNR=0), the estimated dimensionality for c =Number of constituent endmembers=7 was accurate. Therefore, it can be claimed with confidence that RAP is a robust rank estimation method in regard to extremely noisy hyperspectral applications.

Experiment 3: In this experiment the accuracy of RAP for different numbers of con-

Figure 3.15: Rank estimation for SNR=5, 3, 2, 0

stituent endmembers (parameter c) was evaluated. Synthetic data was created using c values equal to 6, 8, 10, 12. The value of SNR was fixed at 5 for this experiment. As shown in Figure 3.16, RAP successfully estimated the number of constituent endmembers with excellent accuracy. It can be seen in the figure that $\|\Delta_{S_m}\|$ has a plateau for $m \geq m^*$. As the number of constituent signals increases, this plateau becomes flatter after the optimum point. Also as the number of constituent endmembers grows, the error will increase. For $c = 12$, estimated dimensionality is equal to 11.

Experiment 4: Evaluation with real data

The proposed method was applied to the real hyperspectral data collected by AVIRIS. A subset of the Indian Pines test site in northwestern Indiana, acquired by AVIRIS in June 1992, was considered. The data set was composed of 185 spectral bands with 10-nm

Figure 3.16: Rank estimation for different numbers of constituent endmembers

bandwidth acquired in the $0.4 - 2.5\mu\text{m}$ region. The data set contained 145×145 pixels (21025 pixels) with a ground resolution of 17m . This observed region contains a mixture of agriculture and forestry. Total of sixteen substances have been listed for this site in USGS library.

Table 3.2 shows the signal subset dimension inferred by HYSIM, HFC, NWHFC, and RAP. Parameter p_f is the false-alarm probability for the HFC and NWHFC methods. This probability was used in a series of Neyman-Pearson tests, each one designed to detect a different orthogonal signal subset direction. The false-alarm probability p_f of each test was specified. A lower value of p_f in each method indicates a lower possibility of error. As it can be seen in Table 3.2, the value obtained by RAP at low SNR values almost coincides with the number of ground truth materials listed in USGS library. While

for all other methods the error was between 18% and 62%, the error for RAP was only 6%. It is interesting to see that even for a very low false-alarm probability in both HFC and NWHFC, RAP still outperforms both methods. It can be seen from the table that MRE using regression or RAP have the same estimation of the rank order. Further results showed that the MSE error obtained by MRE when regression was used as the variance estimation method, was 17.2 while the error obtained by RAP was 13.1. This implies that even though the rank estimation was the same for both methods RAP estimates the endmembers matrix and the abundance vectors 20% more accurately than regression estimate.

Table 3.2: Comparison of rank estimation for SNR=10 and c=16

Method	$\hat{p}$
HFC($P_f = 10^{-3}$)	26
HFC($P_f = 10^{-4}$)	23
HFC($P_f = 10^{-5}$)	19
NWHFC($P_f = 10^{-3}$)	20
NWHFC($P_f = 10^{-4}$)	19
NWHFC($P_f = 10^{-5}$)	19
HYSIME	13
Regression	15
RAP	15

To show the effectiveness of RAP over the Multiple Regression method [55], Table 3.3 is provided. Table 3.3 implies that in high SNR applications (e.g. SNR>20) the noise variance estimation provided by RAP was substantially accurate (error<1%). In these applications, the estimation provided by the Multiple Regression method was less accurate than that of RAP but still in an acceptable range (error<7%) and very close to RAP. As SNR decreases, the estimation provided by RAP remains accurate unlike the results from the Multiple Regression method which are far from accurate. In the case of SNR=2, the estimation error from the Multiple Regression method was 37%, while in the same scenario RAP reported an error of only 2%. The results proved that RAP was a better candidate for estimation of noise variance in low SNR applications.

Table 3.3: Comparison of the Multiple Regression method and RAP noise estimation

SNR	True noise variance $\times 10^{-4}$	RAP estimation $\times 10^{-4}$	Regression estimation $\times 10^{-4}$
20	445	444	412
15	774	761	590
10	1342	1347	987
8	1661	1665	1180
5	2387	2383	1588
2	3391	3316	2119

3.4 Simultaneous Denoising and Order Selection Method

Based on explanations in the introduction section and to address the requirements of hyperspectral applications, this thesis proposes the DIOS method. DIOS analysis structure is sequence of consecutive stages as shown in Fig. 3.17. The analysis structure starts with the hyperspectral data originating from the output of hyperspectral sensors. Two major processing blocks, the MRE and the UFU blocks, communicate with each other to form the structure of the DIOS. The UFU method was primitively introduced in [58] as an unmixing method. Here the UFU is used as the built-in unmixing part of the order selection method. UFU block calculates the endmembers and abundances in an iterative approach to project the hyperspectral data into subsets of different dimensions. The MSE estimate between the noiseless data and the estimated data in different subsets are then calculated for all subsets. For MSE estimation, the proposed method in [52] and [53] was adopted. The MSE estimates are used by the MRE block for dimension estimation and denoising purposes. The MRE block finds the optimum subset through MSE minimization. The optimum subset represents the denoised data and the order of this subset is the estimated intrinsic dimension of the data. Note that denoising and intrinsic dimension estimation are simultaneous processes in DIOS. This is illustrated in Fig. 3.17 with reciprocal arrows between these two blocks. The parallel processing of denoising and dimension estimation is handled by the proposed approach. DIOS algorithm including the integration of MRE and UFU is explained in Algorithm 3.

Algorithm 3: DIOS Algorithm

Input: Mixture data $y = [y_1 \ y_2 \ .. \ y_N] \in \mathbb{R}^{\ell \times N}$ (N is the number of pixels).

Output: Estimates of The number of distinct endmembers (hyperspectral data dimension) $\hat{c}$, endmember matrix $\hat{A}$, Abundance vectors $\hat{S}_i (i = 1..N)$.

- 1) Estimate the noise variance in (C.3)
- 2) Initialize a subset with two endmembers ($m = 2$) and calculate the estimates of these two first endmembers in (3.2)-(3.6)
- 3) Calculate the residuals for each pixel and select the next endmembers in (3.9)-(3.11)
- 4) Find the improved version of endmembers from spectral library (if available) in (2.11)
- 5) Calculate transformation matrix and abundances in (3.12)-(3.14)
- 6) Calculate the lower bound L_m and upper bound U_m , in (3.41) and (3.40)
- 7) Increase m and repeat from step 3 until U_m reaches a minimum
- 8) Calculate $\hat{c} = m^*$ in (3.42)
- 9) Calculate estimated endmember matrix $\hat{A} = [a_1, a_2, .., a_m]$
- 10) Calculate estimated abundance vectors in 3.7

3.4.1 Simulation Results for DIOS

In this subsection, two dimension estimation examples are provided. In the first example, the synthetic data was generated using the spectrum library and a set of random numbers. The second example uses the real data collected by AVIRIS.

Experiment 1: Synthetic Data The spectral reflectance used in the subsequent experiments was selected from the USGS digital spectral library which contains 224 spectral bands covering wavelengths ranging from 0.38 μ m to 2.5 μ m. A set of spectral profiles was selected as the endmembers to create the mixture. To create linear mixtures, a set of abundance vectors were randomly selected and multiplied to spectral endmembers followed by addition of a Gaussian noise. Abundance fractions, s_{ij} , were generated according to a Dirichlet distribution [59]. Generating pure pixels in the image or eliminating the chance of having any pure pixels in data was controllable in this experiment. We had also the option of using a spatial $k \times k$ averaging filter to degrade and mix the synthetic data. The mixing degree was controlled by parameter k . With a small k , only the pixels close to the block boundary were mixed, so the mixture data were very likely to contain

Figure 3.17: DIOS analysis diagram

pure pixels. For simulation purposes, endmembers from the USGS library were used to create the synthetic data. Mixing parameter k was used equal to 7, which indicated a fair degree of mixing and guaranteed that no pure pixel was available in simulation results. The number of pixels N was 3364 which indicates an image size of 58×58 in each band. To investigate the robustness of the system to the number of constituent components in the signal, four different sets of hyperspectral data were created. The true dimension (number of endmembers) was selected as $c = 12, 13, 15$ and 16 respectively. The SNR was fixed to 15dB. In each case, the minimum point of the RE, z_m , was estimated by (3.41) and matched with the dimensionality of the corresponding data. Fig. 3.18 shows that DIOS accurately predicted the dimensionality for all the different numbers of constituent endmembers. Further to check the robustness of the DIOS to the noise power, noisy signals with different SNR values were created and in each step the minimum point of reconstruction error was calculated. It can be seen in Fig. 3.19 that for all SNR values of 2, 5, 8 and 10dB, DIOS accurately estimated $m^* = \hat{c} = 12$ which is the original dimensionality of the data. DIOS was compared to three other methods, HFC, NWHFC (with

Figure 3.18: Error behavior as a function of choice of dimension when SNR is 15dB and for various values for the true dimensions $c=12, 13, 15$, and 16 respectively.

$P_f = 10^{-4}$) and Empirical Indicator Function (EIF) [60] as reference methods. These methods represent the most used eigenvalue decomposition-based methods. The three methods are also embodied in many correlation-based order selection methods. Thus, comparing DIOS to these methods will lead to a reliable conclusion that encompasses a variety of different hyperspectral analysis methods. In HFC and NWHFC, parameter P_f represents the false-alarm probability. Both HFC and NWHFC sort the eigenvalues resulting from the sample correlation matrix to define the contribution of original signal and noise to the eigenvalues. The number of eigenvalues that contribute to the original signal is selected as the dimension of the hyperspectral data. For simulation, in each step mixed data including a specified number of the USGS library spectrum data as the constituent endmembers was created. Each method was used to estimate the dimensionality of mixed data for that step. Figure 3.20 shows the estimation results for $c = 2, 3, 5, 10, 15, 20$ and 23 . In this experiment SNR was considered equal to 10dB. To define the proximity of each method to the exact estimation, a straight line was plotted as an indication of the accurate estimation. The figure shows that the DIOS plot has the least distance from

Figure 3.19: Error behavior as a function of choice of dimension for a fixed value of $c = 12$ and variable SNRs (SNR=2, 5, 8, and 10dB)

the straight line, confirming the superiority of DIOS over other four methods to provide a better estimation. In the range of $c \leq 16$, DIOS is mostly equal to the exact estimation line, indicating the strength of the method for dimension estimation. It can be seen that DIOS slightly underestimates dimensionality for the case of $c \geq 16$. Averaged MSE error for DIOS, NWFHC, HFC and EIF is 0.0363, 0.1112, 0.0662 and 0.0662, respectively. The MSE indicates a better accuracy for DIOS. The constraint for the upper bound in this example is mainly due to the fact that for this range the number of constituent endmembers in the signal is close to the upper end of the spectrum library (USGS library used in the method has maximum of 24 endmembers). However, DIOS still outperformed all compared methods including NWFHC with enforced low error probability ($P_f = 10^{-4}$) in any range covered by the library spectrum. Figure 3.21 shows the similar simulation when SNR=5dB. It can be seen that all methods including DIOS have been affected by the strength of the noise. However, DIOS still was outperforming other methods and showed less sensitivity to the noise variation. In this case, the averaged MSE error for

Figure 3.20: Comparing different methods of dimension estimation at SNR=10dB. Statistical data is presented in Fig. 3.22.

DIOS, NWFHC, HFC and EIF is 0.0771, 0.2065, 0.0870 and 0.1583, respectively. The MSE indicates again a better accuracy for DIOS.

Figure shows the statistical 2σ line intervals for dimension estimation of DIOS when it was run 100 times. Short distance between the average line and the upper and lower bounds indicates a high confidence in DIOS to provide a correct estimation of the dimension.

Table 3.4 provides some statistical results regarding the success rate of the DIOS, HFC and EIF to estimate the correct dimensionality. All tested methods were repeated 100 times for different values of endmembers quantity and SNRs. Success rate was calculated based on the number of times each method performed accurately out of the 100 runs. Success rates confirms that both HFC and EIF have a fast decreasing trend in accuracy with variation of SNR and number of endmembers. For SNR_i10dB these two methods did not provide an accurate estimation of dimensionality. Specifically, for $c > 10$, EIF is exhibiting a substantial error up to 50%. It can be seen that DIOS performs a fairly accurate and robust estimation in different range of SNRs and number of endmembers. The sensitivity of DIOS to SNR and number of endmembers is very low. The lowest

Figure 3.21: Comparing different methods of dimension estimation at SNR=5dB. Statistical data is presented in Fig. 3.22.

Figure 3.22: Averaged dimension estimation at SNR=5dB with 2σ line intervals.

success rate happens at $c = 2$ and SNR=2dB which is equal to 0.95. This success rate is two times higher than the lowest rate for other two methods. Table 3.4 confirms that DIOS can be used in critical hyperspectral applications where an accurate estimation of dimension is required.

Table 3.4: Success rate statistical results of DIOS, HFC and EIF for 100 runs. Histogram has been presented in Fig. 3.23.

SNR (dB)		15	10	5	2
$\hat{c}=15$	HFC	0.89	0.85	0.80	0.69
	DIOS	1	1	0.98	0.97
	EIF	0.75	0.71	0.63	0.57
$\hat{c}=10$	HFC	0.92	0.90	0.81	0.72
	DIOS	1	0.99	0.98	0.97
	EIF	0.81	0.77	0.72	0.65
$\hat{c}=5$	HFC	0.93	0.92	0.83	0.75
	DIOS	0.98	0.98	0.97	0.96
	EIF	0.88	0.85	0.79	0.70
$\hat{c}=2$	HFC	0.94	0.93	0.85	0.76
	DIOS	0.96	0.96	0.95	0.95
	EIF	0.90	0.87	0.81	0.72

Although the above mentioned results are an indication of DIOS' competitiveness, a better assessment of each algorithm performance is presented in Table 3.5. This table covers a variety of different scenarios for evaluation purposes. It can be seen from the table that the NWHFC, HFC, and EIF could be possible solutions for high values of SNRs, but they are far from an accurate estimation for low SNR scenarios. In some cases, the reference methods are not able to differentiate the increase in the number of constituent endmembers. For example, when SNR is equal to 10dB and $c = 10$, the EIF estimate for the number of constituent endmembers is 13. However, when c increases to 15, this estimate decreases to 9. This invalidates the accuracy of the reference methods for low SNR values. The HYSIME method seems to be more stable for low SNR scenarios compared to the other methods. Yet, this method underestimates the true number of endmembers for noisy situations specifically when the number of endmembers is fairly high. The accuracy

of DIOS in the case of noisy data is due to the fact that the denoising and dimension estimation in this method is performed simultaneously and that the MSE error is successfully minimized in DIOS. The last column in Table 3.5 provides the statistical variance for the estimated dimensions when number of endmembers were equal to 10 and SNR varied from 25dB to 10dB. It can be seen that when SNR value decreases the variances for other methods increases substantially while for DIOS the variance increases only from 0.1 to 0.3 when $c = 10$. The strong contribution of the noise in low SNR applications prevents other methods from being able to differentiate between eigenvalues related to the noise and those related to the original signal, leading to an inaccurate estimation of dimensionality. Figure 3.23 shows the histogram of the estimated dimensions for each method for 100 runs when SNR=2dB and $c=15$. In general, histograms provide more statistical information than mean and variance only. This matches with the very top-right column of the Table 3.5. It can be seen that EIF has a fairly uniform distribution for number of incorrect estimations in the range of $11 < \hat{c} < 15$. However, HFC distribution of estimated dimensions has a decreasing trend with the decreasing dimension. DIOS distribution confirms the accuracy of this method in estimating dimension in noisy hyperspectral applications. This feature is owing to the simultaneous denoising and order selection approach proposed in DIOS.

Fig. 3.24 shows the effect of applying DIOS on band 100 of the noisy data. The SNR is equal to 10dB and c is equal to 7. The original noiseless data has been provided in the figure for comparison. It can be seen that the denoised data in this band is very similar to the original noiseless data indicating the strength of DIOS in denoising noisy hyperspectral applications. Note that the synthetic data used in this paper is produced by mixing the pixels using an averaging mixing filter. This guarantees that the data do not include any pure pixels. As Table 3.5 shows, the eigenvalue-based methods perform worse in this scenario, especially as the SNR decreases. This confirms the problem of these methods' sensitivity to both the noise variance and absence of pure pixels in the image. Even with SNR values as low as 5dB the method still provides a close estimate of the original number of constituents, whereas the other methods have a consistent tendency to provide an order estimation lower than the true value.

Pure Pixel Dependency:

Existing methods are mainly suffering from the high level of dependency on the presence of pure pixels in the scene. To test the robustness of the DIOS to the existence level of pure pixels, different set of synthetic data each having a different threshold value for maximum

Figure 3.23: Histogram of estimated dimensions for HFC, EIF and DIOS with SNR=2dB and $c=15$ for 100 runs. Statistical success rate is presented in Table 3.4.

Figure 3.24: a) Original synthetic noiseless data in band 100, b) Noisy data in the same band (SNR=10dB), c) Denoised data using DIOS

of abundance values were created. A maximum threshold value of 1 is identical to having pure pixels in the scene. Lower threshold values indicate more mixing level and no chance of having a pure pixel. Figure 3.25 is showing the result of dimension estimation for

Table 3.5: Comparison of dimension estimation methods with different SNRs and different constituent endmembers

SNR	Method	$c=3$	$c=5$	$c=10$	$c=15$	Variance for $c=15$
25 dB	HFC($P_f = 10^{-4}$)	3	5	7	12	0.9
	NWHFC($P_f = 10^{-5}$)	4	5	11	12	0.5
	EIF	3	5	7	11	1
	HYSIME	3	5	10	14	0.3
	DIOS	3	5	10	15	0.1
15 dB	HFC($P_f = 10^{-4}$)	3	5	7	8	1.2
	NWHFC($P_f = 10^{-5}$)	4	5	11	8	0.8
	EIF	3	5	6	10	1.2
	HYSIME	3	5	8	12	0.5
	DIOS	3	5	9	15	0.2
10 dB	HFC($P_f = 10^{-4}$)	2	3	6	7	2.9
	NWHFC($P_f = 10^{-5}$)	5	7	14	11	2.5
	EIF	3	4	13	9	2.8
	HYSIME	2	4	8	12	1.9
	DIOS	3	5	8	15	0.4

different cases where maximum abundance values are equal to 1, 0.8, 0.4, 0.2. Synthetic test data were created by mixing 10 endmembers from USGS library. SNR was fixed at 2dB. The region including the Z_m minimum for better visualization was magnified. It can be seen that DIOS estimates the correct dimension ($c=10$) even with a maximum abundance value of 0.4. For a maximum value of 0.2 (which is matched to a fairly high mixed level), DIOS estimation is 11 which is still a fairly accurate estimation specifically with an SNR value as low as 2dB. Table 3.6 provides more statistical information regarding the robustness of the DIOS to the pure pixel existence. The DIOS and VCA were repeated 100 times for different maximum abundances of 1, 0.8, 0.4, 0.2. It can be seen from the table that even with a maximum abundance value as low as 0.2, the DIOS is predicting the correct dimension with 94% probability. The success rate of VCA for the same case is only 79%.

SID Criterion for Unmixing Evaluation: As was shown, the DIOS can be used for order selection only or for order selection and unmixing. The unmixing part of the method is handled by UFU. To evaluate the effectiveness of the spectral matching approach, UFU is compared to Vertex Component Analysis VCA and GDME which are commonly used as effective unmixing methods. For this comparison, The definition of Spectral Information

Table 3.6: Statistical results of DIOS and VCA estimation for 100 runs and $C = 10$

Maximum Abundance		1	0.8	0.4	0.2
# of times detected $\hat{c}=9$	DIOS	0	0	1	2
	VCA	1	5	7	12
	GDME	2	8	11	15
# of times detected $\hat{c}=10$	DIOS	100	100	98	94
	VCA	98	91	87	79
	GDME	96	85	80	72
# of times detected $\hat{c}=11$	DIOS	0	1	1	4
	VCA	1	4	6	9
	GDME	2	7	9	13

Divergence (SID) in [61] is used. SID is a divergence indicator for two pixel vectors from spectral point of view. In fact, if two pixels have similar spectral signatures, the SID value for them is close to zero. SID criterion has been used for many years in hyperspectral processing community and for this reason it has been used in this thesis for comparison purposes. The SID for two pixels $F = [f_1 \ f_2 \dots f_\ell]^T$ and $G = [g_1 \ g_2 \dots g_\ell]^T$ is defined as

$$\text{SID}(F, G) = D(F \parallel G) + D(G \parallel F), \quad (3.43)$$

where

$$D(F \parallel G) = \sum_{i=1}^{\ell} p_i \log(p_i \mid q_i) \quad (3.44)$$

$$D(G \parallel F) = \sum_{i=1}^{\ell} q_i \log(q_i \mid p_i) \quad (3.45)$$

$$p_i = f_i / \sum_{i=1}^{\ell} f_i \quad , \quad q_i = g_i / \sum_{i=1}^{\ell} g_i. \quad (3.46)$$

Fig. 3.26 shows how SID changes with noise for the UFU, VCA and GDME approaches. In each case, G is the data vector representing the last estimated pure pixel and F is the vector related to the exact match from the library. It can be seen that the SID variation for UFU is almost three times less than that of VCA and in some points almost

Figure 3.25: Dimension estimation with different maximum abundance values (exact dimension=10)

6 times less than the estimation of GDME. This indicates that estimations in UFU are more correlated to the real spectra compared to those in VCA or GDME. In lower SNRs, GDME is behaving better than VCA, but this is reversed for higher SNRs. In any case for all SNRs, specifically for very low values of SNR, UFU is estimating more accurately. Note that in comparison, the improved version of the endmembers was not used. As a matter of fact, if we use the improved version, the method will work so efficiently such that the SID will be zero in most cases due to the exact matching. Further simulation results showed that error performance was improved by at least 5% to 10% when spectral matching (the improved version of endmembers) was considered. This process was shown to be much more effective especially for high noise levels of $\text{SNR} < 15\text{dB}$. However, a comparison of SID validates only the accuracy of endmembers estimation, not that of abundances. To

evaluate the complete unmixing procedure, results showed that the MSE estimate of the resulting image is also minimum with the UFU compared to the other approaches.

Figure 3.26: Spectral Information Divergence (SID) comparison of VCA, GDME and DIOS

To study the effect of dimensionality on the estimation error, Figure 3.27 has been provided. This figure shows the bar graph of averaged MSE error of endmember estimation for DIOS, VCA, FCLS and GDME when number of endmembers to be extracted is changing. Endmembers to be mixed were selected from USGS library. SNR was fixed at $5dB$. Each method was repeated 100 times and the estimated error was averaged on all runs. It can be seen that DIOS is outperforming all methods used in the comparison. It can also be seen that all methods but FCLS are showing a decreasing error as the number of endmembers in the signal increases. FCLS estimation error shows an increasing trend when number of endmembers exceeds 12. DIOS averaged error is lower than other methods used for this comparison at least by a factor of 3 when $c > 12$ and a factor 2 when $c < 12$.

Experiment 2: Real Hyperspectral Data In this section, the proposed method was applied to real hyperspectral data collected by AVIRIS. A subset of the Indian Pines test site in northwestern Indiana, acquired by AVIRIS in June 1992, was considered. Table 3.7 shows the dimension estimation inferred by HFC, NWHFC, EIF, HYSIME and DIOS. As it can

Figure 3.27: Averaged MSE of endmember estimation error for 100 runs in VCA, GDME, FCLS and DIOS

be seen in Table 3.7, only the values obtained by DIOS and HYSIME exactly coincides with the number of ground truth materials listed in USGS library ($c = 16$). However, the minimum squared error obtained by DIOS in this estimation is almost twice less than of the HYSIME's, which indicates a better accuracy achieved by the proposed method. Other methods exhibit an error that varies in the range of 18% to 62%. It is interesting to observe that even for a very low false-alarm probability in both HFC and NWHFC, DIOS still outperforms both methods.

The real hyperspectral data collected by the AVIRIS sensor over Cuprite site, in Nevada was used as another test data. The test data was obtained from a sub-scene at the eastern center (250 x 190 pixels and 188 bands) of a dataset acquired on the AVIRIS flight on June 19, 1997. The noisy bands as well as the water vapor absorption bands (including bands 1, 2, 104-113, 148-167, 221-224) were removed from the original 224-band data cube. Complete list of relevant minerals in this site is provided in [63]. Figures 3.28 and 3.29 are the unmixing results from the DIOS. A visual comparison between the DIOS results on the Cuprite data set and the ground truth presented in [63] can identify the

Table 3.7: Comparison of dimension estimation for Indian Pines test site with number of end members $c=16$

Method	Dimension Estimate ($\hat{c}$)
HFC($P_f = 10^{-3}$)	26
HFC($P_f = 10^{-4}$)	23
HFC($P_f = 10^{-5}$)	19
NWHFC($P_f = 10^{-3}$)	20
NWHFC($P_f = 10^{-4}$)	19
NWHFC($P_f = 10^{-5}$)	19
EIF	18
HYSIME	16 ($MSE = 4.5 \times 10^{-4}$)
DIOS	16 ($MSE = 2.3 \times 10^{-4}$)

type of endmember materials matched to the library. Minerals shown in the figures match the library data available for Cuprite listed in [63]. To show the accuracy of endmembers estimation, the exact endmembers from the USGS library have been plotted in Fig. 3.29. Since the USGS library has different instances for each material, the number matching to the instances in the library also has been specified as a reference to the name of the endmember (e.g. Kaolinite # 13). Figure 3.29 shows that estimated signatures extracted by DIOS matches the library signatures with a fairly good accuracy. Since some spectral signatures have close relation, comparison of the abundance maps for these signatures will be more accurate for classification purposes. Table 3.8 has been provided as a quantitative support to compare the accuracy of DIOS with GDME, VCA and Minimum Volume Constrained Nonnegative Matrix Factorization (MVC-NMF) [64] methods in estimating endmembers of the Cuprite site. Spectral Angle Distance (SAD) between the estimated endmember $\hat{a}$ and the exact endmember in the library a in this table is calculated using (B.5). The USGS library provides the signature of minerals as well as their instances (e.g. Kaolinite #1, #2,...) as a reference. However, neither the USGS library nor any other valid reference has provided the exact listing of Cuprite data to the extent of the instances. As a result, only the first instance of each detected material has been used for comparison. DIOS has detected 13 endmembers for the Cuprite site. Including the first instance of each material and excluding the rest, resulted in the table 3.8 with 11 materials. Empty

cells in Table 3.8 indicate that the material has not been detected by the corresponding method. Materials detected and listed in Table 3.8 have been extracted by finding the most correlated match in USGS library. From the table, it can be seen that the existing Pyrophyllite (which is different to Pyrope based on USGS library) in the Cuprite material listing was only detected by DIOS. However, DIOS, as well as VCA, didn't detect the Chalcedony which is listed in the table of Cuprite materials. It can be seen from the table that except the Muscovite all other materials have been detected with a better accuracy by DIOS. This is basically due to the fact that DIOS has picked the Muscovite #13 from library which is not exactly the instantiation in Cuprite site. It seems that other method have manually corrected this incompatibility to achieve a better result.

Table 3.8: Spectral angle distance between library spectra and extracted endmembers by DIOS, VCA and GDME

	GDME	VCA	MVC-NMF	DIOS
Detected materials	13	14	9	13
Alunite	3.6	4.1	3.52	3.0
Kaolinite	4.4	4.8	7.16	3.2
Buddingtonite	4.2	3.8	3.95	2.1
Muscovite	2.3	4.1	2.34	3.0
Montmorillonite	2.6	3.0	8.42	1.3
Nontronite	4.7	2.6	7.22	2.2
Pyrophyllite	-	-	-	2.8
Sphene	3.8	3.1	4.22	2.9
Chalcedony	2.5	-	6.86	-
Desert Vanish	3.3	-	7.01	3.1
Pyrope	-	3.1	-	-

3.5 Hyperspectral Correlation Extractor (HYCE) Method: Colored Noise Correlation Calculation

Assumption of the white Gaussian noise has been a key factor in most of the hyperspectral literatures. However, in some studies a whitening method has been adopted to deal with the correlated noise issue. The whitening method in [45] uses the decomposition of the

Figure 3.28: Abundance fractions estimated with the DIOS for Cuprite site data. SAD data is presented in Table 3.8.

sample covariance matrix to equalize the noise variance of each band in the whitened signal. Method used in [65] estimates the signal source by pre-whitening the signal and uses whitening as an interference suppressor. In [66] a whitening transformation has been proposed to enhance the SNR in colored structural noise. Method used in [67] estimates the signal source by pre-whitening the signal followed by a unitary transformation, which

jointly diagonalize a set of correlation matrices. Even though noise whitening seems to be a well known topic in the literatures, a complete and accurate solution for hyperspectral signal unmixing in the presence of a strong correlated noise has not yet been proposed. The method proposed in [45] is lacking the accuracy for noisy applications where SNR is below 15 dB, and yet the method is still not providing the estimation of endmembers and abundance fractions for a hyperspectral application. In this section, the Hyperspectral Correlation Extractor (HYCE) method is proposed to extract the noise correlation in different bands and to calculate the noise autocorrelation matrix. Noise autocorrelation matrix is further used to whiten the noise for processing in next steps. The HYCE algorithm is based on the nature of the correlation between bands. Next section explains briefly about the correlation of the noise in spectral bands and the concept of Partial Correlation Coefficient.

3.5.1 Correlation Between Bands

Lets consider y_{ijk} as the component of pixel vector y_{ij} in k th band. In this section, for simplicity we drop the index ij to refer the y_k as the signal representing the spatial data in k th-band. Estimation of y_k using a linear combination of the remaining bands $y_1, \dots, y_{k-1}, y_{k+1}, \dots, y_\ell$ leads to the error e_k ,

$$e_k = y_k - \hat{y}_k = \sum_{q=1}^{\ell} \alpha_{kq} y_q = \alpha_k^T y \quad \alpha_k = 1. \quad (3.47)$$

The variance of the error is given by

$$Var(e_k) = \alpha_k^T \Sigma \alpha_k, \quad (3.48)$$

where Σ is the covariance matrix of the data. We calculated the correlation coefficient matrix for a data cube acquired by AVIRIS in June 1992 from a subset of the Indian Pines test site in northwestern Indiana. The preliminary calculation suggested a strong correlation between spectral channels that are up to 50 bands away. To further investigate the strength of this correlations, the method used in [68] is adopted to compute the correlation coefficient between the estimation errors e_i and e_j , by

$$PCC_{ij} = \frac{Cov(e_i e_j)}{\sqrt{Var(e_i) Var(e_j)}}. \quad (3.49)$$

This result, which is known as *Partial Correlation Coefficient* (PCC) [69], measures the linear dependency between two spectral bands after the effect of all other bands are removed. Figure 3.30 shows the partial correlation coefficient matrix for the Indian Pine data set. The center part was magnified to reveal that only neighboring bands have significant linear dependency. In a gray scale mode, the bright pixels indicate the higher correlation and the dark pixels indicate no dependency. Further investigation revealed that a two-band correlation described by a second order Markov model will fairly explain the structure of correlation in spectral bands. In such a model, each band is considered to be significantly correlated with two lower bands:

$$w(k+1) = \rho_1 w(k) + \rho_2 w(k-1) + \varepsilon_w(k+1), \quad (3.50)$$

where $\varepsilon_w(k+1)$ is a white Gaussian random process having a variance $\sigma_{\varepsilon_w}^2(1 - \rho_1^2 - \rho_2^2)$. Parameters ρ_1, ρ_2 are reflecting the strength of the correlation in adjacent bands. Please note that this model doesn't ignore the correlation of one band to the other bands rather than the two adjacent one but only emphasizes that the correlation weight is mostly on the two neighbor bands. A successive calculation of (3.50) for lower bands can confirm this claim.

3.5.2 Mathematical Noise Model

The linear mixture model is used due to its effectiveness. Noise is considered as a fully stochastic process. Noise is regarded as a zero-mean Gaussian process independent of $\tilde{y}$, stationary along (i,j) but not along k , and correlated spectrally and spatially. Spatial stationarity of noise implies that the noise variance remains constant in each spectral band for different regions. Noise correlation coefficients are $\rho_x, \rho_y, \rho_\lambda$ in spatial and spectral directions respectively.

3.5.3 Noise Estimation

Since the noise variance is only a function of the spectral band index, the variance of (2.2) on a general region ς on the k th spectral layer is calculated as

$$\sigma_y^2(\varsigma, k) = \sigma_{\tilde{y}}^2(\varsigma, k) + \sigma_w^2(k), \quad (3.51)$$

where the $\sigma_w^2(k)$ is the noise variance in the k th spectral band and $\sigma_y^2(\varsigma, k)$ is calculated over all pixels belonging to the region ς . From (3.51), by considering spatial stationarity,

it appears that $\sigma_w^2(k)$ can be estimated by averaging the $\sigma_y^2(\varsigma, k)$ over a homogenous area where $\sigma_y^2(\varsigma, k) \equiv 0$. In fact, a homogenous area is defined as a scanned region related to the same material on all pixels in that region. Thus, we have $\sigma_y(\varsigma, k) = \sigma_w(k)$ in a general spectral band in homogenous areas. Therefore, an estimate of $\sigma_w(k)$ namely $\hat{\sigma}_w(k)$, is the y -intercept of the horizontal regression line drawn on the scatterplot of estimated variance $\hat{\sigma}_y$ versus estimated mean $\hat{\mu}_y$. We use the fact that a homogenous area originates a cluster of scatter points aligned along a horizontal line having y -intercept equal to $\hat{\sigma}_w(k)$. To avoid the drawback of a supervised method to find the homogenous areas, we find the best fit line as the horizontal line with minimum squared distance error to all points on the scatter plot. Interception of this line with y axis is identical to the variance. We can now consider the covariance of unity lag along either of the coordinate directions, say k . For simplicity, we eliminate the index i, j :

$$C_y(k; 1) = E\{[y(k) - \bar{y}(k)][y(k+1) - \bar{y}(k+1)]\}. \quad (3.52)$$

Replacing (2.2) in (3.52) and using the independency between noise and reflectance signal to eliminate the expected value of the noise multiplicative terms,

$$C_y(k; 1) = C_{\bar{y}}(k; 1) + E\{w(k)w(k+1)\}. \quad (3.53)$$

Multiplying both sides of (3.50) by $w(k)$ and taking the expected value we can obtain

$$E\{w(k)w(k+1)\} = \rho_1 \sigma_w^2(k) + \rho_2 E\{w(k)w(k-1)\}. \quad (3.54)$$

By assuming $E\{w(k)w(k-1)\} = E\{w(k)w(k+1)\}$ and Replacing (3.54) in (3.53),

$$C_y(k; 1) = C_{\bar{y}}(k; 1) + \frac{\rho_1}{1 - \rho_2} \sigma_w^2(k). \quad (3.55)$$

In a homogenous area, the term $C_{\bar{y}}(k; 1)$ is identically zero such that

$$C_y(k; 1) = \frac{\rho_1}{1 - \rho_2} \sigma_w^2(k). \quad (3.56)$$

Hence, $\rho_1/(1 - \rho_2)$ can be estimated from the slope of the best line fit in unity-lag covariance-to-variance scatterplot in those areas. Since both ρ_1, ρ_2 are unknown, an-

other equation is needed to calculate both parameters. Continuing with the two-lag noise covariance calculation, the two-lag covariance can be calculated as

$$C_y(k; 2) = C_{\bar{y}}(k; 2) + E\{w(k)w(k+2)\}. \quad (3.57)$$

Writing second order Markov model for $(k+2)th$ band

$$w(k+2) = \rho_1 w(k+1) + \rho_2 w(k) + \varepsilon_w(k+2). \quad (3.58)$$

Multiplying both sides of (3.58) by $w(k)$ and taking the expected value we can obtain

$$E\{w(k)w(k+2)\} = \rho_1 E\{w(k)w(k+1)\} + \rho_2 \sigma_w^2(k). \quad (3.59)$$

Using (3.55) to replace $E\{w(k)w(k+1)\}$ in (3.59) we have

$$C_y(k; 2) = C_{\bar{y}}(k; 2) + \frac{\rho_1^2 - \rho_2^2 + \rho_2}{1 - \rho_2} \sigma_w^2(k). \quad (3.60)$$

Hence, $(\rho_1^2 - \rho_2^2 + \rho_2)/(1 - \rho_2)$ can be estimated from the slope of the best line fit in two-lag covariance-to-variance scatterplot in homogenous areas. Two combinational fractions of ρ_1 and ρ_2 will result in calculation of ρ_1, ρ_2 . Knowing the correlation parameters ρ_1, ρ_2 we can estimate the noise correlation matrix in spectral direction as $R_{ww} \in R^{\ell \times \ell}$. Since R_{ww} is available, the whitened noisy signal y_{white} would be calculated as

$$y_{white} = R_{ww}^{-1/2} y = (V_w \Lambda_w^{-1/2} V_w') y. \quad (3.61)$$

The term in parentheses is provided based on the eigenvalue decomposition of R_{ww} . Signal y_w will be used in next section as the input to the unmixing process.

3.5.4 Simulation Results for HYCE

The spectral reflectance used in the subsequent experiments was selected from the USGS library. A set of spectral profiles was selected as the endmembers to generate the mixture. Abundance fractions, s_{ij} , were generated according to a Dirichlet distribution. The data cube was 64×64 pixels in 188 bands. In the first step, five endmembers from the USGS library were selected to form the synthetic data. The proposed second order Markov model was used to add a colored noise with correlation coefficient, $\rho_1 = 0.5, \rho_2 = 0.4$,

along the spectral bands. Variance was selected equal to 0.1. All rows were concatenated (in i direction) in the data cube to create a 4096×188 data matrix and divided the new matrix into 9×9 sliding blocks in the direction of the spectral bands (sliding one pixel at a time). For each sliding block, variance and unity-lag and two-lag covariances were calculated. Figure 3.31, 3.32 and 3.33 show the scatter plot of the variance and two covariances for all blocks. A horizontal line was fitted into the variance scatterplot based on the LSE of the distance to the line. The y -interception of the line is equal to 0.11 which matches the variance of the original colored noise ($\sigma_w^2 = 0.1$). Further, a line was fitted into both covariance-variance scatter plots to pass the origin (since we selected a zero-mean noise). Line slopes as shown in the figure are 0.82 and 0.81 which is fairly identical to the $\rho_1/(1 - \rho_2) = 0.8333$ and $(\rho_1^2 - \rho_2^2 + \rho_2)/(1 - \rho_2) = 0.8167$ for the selected $\rho_1 = 0.5, \rho_2 = 0.4$.

To show the robustness of the HYCE to variance and correlation coefficients Fig. 3.34 is provided. Three different variances, 0.2, 0.3, 0.6 were considered. For each variance, selected ρ_1, ρ_2 were shown on the title of the related covariance-variance scatterplot. To save space, only unity-lag covariance plot were shown. It can be seen that estimated variance and the slope of the best fit line match the original variance and $\rho_1/(1 - \rho_2)$. Figure 3.35 shows the sensitivity of the HYCE to the parameters ρ_1 and ρ_2 . These parameters were changed from 0.1 to 0.95 in step of 0.05 to simulate different correlation factors. A set of five endmembers were used from USGS library to be mixed with the correlated noise generated by parameters ρ_1 and ρ_2 . The HYCE and UFU were used to extract the five endmembers. Averaged MSE of estimation error was calculated in each step and the result was plotted on Fig. 3.35. As it can be seen, the maximum error is at $\rho_1 = 0.95$ and $\rho_2 = 0.95$ which simulates the maximum correlated noise. It can be seen that the error is increasing as ρ_1 and ρ_2 increase. However, the degree of increase is different for ρ_1 and ρ_2 . HYCE is more sensitive to ρ_1 compared to ρ_2 . The reason as discussed in Section 3.5.1 is the higher strength of PCC in neighbor bands. Since ρ_1 is representing the correlation in a closer band, the degree of dependency will be higher. Error is in the order of 10^{-5} when parameters are changed from minimum to maximum in the range of 0.1 to 0.95. A comparison of HYCE and NWFHC whitening method used in [45] is provided in Fig. 3.36. NWFHC method is based on finding the matrix $K_n = \text{diag}\{\zeta_1^{-1}, \zeta_2^{-1}, \dots, \zeta_\ell^{-1}\}$ where ζ_i^{-1} s are the diagonal elements of inverse sample covariance matrix, $K_{\ell \times \ell}^{-1}$. Further, NWFHC uses $K_n^{-1/2} K K_n^{-1/2}$ to whiten the colored hyperspectral signal. This method has been used for many years in hyperspectral image analysis to perform whitening. For this reason, HYCE was compared to NWFHC

in this thesis. A set of six endmembers were used from USGS library to be mixed with the correlated noise generated by parameters ρ_1 and ρ_2 . Each bar graph set on the figure is indicating a set of ρ_1 and ρ_2 . Averaged MSE was calculated based on 100 runs for each set of correlation coefficient. It can be seen that HYCE is outperforming NWHFC by 20 – 100%. HYCE is performing much better than NWHFC in higher range of correlation factors. The reason is that HYCE is capable of extracting the between-band correlation while NWHFC is providing a general whitening solution incapable of extracting such a correlation. When PCC is strong enough, the method which is relying on the structure of the noise would be more efficient leading to a lower estimation error.

3.5.5 Simulation Results of HYCE with MRE and UFU

We selected twelve endmembers from the USGS library to form the synthetic data. Second order Markov model was used to add a colored noise with the correlation coefficient, $\rho_1 = 0.5, \rho_2 = 0.4$, along the spectral bands. Variance was selected equal to 0.2. To simulate the presence of a strong noise, SNR was selected to be 5dB. Figure 3.37.a shows the application of the MRE on the synthetic data. It can be seen that the nature of the correlated noise highly distract the result of the order dimension estimation in MRE. In fact, there is no optimum point on the graph of reconstruction error.

Figure 3.37.b shows the application of the MRE further to the whitening process performed on the synthetic data by HYCE. Using HYCE, ρ_1, ρ_2 was calculated as 0.53 and 0.38 respectively. Noise correlation matrix was calculated to whiten the correlated hyperspectral signal. It can be seen that MRE was able to predict the order of data on the minimum point of z_m graph which is identical to the twelve endmembers originally selected in the simulation.

3.6 Summary

The DIOS was proposed to estimate the dimensionality of hyperspectral data and unmix the hyperspectral data. To provide this estimate, the method minimizes the MSE estimate among subspaces of different orders. Unlike the existing approaches that use different criteria for each stage of denoising and order selection, it was shown that the adaptive order selection goal was to minimize one criterion (MSE) for all stages. Other advantages such as robustness to noise in low SNR applications, as well as higher efficiency when pure materials are represented in very few pixels, were explained. The results showed that as the noise exhibits more strength, DIOS detects the endmembers more accurately compared to the other methods that lose their discriminating power with increasing noise. It was shown that the spectral matching method increase the accuracy of endmember estimation by 10-15%. DIOS was proved to be fairly insensitive to the maximum abundance values of greater than 0.2. Consistency of DIOS were studied by statistical results to prove that dimension estimation is at least 95% accurate and that the MSE error is lower than other existing accurate methods by a factor of 2 to 3. This chapter also discussed the proposed HYCE method to extract the noise correlation structure and to unmix a hyperspectral signal when noise is highly strong and spectrally correlated. A second order Markov model and an unsupervised scatterplot based approach was proposed in HYCE to find the spectral correlation coefficients. It was shown that HYCE could use the information from the correlation extraction stage to unmix the hyperspectral data accurately even in the presence of a strong noise. Simulation results showed that HYCE was robust to the variation of second order statistics. Unmixing outputs confirmed that HYCE accurately estimated the endmembers spectral signatures as well as endmembers mapping. HYCE unmixing approach was compared to other methods for accuracy and speed. Comparison results verified that HYCE outperforms other existing method.

Figure 3.29: Extracted signatures from Cuprite site data using the DIOS: Library endmembers (dashed line), Estimated endmembers (Solid line). SAD data is presented in Table 3.8.

Figure 3.30: Partial correlation coefficient plot for Indian Pine image (area in the square is magnified on the right side)

Figure 3.31: Variance versus mean squared scatterplot

Figure 3.32: Unity-lag covariance and variance scatterplot

Figure 3.33: Two-lag covariance and variance scatterplot

Figure 3.34: Estimation of different variances and correlation factors

Figure 3.35: Sensitivity of HYCE to the variation of correlation factors

Figure 3.36: HYCE comparison to NWHFC for different ρ_1 and ρ_2

Figure 3.37: a) MRE estimation when correlated noise is ignored (SNR=5 dB, $c=12$) with no minimum point to detect the correct dimension, b) MRE estimation for a correlated noise (SNR=5 dB, $c=12$) after HYCE is applied. Minimum point is representing the correct dimension.

Chapter 4

Concluding Remarks and Future Work

The following is a summary of the main contributions of this thesis:

- **Ultra-Fast Unmixing**

The new unmixing method based on least square approach and a new innovative library matching was introduced. It was shown that the concept of transformation matrix could be utilized to overcome the speed performance issue. Experimental results over samples of real and synthetic data showed that the method is about 10 times faster than the fastest existing methods for a data cube size of 4096 pixel with 224 bands. The accuracy problem was addressed by implementing the concept of spectral matching method in this contribution. The spectral matching method increased the accuracy of endmember estimation. Using the SID criterion for evaluation purposes, it was concluded that the level of accuracy in ultra-fast unmixing method was higher than that of well known existing methods. When UFU was used as part of DIOS, it was shown an improvement up to 60% in MSE. Furthermore, it was illustrated that unlike state of the art approaches, processing time in the proposed method was independent from the noise level. Proposed unmixing method is definitely a departure from the trade off between speed and accuracy and has potential application where fast hyperspectral imaging is in demand.

- **Simultaneous Denoising and Order Selection**

Proposed DIOS method implemented minimizing the reconstruction error for subset of different orders with the purpose of simultaneous order selection and denoising.

Unlike the existing approaches that use different criteria for the two stages of denoising and order selection, the adaptive order selection goal was to minimize one criterion MSE for all stages. In addition, other advantages of this method over the ones with eigenvalue decomposition was its robustness to noise in low SNR scenarios, as well as its higher efficiency when pure materials were represented in very few pixels in a scene. Sensitivity of the proposed method to the pure pixel index (ratio of the pure pixel to the mixed pixel) was evaluated. It was shown that even a high mixing level with a low pixel index could be processed with the proposed method. It is worth mentioning that, the order selection and denoising steps in DIOS work in parallel with the UFU unmixing method.

- **Hyperspectral Correlation Extractor**

A second order Markov model was considered for the scenarios with the colored noise. We proposed a new method to estimate the noise autocorrelation matrix. Simulation results showed that correlation extractor method was robust to the variation of correlation factors. Combining the proposed whitening method with DIOS provided more accurate estimates of endmembers, abundances and dimension. In addition, simulations confirmed the advantages of the proposed method over the existing ones in the sense of accuracy and speed in the presence of high correlation coefficients.

4.1 Some Potential Future Work

Expanding the method used in this research to non-linear unmixing model is a possible direction for future work. On the other hand, results confirmed that the proposed method was more accurate for a fairly high mixing level and in the absence of pure pixels. However, it has been observed that for scenarios where the maximum abundance value is lower than 0.2, the proposed method is not providing fairly accurate results. Improving this thesis for these rare cases is another path for future work.

Expanding the proposed idea for additive noises rather than Gaussian, is another potential area of research for future work. For example, application of the Generalized Gaussian Density (GGD) distribution [55] in the proposed research can be considered as a future work.

Appendix A

Upper Bound and Lower Bound on Δ_{S_m}

The probability density function of the Chi-square distribution is

$$f(x; k) = \begin{cases} \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2} & \text{if } x \geq 0; \\ 0 & \text{if } x < 0 \end{cases}$$

If the order of Chi-square random variable X_m is large, it can be well estimated by a Gaussian distribution. In this case the validation probability p_2 is considered in the form $p_2 = Q(\alpha)$ where $Q(\alpha) = \int_{-\alpha}^{\alpha} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$. It is shown in [52] that if p_2 , or equivalently α , is chosen large enough ($p_2 \geq 0.9$) such that

$$\alpha \geq \frac{N}{\sqrt{2(N-m)}} \left(1 - \frac{m}{N} - \frac{x_m}{\sigma_w^2} \right), \quad (\text{A.1})$$

then with validation probability $p_2 = Q(\alpha)$ we have

$$L_m(y, Q(\alpha)) \leq \frac{1}{N} \|\Delta_{S_m}\|_2^2 \leq U_m(y, Q(\alpha)). \quad (\text{A.2})$$

where

$$U_m(y, Q(\alpha)) = x_m - m_w + \frac{2\alpha^2 \sigma_w^2}{N} + K_m(\alpha) \quad (\text{A.3})$$

$$K_m(\alpha) = 2\alpha \frac{\sigma_w}{\sqrt{N}} \sqrt{\frac{\alpha^2 \sigma_w^2}{N} + x_m - \frac{1}{2} m_w} \quad m_w = (1 - \frac{m}{N}) \sigma_w^2. \quad (\text{A.4})$$

The lower bound $L_m(y, Q(\alpha))$ is zero if

$$(m_w - \alpha\sqrt{v_m}) \leq x_m \leq (m_w + \alpha\sqrt{v_m})v_m = \frac{2}{N}(1 - \frac{m}{N})\sigma_w^4. \quad (\text{A.5})$$

Otherwise the lower bound is

$$L_m(y, Q(\alpha)) = x_m - m_w + \frac{2\alpha^2\sigma_w^2}{N} - K_m(\alpha). \quad (\text{A.6})$$

Appendix B

GA-LSE Algorithm

The mathematical intractability of the abundance non-negative constraint results in complex and extensive numerical approaches. Due to such mathematical intractability, many LSE based methods are unconstrained and can only produce suboptimal solutions. Some efforts were devoted to solving fully constrained linear mixing problems. However, the approaches used to implement these constraints were designed mainly for a small number of material signatures. In [70] constrained least square solution is obtained by solving an overdetermined system that consisted of m equations with w unknowns ($n < m$), where m is the number of bands and w is the number of signatures. Since there are no closed-form solutions, one must examine possible solutions in a feasible region bounded by the Abundance Sum-to-one Constraint (ASC) and Abundance Nonnegativity Constraint (ANC). The use of quadratic programming techniques to impose the ASC and ANC were investigated in [71]. Nevertheless, the algorithm used were computationally expensive. Other methods presented also suffered from excessive computational complexity as the number of materials increases. Some evolutionary method have been proposed in hyperspectral to ease the mathematical complexity. In [72] a method for estimating the partial abundance of spectrally similar minerals in complex mixtures using simulated annealing. The method requires formulation of a linear function of individual spectra of individual minerals. The first and second derivatives of each of the different sets of mixed spectra and the individual spectra are determined. The error is minimized by means of simulated annealing. In [73] Genetic Algorithm (GA) was proposed to generate multiple mixture models using subsets of reference spectra, providing the analysts with alternative explanations of the hyperspectral data. In [74] a selective principal component analysis method based on GA with subgroups is proposed. This method has the distinct characteristic of combining feature selection and extraction together via the introduction of subgroup concept and

tries to solve the problem of how many and which bands should be selected for feature extraction. Further, the transformed features are used for classification. In general, the simplicity of evolutionary methods such as GA to implement the equality and inequality constraints could ease the complexity of the process assuming that the processing time is handled carefully. In this chapter, a GA algorithm is proposed to estimate the number of spectrally distinct EMs and then a mixed GA and LSE-based (GA-LSE) estimation method is proposed to extract the EM matrix and related abundances vectors. The proposed GA-LSE method is applied to the subject of unmixing hyperspectral data. GA-LSE summarizes in three steps as follows.

B.0.1 Step 1: Estimation of Virtual Dimensionality

VD is utilized to estimate the number of endmembers in this step. In passive sensor array processing, a similar problem to VD estimation also arises from the estimation of the number of signal sources impinging upon the array. This problem is related to how to select an appropriate model for a parameterized family of probability density functions used to best fit the sensor array data. It is known that the commonly used Minimum Description Length (MDL) method proposed by Rissanen [75], can be used for model selection. Following formula was obtained in [76]:

$$MDL(p) = -\log \left(\frac{\prod_{j=p+1}^L \lambda_j^{\frac{1}{L-p}}}{\frac{1}{L-p} \sum_{j=p+1}^L \lambda_j} \right)^{(L-p)N} \quad (\text{B.1})$$

$$+ 0.5p(2L - p) \log N, \quad (\text{B.2})$$

where p is the number of free parameters that specifies a family of probability density functions. $\lambda_1 > \lambda_2 > \dots > \lambda_L$ are eigenvalues generated by sample covariance matrix $K_{L \times L}$. A linear unmixing problem could be formulated as a passive sensor array problem, where signal sources impinging on a sensor array were interpreted as desired target sources to be detected from a set of co-registered images acquired by a bank of spectral channels. With this interpretation, a linear mixing problem can be solved by techniques that are used to solve sensor array problems. In order to take advantage of these two criteria and apply them to the estimation of the VD, we need to examine their underlying assumptions, i.e., 1) noise must be independent identically distributed (i.i.d.), and 2) the observation process is a zero-mean Gaussian random process. When noise is Gaussian, the first assumption

can be taken care of by the whitening technique, in which case the noise variance will be normalized to unity. Then, the resulting sample covariance matrix can be used in the Gaussian process made in the second assumption. Now, the problem of estimating VD can be solved by minimizing the $MDL(p)$ as follows :

$$VD_{MDL} = \arg \left\{ \min_p MDL(p) \right\}. \quad (B.3)$$

B.0.2 Step 2: LSE-Based Endmember Extraction

In order to detect endmembers from the given hyperspectral data cube automatically, we apply the proposed UFU method explained in 3. Following the estimation of the first two endmembers using UFU the following Genetic algorithm based abundance estimation method is then applied. The new endmember is selected from the pixels yielding the largest residuals. Since the averaging scheme that normally is adopted tends to smooth out the spectral details when the noise level is low, Maximum Entropy method first estimates the noise level of the given data. If the SNR is higher than a predetermined threshold, only the pixel with the largest LSE is used as the new endmember. Otherwise, the method takes the average of a set of pixels with the largest LSEs as the new endmember. In this way, the noise effect is successfully suppressed when the SNR is low and the smoothness effect is avoided when the SNR is high. The estimation of SNR is given by:

$$SNR = 10 \log \frac{P_s}{P_t - P_s}. \quad (B.4)$$

where $P_t = E[Y^T Y]$, $P_s = E[Y^T U_c U_c^T Y]$ and $Y = [y_1, y_2 \dots y_N]$. U_c is $l \times c$ matrix formed by the leftmost singular vectors of singular value decomposition of the given data matrix.

B.0.3 Step 3: GA estimation of abundance vectors

GA is an adaptive search technique which derives the name from the fact that its operations is similar to the mechanics of genetic model of natural systems. GA typically maintains a constant-sized population of individuals which represent samples of the space to be searched. Each individual is evaluated on the basis of its overall fitness with respect to the specific application domain. New individuals(samples of the search space), which have higher performance than their "parents", are generated by mainly using three genetic operators: selection, crossover and mutation. This eventually leads to a population that

has improved fitness with respect to the given goal and continue until it reaches the best estimation limited by the stopping factors. In this paper, GA is used to work in parallel with the LSEM method to estimate the abundance factors. The details and the steps of the GA algorithm is as bellows:

[Encoding] Since abundance factors are real values "value Encoding" is used. Process is started by generating a random uniform population of 40 chromosomes.

[Fitness] Least square method is used to minimize the final error. The fitness function is $abs\{(As - Y)^T(As - Y)\}$. As GA and LSEM are working in parallel a parameterized vector fitness function is used. Sum-to-one constraint function is selected.

[New population] A new population is created by repeating following steps until the new population is complete.

[Selection] Stochastic uniform selection method is chosen as it shows a better result in this case.

[Crossover] Crossover probability is chosen equal to 0.6.

[Mutation] Uniform mutation is selected with the probability of 0.02.

[Accepting] New offspring is placed in a new population.

[Replace] New generated population is used for a further run of algorithm.

[Test] If the end condition (error tolerance, maximum generation) is satisfied, stop.

[Loop] Go to the [New population] step.

B.1 Simulation Results

The spectral reflectance used in the subsequent experiments are selected from the USGS digital spectral library which contains 224 spectral bands covering wavelengths ranging from 0.38 to 2.5 μ m. A set of four spectral profiles were selected as the endmembers to create the mixture as shown in Figure B.2 (indicated by true EMs). To create linear mixtures, abundance vectors were randomly selected assuming that they were positive and the summation was equal to one. Then spectral endmembers were multiplied with this abundances to create the hyperspectral image. We added a white Gaussian noise with SNR=20. The resulting image was then degraded by a spatial low-pass filter to produce mixed pixels. With this mixing method, it was aimed to simulate a hyperspectral scene with endmembers arranged in discrete patches so that a linear mixture model would be appropriate. The low-pass filter was a simple $k \times k$ average filter. The value of parameter k controls the degree of mixing. With a small k , only the pixels close to the block boundary were mixed, so the mixture data were very likely to contain pure pixels. On the

other hand, a large size might remove all the pure pixels and more endmembers may be involved when the mixed pixel is generated. Minimizing the *MDL* function, the number of distinct components in the mixed data found to be equal to 4 which matched to what was preliminary considered to create the mixed data. The GA algorithm was tested with several population size to find out the optimum value for this parameter. In each run, the LSE error was measured for estimation of the abundance factors. The result is shown in Figure B.1. As it can be seen the "population size=40" is an optimum for this parameter so it was chosen for the simulation.

Figure B.1: LSE error versus the population size

Figure B.2 shows the simulation result. As it can be seen, all four spectral endmembers have been successfully extracted from the image. To show the abundance values, a gray scale mapping is adopted. Therefore, the brightness of the pixels is proportional to the abundances values. Obviously a black point means that the material which is identified with the related endmember doesn't exist in that location of the image. The abundance factors also have been successfully estimated with the LSE error equal to the 0.017. This error is much lower than what we normally expect from an acceptable hyperspectral analysis process.

In figure B.3, the effect of the Gaussian noise is shown with $SNR = 10$. As it can be seen, noise causes the GA-LSE method to be degraded in a way that some sharp changes in the spectral endmembers are not completely followed by the algorithm. Specifically in the case of fourth endmember this is more clear. Regardless, the purposes of analysis was not to extract the exact endmember but to recognize the signature and type of material in the image. Having said that, GA-LSE method was still capable of recognizing

Figure B.2: The GA-LSE estimation of endmembers and abundances (SNR=20)

the spectral signature. Proposed correlation based method successfully matched the estimated endmember with one of the known spectral signatures in USGS library. As for the abundances, noise is affecting in a way that most of the values are shifted to a higher value. This can be easily noticed by seeing a brighter mapping of pixels compared to the true endmembers of the image.

Figure B.4 shows the scattered noise plot for SNR=7. It can be seen that for this SNR the plot is linear, meaning that with any variation of noise the prediction is very close to the correct values. Results showed that for SNR=10, 15, 25, 30 the same linear scatter plot confirming the robustness to noise in the proposed method.

To compare GA-LSE accuracy with other methods, concept of the SAD is used. The SAD between the estimated endmember $\hat{a}$ and the original endmember a is calculated as

$$\text{SAD} = \cos^{-1} \left(\frac{a^T \hat{a}}{\|a\|_2 \|\hat{a}\|_2} \right). \quad (\text{B.5})$$

In fact, the SAD is an indication of the similarity or the correlation between the real

Figure B.3: Gaussian noise effect on unmixing: SNR=10

Figure B.4: Noise scattered plot for SNR=7

Table B.1: Spectral angle distance between USGS spectra and estimated spectra by UFU, VCA and GA-LSE

Endmember #	VCA	UFU	GA-LSE
1	3.5	3.2	3.0
2	3.6	3.25	3.3
3	3.7	3.4	3.5
4	3.9	3.5	3.4

vector and the estimated one. A low SAD value is an indication of the closeness between the original and estimated spectral signature. Table B.1 shows the SAD values for the UFU, GA-LSE and VCA methods for all four endmembers originally selected from USGS library for mixing. It can be seen that GA-LSE is outperforming VCA by exhibiting a better SAD values. However, GA-LSE is not a major improvement to the UFU. In fact, GA-LSE is easing the implementation of the constraints to the cost of extremely low convergence speed.

Appendix C

Regression Method and RAP for Noise Variance Estimation

C.1 Regression Method

The multiple regression concept is based on the statistical fact that if two variables are correlated, then knowing the score on one variable will allow the prediction of the score on the other variable. The stronger the correlation, the closer the scores will be to the regression line indicating a better accuracy in the prediction. Multiple regression is an extension of this principle, where we predict one variable on the basis of several other variables. This method considers a white noise but there has been some extensions to the multiple regression method that use a colored noise. Let $Y = [y_1 \ y_2 \dots y_N]$ denote an $\ell \times N$ matrix holding the N spectral observed vectors of size ℓ . Define the matrix $\Psi = Y^T$ and the $N \times 1$ vector $\psi_i = [\Psi]_{:,i}$, where $[\Psi]_{:,i}$ stands for the i th column of Ψ (i.e., ψ_i contains the data read by the hyperspectral sensor at the i th band for all image pixels), and the $N \times (\ell - 1)$ matrix $\Psi_{\partial_i} = [\psi_1, \dots, \psi_{i-1}, \psi_{i+1}, \dots, \psi_\ell]$. Assume that ψ_i is explained by a linear combination of the remaining $\ell - 1$ bands. Formally, this consists in writing

$$\psi_i = \Psi_{\partial_i} \delta_i + \varrho_i, \tag{C.1}$$

where Ψ_{∂_i} is the explanatory data matrix, δ_i is the regression vector of size $(\ell - 1) \times 1$, and ϱ_i is the modeling error vector of size $N \times 1$. For each $i \in 1, \dots, \ell$, the least squares

estimator of the regression vector δ_i is given by

$$\hat{\delta}_i = (\Psi_{\partial_i}^T \Psi_{\partial_i})^{-1} \Psi_{\partial_i}^T \psi_i. \quad (\text{C.2})$$

the noise value estimate $\hat{\varrho}_i$ and its variance estimate are

$$\hat{\varrho}_i = \psi_i - \Psi_{\partial_i} \hat{\delta}_i, \quad \text{Var}(\hat{\varrho}) = \frac{\sum_{i,j} (\varrho_{ij} - \bar{\varrho})^2}{N-1}, \quad (\text{C.3})$$

and the noise correlation matrix is given by

$$\widehat{R}_{ww} = [\hat{\varrho}_1, \dots, \hat{\varrho}_N]^T [\hat{\varrho}_1, \dots, \hat{\varrho}_N] / N. \quad (\text{C.4})$$

In [55] a computational method is proposed to overcome the complexity of calculating the pseudo inverse $\Psi_{\partial_i}^\# = (\Psi_{\partial_i}^T \Psi_{\partial_i})^{-1} \Psi_{\partial_i}^T$.

C.2 Residual Autocorrelation Power

RAP is a recent noise variance estimation method for BayesShrink denoising. BayesShrink provides a threshold that empirically minimizes the Bayesian Risk based on the assumption that the images show properties of Generalized Gaussian Distributions (GGD) [42]. Although many approaches emerged after BayesShrink that outperform this method in some sense, the simplicity and efficiency of this threshold still keep it among the well used image denoising approaches [77][78]. It is known that the autocorrelation of a white noise has the following form:

$$R_{ww}[i, j] = \sigma^2 \delta[i, j], \quad (\text{C.5})$$

where δ is the Dirac-delta function. While the autocorrelation of the noise is an impulse, the autocorrelation of the noise-free image is colored. In RAP, the autocorrelation of the residuals of the denoising process is used to estimate the noise variance. First a range of candidates for the noise standard deviation $\Omega = [\sigma_1, \sigma_2, \dots, \sigma_\kappa, \dots, \sigma_K]$ is selected. Further, each noise standard deviation σ_κ in the BayesShrink approach is used to denoise the data with threshold defined in (2.22). The error associated to the denoising with the κ th stddev candidate is

$$\hat{w}_\kappa = y - \hat{y}_\kappa, \quad (\text{C.6})$$

where $\hat{y}_\kappa$ is the resulting denoised data with the κ th threshold in (2.22). Calculating the autocorrelation of the residuals, $R_{\hat{w}_\kappa \hat{w}_\kappa}$, the residual autocorrelation power (RAP) is

$$RAP_\kappa = \frac{1}{N} \sum_{i,j} R_{\hat{w}_\kappa \hat{w}_\kappa}^2[i, j], \quad (\text{C.7})$$

where N is the number of points in the autocorrelation. Figure C.1 shows the RAP of a noisy Lenna image for a range of stddevs candidates from zero to 20 with steps of 0.1, when the true stddev is 15. As the figure shows, there is a rapid growth of the RAP right before the true stddev and the slope of P_κ is much smaller around that stddev. To capture this behavior, RAP suggests using the difference of RAP as seen in Figure C.2:

$$D_\kappa = RAP_{\kappa+1} - RAP_\kappa. \quad (\text{C.8})$$

Consequently, RAP estimates the noise variance through the following steps:

$$\kappa_{max} = \arg_\kappa \max D_\kappa. \quad (\text{C.9})$$

$$\kappa^* = \min_{\kappa > \kappa_{max}} \{ \kappa : D_\kappa < 10^{-2} D_{\kappa_{max}} \}, \quad (\text{C.10})$$

and the RAP estimate of the noise stddev is:

$$\hat{\sigma}(\text{RAP}) = \sigma_{\kappa^*}. \quad (\text{C.11})$$

To choose a range of stddev candidates, RAP starts with the preliminary estimation of the noise variance calculated by the BayesShrink in (2.19). In BayesShrink, the unknown noise variance is estimated by MAD which is the median of absolute value of the wavelet coefficients at the finest decomposition level (level one of the diagonal decomposition HH_1). The RAP method searches around the MAD value with samples a portion of this value apart (for example one tenth) until the difference in the RAP reaches a maximum. Due to the importance of RAP in the comparison, RAP's Algorithm is presented as follow:

Figure C.1: Residuals autocorrelation power (RAP) for a range of standard deviation candidates when the true stddev is 15 for the Lenna image.

Figure C.2: Difference of RAP, D_κ in (C.8), for the RAPs in Figure C.1.

Algorithm 1: RAP Algorithm

Input: Mixture data $y = [y_1, y_2, \dots, y_N] \in \mathbb{R}^{\ell \times N}$ (N is the number of pixels).

Output: Noise variance σ_w^2

- 1) Calculate the initial MAD estimation of noise variance from (2.19)
- 2) Choose a candidate set for the noise stddev $\Omega_k = [\sigma_1, \sigma_2, \dots, \sigma_\kappa, \dots, \sigma_K]$

- 3) Repeat
 - 4) Denoise the data using the standard deviations in the candidate set as a threshold in (2.22)
 - 5) Calculate the denoising error (C.6)
 - 6) Calculate the residual autocorrelation power (C.7)
 - 7) Calculate the RAP difference (C.8)
- 8) until the RAP difference reaches a maximum
- 9) Calculate the noise variance using (C.9)-(C.11)

Appendix D

Hyperspectral Robust Analysis Package (HyRap)

Methods presented in this thesis, are part of the newly developed commercial package denoted by **HyRap**. Most of the hyperspectral softwares are developed to facilitate the visualization of the hypercube or to perform some straight forward manipulations. The potential competitors of HyRap are:

- ENVI software from ITT Visual Information Solutions (ITT VIS)
- Hypercube from US Army geospatial [79]
- Spectral Viewer from Geist Software Labs [80]
- Hyperspectral Viewer from Microimages [81]

Among the existing softwares, ENVI is the most comprehensive and most commonly used hyperspectral software. For this reason, we will concentrate on ENVI and briefly compare it to HyRap.

D.1 ENVI Software

ENVI was originally developed to address the geospatial, urban planning, mining, geology, and space and earth science imagery. Figure D.1 shows the hyperspectral processing flow implemented in ENVI and HyRap. It is worth mentioning that same ENVI is still used for recent application of hyperspectral imaging. In next sections, each stage of the processing flow in ENVI and HyRap will be explained briefly. Further, an example to compare ENVI and HyRap in a biochemistry application will be provided.

Spectral and spatial browsing: In this stage, data is examined using spectral and spa-

Figure D.1: Hyperspectral processing flow implemented in ENVI and HyRap

tial browsing, and color composites to characterize spectral variability and to determine residual errors. Reflectance signatures for water, vegetation, urban areas, and geologic materials are extracted and compared to the spectral libraries.

Spectral and spatial data reduction: ENVI has the option of dimension reduction as one of the beginning stages in the analysis. MNF and Pixel Purity Index (PPI) [82] methods are used for spectral and spatial dimension reduction respectively.

Visualization and identification: In each dimension, visualization is provided for the user to get a better understanding of the new processed data. User can review MNF eigenvalue plot to determine the break in slope and relate to spatial coherency in MNF eigenvalue images and to determine the MNF cut-off between signal and noise to define the dimensionality. In ENVI, the PPI method is used to rank the pixels based on relative purity and spectral extremity. Based on the PPI image, the histogram and threshold are examined to create a list of the purest pixels. Figure D.2 shows a snapshot of the MNF eigenvalue plot used in ENVI.

Classification: ENVI uses Spectral Angle Mapper (SAM) method [83] and available

Figure D.2: Snapshot of the MNF eigenvalue plot used in ENVI (adopted from [3])

Figure D.3: Snapshot of an ENVI mapping output (adopted from [3])

library spectra to perform the classification. The angle between the pixel in datacube and the predefined library spectrum is a measure of classification in ENVI.

Unmixing and mapping: As the final analysis stage, different optional unmixing methods could be used to identify the endmembers. Spatial occurrence and abundance of materials are mapped in the scene. SAM is normally used to determine the spectral similarity to image endmember spectra. A manual SAM classification can also be performed. Error image is examined to evaluate linearity, particularly whether the physical constraints of non-negative and sum to unity have been satisfied. Figure D.3 shows a snapshot of an ENVI mapping output.

D.2 HyRap: Hyperspectral Analysis Package

HyRap has been designed to meet the requirements of hyperspectral applications. Existing version of the HyRap is able to read a hypercube in ENVI, BIL or MAT format. ENVI format is the standard format used originally by the ENVI software and BIL format is a standard multispectral format to show a 3D data structure. MAT file format is used by Matlab to save or load data. HyRap features will be explained in next section.

HyRap spectral/spatial browsing and visualization: Figure D.4 shows a snapshot of HyRap including some manipulation features. In this snapshot, a hypercube of cells has been opened. User can crop the image to the desired area and scan the image through the spectral bands. As a very usefull feature, user can move the cursor on the image and observe a dynamic spectral plot for each pixel on the image. At last, user can accumulate the spectral plots on a separate figure for comparison purposes. This is a very useful feature in manipulation of a hyperspectral cubes to provide a sense of scattered signatures on the image. Some other summarized manipulation features of HyRap are: Band removal, Correlation calculator, divide datacube by reference, add signature to the library, variation classifier, point classifier and independent denoising.

HyRap identification: Figure D.5 shows a snapshot of the Hyrap dimension estimation

Figure D.4: Snapshot from Hyrap datacube manipulation features

feature. Quantum dots hypercube with three endmembers was opened by the Hyrap and

the dimension estimation algorithm has been performed to create the normalized error graph as shown on the bottom of the figure. MRE method is used in this feature. It can be seen that the minimum point of the graph is at 3 which matches to the number of endmembers in quantum dots data.

Figure D.5: Hyrap dimension estimation feature

HyRap unmixing and mapping: Further to the estimation of dimension by Hyrap the unmixing algorithm can be performed to extract those three endmembers. UFU is used at this stage. Figure D.6 shows the unmixing result provided by Hyrap for the quantum dots data. Green areas (bright areas in the monocolor print) are the mapping of each endmembers on the figure.

Figure D.6: Hyrap unmixing feature

HyRap classification: Depend on the availability of spectral signatures from any existing library and assuming that the same signature has been preserved in the hypercube, HyRap, same as the other existing hyperspectral softwares, is able to classify that end-member. Spectral correlation matching is used in this feature. Figure D.7 is showing the snapshot of HyRap for classifying a specific point marked by HyRap on the hypercube. In this case, HyRap is showing the spectral signature of the classification point on the right side of the figure.

D.3 ENVI versus HyRap

Advantages of HyRap over ENVI are,

- HyRap provides blind unmixing and order estimation.
- HyRap provides a better classification for endmembers with similar spectral characteristics (Due to lack of a proper criteria in ENVI to differentiate the similar spectra).

Figure D.7: Hyrap classification feature

- HyRap provides automatic denoising versus a heuristic denoising for ENVI.
- Learning curve for ENVI is much longer than HyRap.

The second listed advantage will be elaborated in detail here. This section starts by providing an example using the test data given by Cytoviva for a sample of salmonella bacteria and peanutbutter. Figure D.8 shows the sample of peanut butter which has been infected by the salmonella bacteria. Salmonella is spread in the sample but is mostly concentrated in the circled area. Detection of salmonella in any food product is very important from the safety point of view. Any false detection could lead to a public health issue.

Figure D.8: Peanut butter infected by the salmonella bacteria

Figure D.9: ENVI mapping of the salmonella bacteria

Figure D.10: Hyrap classification results

Salmonella and peanut butter sample was examined by the ENVI at Cytoviva's facility. Figure D.9 and D.10 show the result of detection for ENVI and HyRap respectively. It can be seen that ENVI hasn't been able to distinguish between peanut butter and the salmonella. In fact, the whole sample has been detected as the salmonella which indicates a false detection due to the lack of accuracy. To explain the source problem in this scenario,

the spectral signatures of the peanut butter and salmonella are provided when they are not mixed together as shown in Figure D.11. It can be seen that the signatures of the peanut butter and salmonella are highly correlated and their peaks are very close to each other. Figure D.12 shows how signatures of PB and salmonella spectrally shifted when

Figure D.11: Salmonella and PB separate signatures

Figure D.12: Salmonella and PB mixed signatures

they were mixed. This implies that a direct spectral matching of individual signatures would not result to the classification of salmonella in PB. As another issue, it can be seen that isolated salmonellas have different signatures compared to salmonellas enclosed by peanut butter. The SAM method used in ENVI utilizes the concept of the SAD to unmix the endmembers in a hyperdata cube. SAD uses the angle between two spectra as a measure of divergence between the two spectra. If two spectra are fairly similar, the SAD measure will be small making them impossible to be differentiated by SAD criteria.

Since the spectral signatures of the peanut butter and salmonella are highly correlated the SAM is not able to unmix them from the hypercube using SAD. Another issue is that the signatures in the library were shifted in the mixture of peanut butter and salmonella so SAD can not be used as the reference for comparing either salmonella or the peanut butter to their library spectral signature.

Further, HyRap was used to achieve the classification in Fig. [D.10](#). It can be seen that salmonella has been successfully unmixed from the data. No Peanut butter has been classified as salmonella as it was previously classified in the mapping output provided by the ENVI. We are hoping to use the HyRap potential in the field of medical science, biology and pharmaceutical industry in near future.

Appendix E

List of Publications

Patent

Masoud Farzam, Soosan Beheshti, Hyperspectral Analysis Package (HyRap), Canadian patent pending, Nov 2010

Journal Articles

1. Masoud Farzam, Soosan Beheshti, “ Simultaneous denoising and intrinsic order selection in hyperspectral imaging”, IEEE Transactions on Geoscience and Remote Sensing, issue 10, vol. 49, 2011
2. Masoud Farzam, Soosan Beheshti, “ Adaptive noise variance estimation and intrinsic order selection for low SNR hyperspectral signals”, Accepted for publication in the IEEE Canadian Journal of Electrical and Computer Engineering, Nov. 2010
3. Masoud Farzam, Soosan Beheshti, “ Endmember transformation and replacement in real rime hyperspectral unmixing”, International Journal of Circuits, Systems and Signal Processing, issue 1, vol. 3, 2009
4. Masoud Farzam, Soosan Beheshti, “ Hyperspectral unmixing in the presence of colored noise”, In preparation, To be submitted to the IEEE Geoscience and Remote Sensing Letters in Sep. 2011

Conference Papers

1. Masoud Farzam, Soosan Beheshti, “ Robust hyperspectral signal unmixing in the presence of correlated noise”, IEEE Conference on Acoustic and Signal Processing (ICASSP), Prague, May 2011

2. Masoud Farzam, Soosan Beheshti, “Information theoretic assessment of correlated noise in hyperspectral signal unmixing”, CCECE conference, Niagara falls, Canada, May 2011
3. Masoud Hashemi, Soosan Beheshti and Masoud Farzam, “Two stage quantization of noisy hyperspectral images”, IEEE 25th Biennial Symposium on Communications, Kingston, Canada, May 2010
4. Masoud Farzam, Soosan Beheshti, “The noiseless code-length concept in subspace estimation for low SNR hyperspectral signals”, IEEE conference on Communications, Computers and Signal Processing, Victoria, August 2009

Received the Gold paper award in this conference

5. Masoud Farzam, Soosan Beheshti, “A noiseless code length method (NCLM) to estimate dimensionality of hyperspectral data”, IEEE Conference on Acoustic and Signal Processing (ICASSP), Taipei, Taiwan, April 2009
6. Masoud Farzam, Soosan Beheshti, “A fast transition and replacement (FTR) algorithm for fast hyperspectral imaging applications”, Remote08 Conference, Italy, Nov. 2008

Received the best paper award in this conference

7. Masoud Farzam, Soosan Beheshti, “Calculation of the abundance factors in hyperspectral imaging using genetic algorithm”, CCECE conference, Niagara falls, Canada, May 2008

Bibliography

- [1] R. N. Clark, G. A. Swayze, A. Gallagher, T. V. King, and W. M. Calvin, "The U.S. geological survey digital spectral library," *Open File Rep. 93-592, U.S. Geological Survey*, 1993.
- [2] Cytoviva Incorporation, "<http://cytoviva.com>", Last time visted: Jul. 2011.
- [3] ENVI software, "<http://www.itvis.com>", Last time visted: Jul. 2011.
- [4] N. Raksuntorn and Q. Du, "Nonlinear spectral mixture analysis for hyperspectral imagery in an unknown environment," *IEEE Geoscience and Remote Sensing letters*, vol. 7, no. 4, Oct. 2010.
- [5] J. F. Mustard, L. Li, and G. He, "Nonlinear spectral mixture modeling of lunar multispectral data: Implications for lateral transport," *Journal of Geophysical Research*, vol. 103, no. E8, pp. 19419-19425, Aug. 1998.
- [6] A. Moody, S. Gopal, and A. H. Strahler, "Artificial neural network response to mixed pixels in coarse-resolution satellite data," *Remote Sensing Environment*, vol. 58, no. 3, pp. 329-343, Dec. 1996.
- [7] L. Zhang, B. Wu, B. Huang, and P. Li, "Nonlinear estimation of subpixel proportion via kernel least square regression," *International Journal of Remote Sensing*, vol. 28, no. 18, pp. 4157-4172, 2007.
- [8] N. H. Broge and E. Leblancb, "Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density," *IEEE Transactions on Geoscience and Remote Sensing*, vol 41, issue 6, Jun. 2003.
- [9] J. Ellis, "Searching for oil seeps and oil-impacted soil with hyperspectral imagery," *Earth Observation Magazine*, pp. 25-28, Jan. 2001.

- [10] I. B. Strachan, E. Patteyb, and J. B. Boisvert, "Impact of nitrogen and environmental conditions on corn as detected by hyperspectral reflectance," *Journal of Remote Sensing of Environment*, vol. 80, issue 2, pp. 213-224, May 2002.
- [11] L. W. Lass, D. C. Thill, B. Shafii, and T. S. Prather, "Detecting spotted knapweed (*centaurea maculosa*) with hyperspectral remote sensing technology," *Journal of Weed Technology*, 16(2), pp. 426-432, 2002.
- [12] V. E. Brando and A. G. Dekker, "Satellite hyperspectral remote sensing for estimating estuarine and coastal water quality," *IEEE Transactions on Geoscience and Remote Sensing*, 41:1378-87, 2003.
- [13] W. A. Marcusa, C. J. Legleiterb, R. J. Aspinallb, J. W. Boardmanc, and R. L. Crabtree, "High spatial resolution hyperspectral mapping of in-stream habitats, depths, and woody debris in mountain streams," *Journal of Geomorphology*, vol 55, issue 1-4, pp. 363-380, Sep. 2003.
- [14] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 8, Aug. 2004.
- [15] M. Meroni, R. Colombob, and C. Panigada, "Inversion of a radiative transfer model with hyperspectral observations for lai mapping in poplar plantations," *Journal of Remote Sensing of Environment*, vol 92, issue 2, pp. 195-206, Aug. 2004.
- [16] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, issue 6, pp. 1351-1362, Jun. 2005.
- [17] A. Plaza, P. Martinez, J. Plaza, and R. Perez, "Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, issue 3, pp. 466-479, Mar. 2005.
- [18] J. M. P. Nascimento and J. M. B. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898-910, Apr. 2005.

- [19] J. Wang and C.-I. Chang, "Independent component analysis-based dimensionality reduction with applications in the hyperspectral image analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 6, pp. 1586-1600, Jun. 2006.
- [20] B. M. Nicolaia, E. Lotzeb, A. Peirsa, N. Scheerlincka, and K. I. Theron, "Non-destructive measurement of bitter pit in apple fruit using nir hyperspectral imaging," *Journal of Post Harvest Biology and Technology*, vol. 40, issue 1, pp. 1-6, Apr. 2006.
- [21] F. Montmessin, B. Gondet, J.-P. Bibring, Y. Langevin, and P. Drossart, "Hyperspectral imaging of convective co2 ice clouds in the equatorial mesosphere of mars," *Journal of Geophysical Research*, vol. 112, E11S90, 2007.
- [22] L. Miao, H. Qi, and H. Szu, "A maximum entropy approach to unsupervised mixed-pixel decomposition," *IEEE Transactions on Image Processing*, vol. 16, no. 4, pp. 1008-1021, Apr. 2007.
- [23] W. F. J. Vermaas, J. A. Timlin, H. D. T. Jones, M. B. Sinclair, and L. T. Nieman, "In vivo hyperspectral confocal fluorescence imaging to determine pigment localization and distribution in cyanobacterial cells," *Proceeding of National Academy of Science of U.S.*, vol. 105, no. 10, Mar. 2008.
- [24] T. H. Chan, C. Y. Chi, Y. M. Huang, and W. K. Ma, "A convex analysis-aased minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Transactions on Signal Processing*, vol. 57, issue 11, pp. 4418-4432, Nov. 2009.
- [25] M. Marconcini, G. Camps-Valls, and L. Bruzzone, "Hyperspectral signal models and implications to material detection algorithms," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, issue 2, pp. 234-238, Apr. 2009.
- [26] J. Qina, T. F. Burksa, M. A. Ritenourb, and W. G. Bonn, "Detection of citrus canker using hyperspectral reflectance imaging with spectral information divergence," *Journal of Food Engineering*, vol. 93, issue 2, pp. 183-191, Jul. 2009.
- [27] R. T. Kester, L. Gao, N. Bedard, and T. S. Tkaczyk, "Real-time hyperspectral endoscope for early cancer diagnostics," *Proc. SPIE*, vol. 7555, Feb. 2010.
- [28] L. Grosberg, A. J. Radosevich, S. Asfaha, X. Yang, T. C. Wang, and E. M. Hillman, "3-d visualization of intrinsic contrast in neoplastic colon tissue using hyperspectral two-photon microscopy," *Biomedical Optics Conference Proceeding*, Miami, Florida, Apr. 2010.

- [29] D. Farrell, J. Alper, K. Ptak, N. J. Panaro, P. Grodzinski, and A. D. Barker, "Recent advances from the national cancer institute alliance for nanotechnology in cancer," *ACS Nano* 4 (2), pp. 589-594,, Feb. 2010.
- [30] S. Lim, K. H. Sohn, and C. Lee, "Principal component analysis for compression of hyperspectral images," *IEEE Geoscience and Remote Sensing Symposium, IGARSS* 2001.
- [31] A. 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 Transactions on Geoscience and Remote Sensing*, 26, pp. 65-74, 1988.
- [32] J. B. Lee, A. S. Woodyatt, and M. Berman, "Enhancement of high spectral resolution remote sensing data by a noise-adjusted principal components transform," *IEEE Transactions Geoscience and Remote Sensing*, 28 (3), pp. 295-304, 1990.
- [33] G. Vane, R. Green, T. Chrien, H. Enmark, E. Hansen, and W. Porter, "The airborne visible/infrared imaging spectrometer (aviris)," *Remote Sensing Environment*, vol. 44, no. 2/3, pp. 127-143, May/Jun. 1993.
- [34] J. Bowles, J. Antoniadis, M. Baumback, J. Grossman, D. Haas, P. Palmadesso, and J. Stracka, "Real time analysis of hyperspectral data sets using nrls orasis algorithm," *SPIE* 3118, pp. 38-45, 1997.
- [35] M. R. Gupta and N. P. Jacobson, "Wavelet principal componenet analysis and its application to hyperspectral iamges," *IEEE International Conference Proceeding on Image Processing*, Atlanta, Oct. 2006.
- [36] D. L. Donoho, "Denoising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613-627, May 1995.
- [37] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, 81:425-455, 1994.
- [38] I. Atkinson, F. Kamalabadi, and D. L. Jones, "Wavelet-based hyperspectral image estimation," *Proc. of IGARSS Conference Proceeding*, 1993.
- [39] M. E. Wall, A. Rechtsteiner, and L. M. Rocha, "Singular value decomposition and principal component analysis," 7260-4, Springer, ISBN 978-1-4419-1226-8, pp. 1-4020, 2003.

- [40] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaption via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425-455, 1994.
- [41] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol 90, pp.1200-1224, 1995.
- [42] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Transactions on Image Processing*. vol. 9, issue 9, pp. 1532-1546, Sep. 2000.
- [43] E. R. Malinowski, "Theory of error in factor analysis," *Analytical Chemistry*, vol. 49, pp. 606-612, 1977.
- [44] J. Harsanyi, W. Farrand, and C.-I. Chang, "Determining the number and identity of spectral endmembers," *Proceeding of 9th Thematic Conference on Geologic Remote Sensing*, Feb. 1993.
- [45] C.-I. Chang, , and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 3, , pp. 608-619, Mar. 2004.
- [46] J. M. P. Nascimento and J. M. B. Dias, "Hyperspectral subspace identification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 8, pp. 2435-2445, Aug. 2008.
- [47] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, 10(3):626-634, 1999.
- [48] C.-I. Chang and D. C. Heinz, "Constrained subpixel target detection for remotely sensed imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 3, pp. 1144-1159, May 2000.
- [49] D. Manolakis, "Hyperspectral signal models and implications to material detection algorithms," *IEEE International Conference Proceeding on Acoustics, Speech, and Signal Processing (ICASSP)*, 2004.
- [50] M. Pal and G. M. Foody, "Feature selection for classification of hyperspectral data by SVM," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 48, no. 5, May 2010.

- [51] A. Plaza and C.-I. Chang, "Impact of initialization on design of endmember extraction algorithms," *IEEE Transactions on Geoscience and Remote Sensing*, 44, pp. 3397-3407, 2006.
- [52] S. Beheshti and M. A. Dahleh, "A new information-theoretic approach to signal denoising and best basis selection," *IEEE Transactions on Signal Processing*, vol. 48, issue 1, Jan. 2010.
- [53] S. Beheshti and M. A. Dahleh, "Noisy data and impulse response estimation," *IEEE Transactions on Signal Processing*, vol. 58, no.2, pp. 510-521, Feb. 2010.
- [54] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, 19(6), 716-723, 1974.
- [55] B. Aiazzi, L. Alparone, A. Barducci, S. Baronti, and I. Pippi, "Estimating noise and information of multispectral imagery," *Opt. Eng.*, vol. 41, no. 3, pp. 656-668, Mar. 2002.
- [56] L.-R. Gao and B. Zhang, "A new operational method for estimating noise in hyperspectral images," *IEEE Geoscience and Remote Sensing letters*, vol. 5, no. 1, pp. 83-87, Jan. 2008.
- [57] M. Hashemi and S. Beheshti, "Adaptive noise variance estimation in bayesshrink," *IEEE Signal Processing Letters*, vol 17, issue 1, Jan. 2010.
- [58] M. Farzam and S. Beheshti, "A fast transition and replacement algorithm for real time hyperspectral imaging applications," *International Remote Sensing Conference Proceeding (REMOTE'08)*, Nov. 21-23, 2008.
- [59] R. J. Connor, "Concepts of independence for proportions with a generalization of the dirichlet distribution," *Journal of the American Statistical Association*, 64 (325): 194-206, 1969.
- [60] E. R. Malinowski, "Determination of the number of factors and experimental error in a data matrix," *Analytical Chemistry*, vol. 49, pp. 612-617, 1977.
- [61] C.-I. Chang, "Spectral information divergence for hyperspectral image analysis," *Geoscience and Remote Sensing Symposium, IGARSS'99*, 1999.
- [62] D. Landgrebe, "Multispectral data analysis: A signal theory perspective," *Purdue Univ*, West Lafayette, Technical Report, 1998.

- [63] G. A. Swayze, "The hydrothermal and structural history of the cuprite mining district, southwestern nevada: An integrated geological and geophysical approach," *Ph.D. dissertation, Univ. Colorado, Boulder*, 1997.
- [64] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Transactions on Geoscience and Remote Sensing*, 45, pp. 765-777, 2007.
- [65] H. Li and J. H. Michels, "Parametric adaptive signal detection for hyperspectral imaging," *IEEE Transactions on Signal Processing*, vol 54, issue 7, pp. 2704-2715, 2006.
- [66] M. A. G. Izquierdo, M. G. Hernandez, O. Graullera, and J. J. Anaya, "Signal-to-noise ratio enhancement based on the whitening transformation of colored structural noise," *Ultrasonics*, vol 38, issue 1-8, pp. 500-502, Mar. 2000.
- [67] S. Choi and A. Cichoki, "Blind separation of nonstationary sources in noisy mixtures," *Electronic letters*, vol. 36, pp. 848-849, 2000.
- [68] D. Manolakis, R. Lockwood, and T. Cooley, "On the spectral correlation structure of hyperspectral imaging data," *IEEE Geoscience and Remote Sensing Symposium, IGARSS 2008*.
- [69] T. W. Anderson, "An introduction to multivariate statistical analysis," 2003.
- [70] Y. E. Shimabukuro and J. A. Smith, "The least-squares mixing models to generate fraction images derived from remote sensing multispectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 29, pp. 16-20, Jan. 1991.
- [71] C.-I. Chang, T. L. E. Sun, and M. L. G. Althouse, "An unsupervised interference rejection approach to target detection and classification for hyperspectral imagery," *Opt. Eng.*, vol. 37, pp. 735-743, Mar. 1998.
- [72] P. Debba, E. J. M. Carranza, and F. D., "van der meer and alfred stein, abundance estimation of spectrally similar minerals by using derivative spectra in simulated annealing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 12, pp. 3649-3658, Dec. 2006.
- [73] R. S. Roberts, "Characterization of hyperspectral data using a genetic algorithm," *Thirty-Fourth Asilomar Conference Proceeding on Signals, Systems and Computers*, USA, 2000.

- [74] L. Ying, G. Yanfeng, and Z. Ye, "Hyperspectral feature extraction using selective pca based on genetic algorithm with subgroups," *International Conference Proceeding on Innovative Computing, Information and Control, ICIC'06*, 2006.
- [75] J. Rissanen, "Minimum description length denoising," *IEEE Transactions Information Theory*, vol. 46, pp. 2537-2543, 2000.
- [76] M. Wax and T. Kailath, "Detection of signals by information criteria," *IEEE Transactions on Acoustic, Speech, and Signal Process.*, vol. 33, pp. 387-392, Apr., 1985.
- [77] N. N. Kachouie and P. Fieguth, "A combined bayesshrink wavelet-ridgelet technique for image denoising," *IEEE International Conference Proceeding on Multimedia*, pp. 1917-1920, Jul. 2006.
- [78] N. N. Kachouie and P. Fieguth, "A gabor based technique for image denoising," *Canadian Conference Proceeding on Electrical and Computer Engineering*, pp. 980-983, May 2005.
- [79] Hypercube software, "<http://www.agc.army.mil/Hypercube/>", Last time visted: Jul. 2011.
- [80] Spectral Viewer software, "<http://www.geist3d.org>", Last time visted: Jul. 2011.
- [81] Hyperspectral explorer software, "<http://www.microimages.com/getstart/hypanly.htm>", Last time visted: Jul. 2011.
- [82] J. Boardman, "Automating spectral unmixing of aviris data using convex geometry concepts," *Summaries 4th Annu. JPL Airborne Geoscience Workshop*, vol. 1, JPL Pub. 93-26, pp. 11-14, 1993.
- [83] R. H. Yuhas, A. F. H. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (sam) algorithm," *Summaries of the 4th JPL Airborne Earth Science Workshop*, JPL Publ.92-41, pp. 147-149, 1992.