UNSUPERVISED UNMIXING OF HYPERSPECTRAL IMAGERY USING THE CONSTRAINED POSITIVE MATRIX FACTORIZATION

By

Yahya M. Masalmah

A dissertation submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY
in

COMPUTING AND INFORMATION SCIENCE AND ENGINEERING

UNIVERSITY OF PUERTO RICO
MAYAGÜEZ CAMPUS

July, 2007

Approved by:

Miguel Veléz-Reyes, Ph.D President, Graduate Committee	Date
Shawn Hunt, Ph.D Member, Graduate Committee	Date
Nayda Santiago, Ph.D Member, Graduate Committee	Date
Wilson Rivera, Ph.D Member, Graduate Committee	Date
Fernando Gilbes, Ph.D Representative of Graduate Studies	Date
Nestor Rodríguez, Ph.D Program Coordinator	Date

Abstract of Dissertation Presented to the Graduate School
of the University of Puerto Rico in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

UNSUPERVISED UNMIXING OF HYPERSPECTRAL IMAGERY USING THE CONSTRAINED POSITIVE MATRIX FACTORIZATION

By

Yahya M. Masalmah

July 2007

Chair: Miguel Veléz-Reyes, Ph.D
Major Department: Computing and Information Science and Engineering

In hyperspectral imaging, hundreds of images are taken at narrow and contiguous spectral bands providing us with high spectral resolution spectral signatures that can be used to discriminate between objects. In many applications, the measured spectral signature is a mixture of the object of interest, and other objects within the field of view of the sensor. To determine which objects are in the field of view of the sensor, we need to decompose the measured spectral signature in its constituents and their contribution to the measured signal. This research dealt with the unsupervised determination of the constituents and their fractional abundance in each pixel in a hyperspectral image using a constrained positive matrix factorization (cPMF). Different algorithms are presented to compute the cPMF. Tests and validation with real and simulated data show the effectiveness of the method. Application of the approach to environmental remote sensing and microscopic imaging is shown.

Resumen de Disertación Presentado a Escuela Graduada
de la Universidad de Puerto Rico como requisito parcial de los
Requerimientos para el grado de Doctor en Filosofía

UNSUPERVISED UNMIXING OF HYPERSPECTRAL IMAGERY USING THE CONSTRAINED POSITIVE MATRIX FACTORIZATION

Por

Yahya M. Masalmah

Julio 2007

Consejero: Miguel Veléz-Reyes, Ph.D
Departamento: Ciencias e Ingeniería de la Computación y la Información

En imágenes hiperespectralres cientos de imágenes son capturadas en bandas estrechas y contiguas en el espectro proveyendo firmas espectrales a alta resolución que pueden ser utilizadas para discriminar entre objetos. En muchas aplicaciones, la firma espectral medida es una mezcla del objetos de interés y otros objetos en el campo de visión del sensor. Para determinar que objetos estan en el campo de visión del sensor, necesitamos descomponer la firma espectral medida en sus constituyentes y su contribución a la señal medida. Esta investigación trata sobre la determinación no supervisada de los constituyentes y sus abundancias en cada pixel de imágenes hiperespectrales utilizando factorazion de matrices positivas con restriciones (cPMF). Diferentes algoritmos se presentan para el computo del cPMF. Pruebas y validación con datos simulados y reales muestran la efectividad del método. Aplicación del método a percepción remota del medio ambiente y imágenes microscopicas es presentada.

DEDICATION

This work is dedicated to my great wife for her support and help during this period of my life. To my great mother for her prayers and support. Finally, to my beautiful Kids: Sameh, Malaak, and Suleiman.

ACKNOWLEDGMENTS

First and foremost, I extend my thanks to my advisor, Professor Miguel Veléz-Reyes, with whom I have had the pleasure of working as both a Master's and a Ph.D. candidate. We have had countless discussions along the way, he was very helpful and patient. I would also like to thank my Ph.D. committee members, Dra. Nayda Santiago, Dr. Shawn Hunt, and Dr. Wilson Rivera, for their help. I further wish to acknowledge the Gordon Center for Subsurface Sensing and Imaging Systems (CenSSIS) for their financial support during my study. I would like also, to convey my thanks to the LARSIP people for their computer and technical support. Additionally, I would like to thank the Puerto Rico NASA Space Grant for their financial support during the last two years.

Thanks to Dr. R. Lockwood from AFRL at Hanscom AFB, MA for providing the Fort A.P. Hill and the simulated cube Hyperspectral data used in this dissertation.

I would like to thank Dr. Max Diem from the Northeastern University and his group for giving me the opportunity to work in their laboratory during the last summer, and for supplying the biochemical data.

Thanks to my great mother for her support and prayers to finish this thesis. My wife, you were extremely supportive in my work, so thank you very much. I would like to acknowledge my brothers and sisters, for following up my work. I do not forget my kind mother-in-law; I greatly respect your support during my work. Thanks for my friends for their prayers and support.

This work was partially supported by NSF Engineering Research Centers Program under grant EEC-9986821 and a Fellowship from the PR NASA Space Grant Program.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure page

List OF ACRONYMS

HSI	Hyperspectral Imagery.
LMM	Linear Mixing Model.
PP	Projection Pursuit.
PI	Projection Index.
PCA	Principal Component Analysis.
ICA	Independent Component Analysis.
PPI	Pixel Purity Index.
MNF	Minimum Noise Fraction.
VCA	Vertex Component Analysis.
NNLS	Non-Negative Least Squares.
PMF	Positive Matrix Factorization.
NNMF	Non-Negative Matrix Factorization.
ALS	Alternating Least Squares.
cNMF	Constrained Non-negative Matrix Factorization.
FA	Factor Analysis.
SVD	Singular Value Decomposition.
VD	Virtual dimensionality.
KKT	Karush-Kuhn-Tucker.
CPMF	Constrained Positive Matrix Factorization.
NNSTOLS	Non-Negative Sum-To-One Least Squares.
PA	Penalty Approach.
SVDSS	Singular Value Decomposition Subset Selection.
RMS	Root Mean Square.
MSE	Mean Squared Error.
SNR	Signal-to-Noise Ratio.
IR	Infra Red.
HCA	Hierarchical Cluster Analysis.
ML	Maximum Likelihood.
MAP	Maximum A Posterior Probability.
AEP	Abundance Estimation Problem.
CSD	Continuous Significant Dimension.
KG	KaiserGuttman.
USGS	U.S. Geological Survey.

CHAPTER 1

Introduction

1.1 Motivation

Hyperspectral imaging (or imaging spectroscopy) is increasingly being employed as a significant technology in the monitoring and management of terrestrial and aquatic ecosystems. Advantages of this technology include both the qualitative benefits derived from a visual overview (imaging), and more importantly, the quantitative abilities (spectroscopy) for systematic assessment and monitoring. In addition to the spatial and temporal capabilities of traditional remote sensing instruments, hyperspectral instruments offer the spectral resolution necessary to extract multiple layers of information from optically complex environments such as those present in many environmental remote sensing scenarios. One important issue in hyperspectral remote sensing is that, due to design tradeoffs, the spatial resolution of sensors flown nowadays might not be enough to resolve objects of interest in the earth surface. In those cases, the measured spectral signature is the mixture of the spectral responses of the objects present in the field of view of the sensor. The problem of interest is then to determine the objects in the field of view of the sensor and, in some applications, their fractional contribution. In the hyperspectral image processing literature, this is the so-called *unmixing problem*.

The general unmixing problem is an ill-posed problem. Well-posedness is achieved by assuming that all the pixels in the hyperspectral image are generated by mixtures of a limited number of materials (or endmembers). The unmixing problem is then to determine these endmembers and their corresponding fractional area coverage (or abundance). This problem had been approached in different ways as discussed later. In most approaches, the endmembers are determined with significant interaction from the image analyst. We call these approaches supervised unmixing. Our interest in this research is to fully automate the unmixing process by having the machine perform both the endmember selection and the abundance estimation. We call this unsupervised mixing. This research focused on extending the positive matrix factorization (PMF) to solve this problem. PMF has been used in the context of machine learning and factor analysis to solve problems with similar mathematical structure. The particular physical structure of the unmixing problem adds additional physical constraints in the abundances which further complicates the application of the traditional PMF to this problem.

A very important component of this work is to develop algorithms and implementations that will produce results in a timely manner. Hyperspectral images are of large size (several hundreds of Megabytes per scene), and it is of importance, to use this technology in many environmental applications, where the analyst uses data products such as fractional abundance and thematic maps, as an inputs in a decision making process.

1.2 Research Objectives

The main objectives of this research were to develop models and find computationally efficient solution to solve the unsupervised unmixing problem. The specific objectives of this research were:

- To develop algorithms to solve the constrained positive matrix factorization (cPMF).
- To demonstrate the applicability of the developed algorithms for monitoring terrestrial and coastal ecosystems.
- To study different approaches to determine the number of endmembers.
- To study the characteristics of the algorithm for an efficient implementation.
- To validate the developed algorithms using simulated and real hyperspectral imagery.

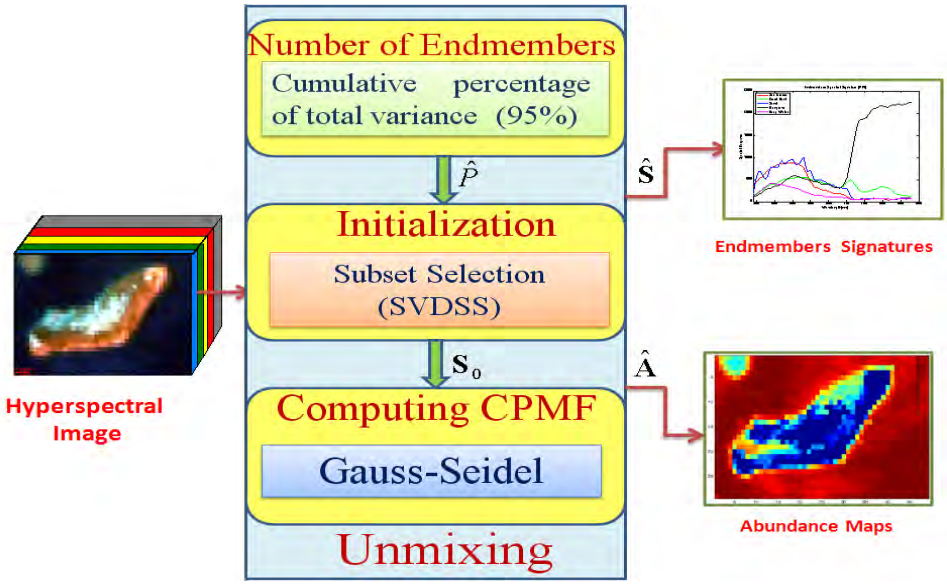


Figure 1–1: Unsupervised unmixing algorithm

1.3 Contributions of the Work

The contributions of this work are as follows. Our first contribution is an unsupervised algorithm to solve the *unmixing problem* in hyperspectral imagery using the constrained positive matrix factorization (cPMF). This algorithm consists of three different steps: *determination of number of endmembers*, *initialization*, and *computing the constrained positive matrix factorization*. The schematic diagram of the algorithm is shown in Figure 1–1.

Second, we have proposed to use the linear dimensionality of hyperspectral imagery as the number of endmembers. The linear dimensionality of the hyperspectral data is the dimension of space spanned by a set of bands. The linear dimensionality of hyperspectral is smaller than the component dimension which is the number of spectral bands.

Third, we have proposed to use the SVD-subset selection approach to initialize the endmembers. This approach selects the most independent pixels from the image as initial endmembers.

Fourth, we showed that cPMF could be used to solve the unsupervised *unmixing problem* by extracting simultaneously, the endmembers, and their abundances, by solving the optimization Problem (2.3). We used the two-stage Gauss-Seidel iterative approach to compute the constrained positive matrix factorization. Unlike other known endmembers extraction algorithms, our algorithm does not assume that endmembers are pure pixels in the image.

We present the evaluation of the effect of number of endmembers, number of spectral bands, and the signal-to-noise ratio on the results. We present analysis for the quality of estimates using the percentage error, and the spectral similarity metrics in simulated data. We present analysis to measure the model fitness using the square of sample correlation ($\mathbf{R}^2$) value, and the root-mean-squared (RMS) residual values. The applicability of the proposed approach to hyperspectral imagery of different domains was shown with environmental and microscopy images demonstrating the diverse problems similar solution vision of the Center for Subsurface Sensing and Imaging Systems. The research combined elements of optimization, linear algebra, and computing to solve several relevant problems.

1.4 Dissertation Outline

This dissertation is structured as follows. Chapter 2 presents a background and previous work on spectral unmixing. Chapter 3 presents a detailed description of Positive Matrix Factorization and its application in unsupervised unmixing. Chapter 4 presents the developed of algorithms to compute the PMF, it also presents the metrics used to evaluate performance. Chapter 5 presents experimental results using simulated data. Chapter 6 presents experimental results using real data. Chapter 7 presents ethical issues. Finally, Chapter 8 presents summary , conclusions, and suggests directions for future work.

CHAPTER 2

Background

This chapter presents background material and review of previous work related to this research.

2.1 Introduction

Materials reflect, absorb, and emit electromagnetic energy, at specific wavelengths, in distinctive patterns related to their composition. Imaging Spectroscopy is the acquisition of images where, for each spatial resolution element or pixel in the image, a spectrum of the energy arriving at the sensor is measured. These spectra are used to derive information based on the signature of the interaction of matter and energy expressed in the spectrum. The imaging spectroscopy concept is illustrated in Figure 2–1 that presents the so-called hyperspectral cube which is the way in which hyperspectral data is usually represented.

Examples of hyperspectral scanners are the 224-band Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) developed by the NASA Jet Propulsion Laboratory [21], the 210-band Hyperspectral Digital Imagery Collection Experiments (HYDICE) [24] developed by the Naval Research Laboratory, and the HYPERION satellite sensor [25] developed by NASA.

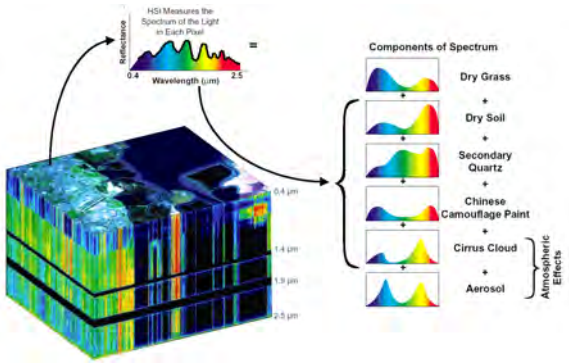


Figure 2–1: Hyperspectral concept illustration

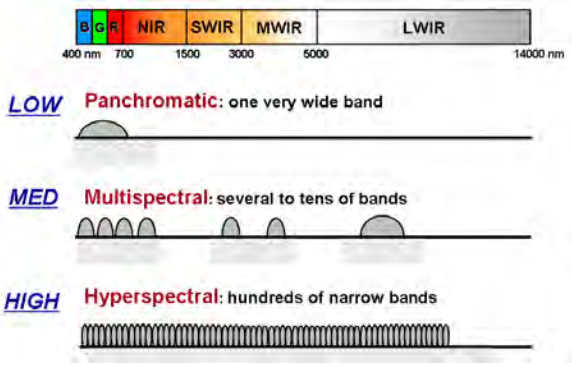


Figure 2–2: Types of spectral sampling in spectral imaging [69].

Multispectral sensors acquire images simultaneously at separate non-contiguous wavelength intervals or bands. Multispectral scanners typically record up to 10, or so, spectral bands with bandwidths on the order of 0.1 μm. More recently, remote sensing imaging technology has advanced in two major ways: improvement in the spatial resolution of images and improvement in the spectral resolution of images. The hyperspectral images samples the spectra at hundreds contiguous wavelengths with a spectral resolution in the order of 1-10 nm. Figure 2–2 illustrates the difference between multispectral and hyperspectral sensing.

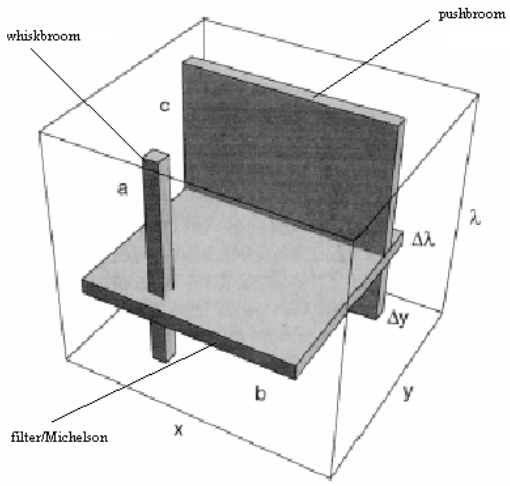


Figure 2–3: Spectral image cube methods [57]

2.2 Building up the Image Cube

It worths to talk briefly about the way the hyperspectral cube is constructed. Most spectral imaging systems build up the cube by scanning through it. There are three conventional methods to scan through the cube. Whiskbroom, filter/ Fourier transform, and pushbroom. An illustration of the three conventional methods is shown in Figure 2–3

We will talk briefly about some of these methods in the following sections.

2.2.1 Whiskbroom Scanning Method

Whiskbroom is a scanning method used in imaging systems to build the spectral cube. It uses a one-dimensional array optical arrangement as shown in Figure 2–4. Examples of whiskbroom spectral sensors are AVIRIS [21], and LANDSAT [66]. The whiskbroom scans in the direction orthogonal to the direction of sensor motion (Cross Track) to obtain the first spatial dimension x. The second spatial dimension, y, is obtained as the sensor is moved across the scene (Along Track) as shown in Figure 2–5. This sensor acquires each pixel in each spectral image with the same

detector array which simplifies calibration . Resulting images from this sensor have relatively low spatial resolution. The artifacts in resulting images are simple and easy to correct [57].

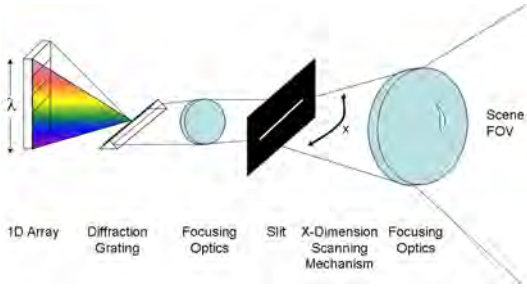


Figure 2–4: One dimensional array arrangement used in whiskbroom scanning method [57]

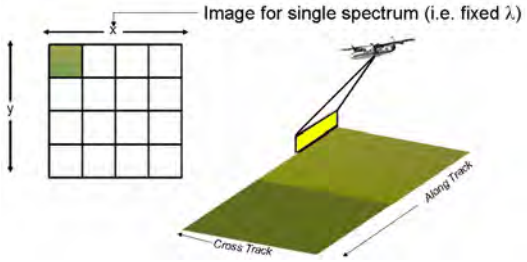


Figure 2–5: Single spectrum image using whiskbroom scanner [57]

2.2.2 Pushbroom Scanning Method

The pushbroom scanner or linear array sensor is a scanner without any mechanical scanning mirror but with a linear array of solid semiconductor elements which enables it to record one line of an image simultaneously, as shown is Figure 2–6. The pushbroom scanner has an optical lens through which a line image is detected simultaneously perpendicular to the flight direction. Though the optical mechanical scanner scans and records mechanically pixel by pixel. The pushbroom scanner scans and records electronically line by line as shown in Figure 2–7.

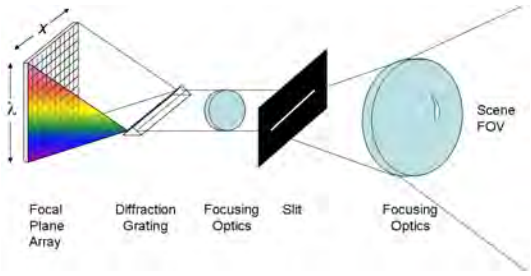


Figure 2–6: Two dimensional array arrangement used in pushbroom scanning method [57]

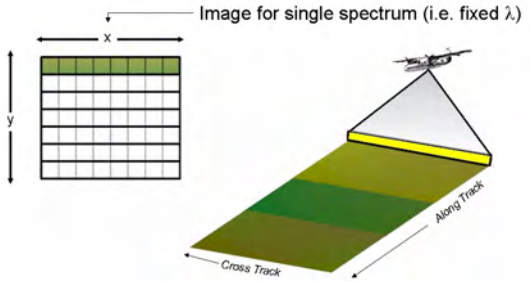


Figure 2–7: Single spectrum image using pushbroom scanner [57]

Pushbroom acquires each pixel of each image line with a different detector which leads to the striping artifacts due to non perfect calibration. Examples of pushbroom sensors are HYDICE [24], and HYPERION [25].

2.3 Linear Mixing Model

Classification is a common and recognized tool in information extraction from remote sensing. To classify a given image, each pixel in the given image is assigned one of several classes. Many classification algorithm assumes that each pixel is pure; that is the response of that pixel is from a single material. This assumption is not always valid since, for example, each pixel produced by the HYPERION instrument covers approximately 900 (30mx30m) square meter area on the ground and many spatial features of interest are smaller than the sensor resolution. Spectral unmixing provides a method to identify elements at subpixel resolution by their contribution to the measured spectral signature. Figure 2–8 illustrates the spatial degradation

caused by low spatial resolution by comparing a visible color composite 1m resolution image of Enrique Reef in Parguera collected using the IKONOS multispectral sensor with a 30m resolution HYPERION image of the same area. The effect of low spatial resolution is evident. Our objective is to derive some of the sub-pixel spatial information from the low spatial resolution by taking advantage of the information available in the spectral dimension.

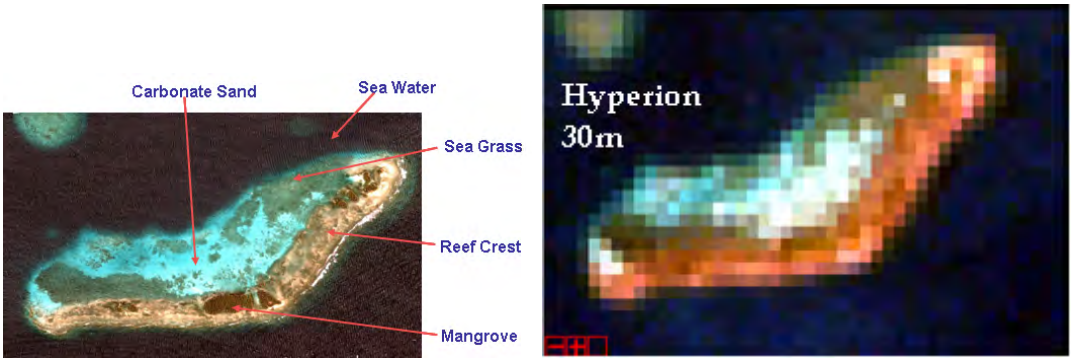


Figure 2–8: Enrique Reef Acquired by (a)IKONOS 1m Resolution Sensor (b)HYPERION 30m Resolution Sensor

The simplest model of a mixed spectrum is a linear one where a single pixel is constructed in a checkboard pattern (see Figure 2–9) of the different materials in the pixel so the measured spectral response of a pixel is a linear combination of the individual materials weighted by the fractional area coverage or abundances. Let $\mathbf{s}_i$ $(1 \leq i \leq P)$ be the i-th endmember corresponding to one of the contributing material spectral signatures, and let $\mathbf{x}_j$ be the observed spectral signature at the j-th pixel. The linear mixing model (LMM) relates the measured spectra $\mathbf{x_j}$ with the endmembers, $\mathbf{s_i}$ as follows

$$\mathbf{x}_j = \sum_{i=1}^{P} \mathbf{s}_i a_{ij} + \mathbf{w}_j = \mathbf{S}\mathbf{a}_j + \mathbf{w}_j, \qquad j = 1, 2, ..., N \tag{2.1}$$

where $\mathbf{S} \in R_+^{m \times P}$ is the matrix of endmembers, $\mathbf{a_j} \in R_+^{P \times 1}$ the vector of fractional abundances, and $\mathbf{w_j}$ is an additive noise term at the j-th pixel and N is the number

of pixels in the image. Furthermore, since the abundances a_{ij} represent fractional area coverage they need to satisfy the following constraints

$$a_{ij} \geq 0 \quad and \quad \mathbf{a}_j^T \mathbf{1}_P = \sum_{i=1}^{P} a_{ij} = 1$$

where $\mathbf{1_P}$ is a vector of 1's of dimension P.

For the entire image, the linear mixture model given in Equation 2.1 above can be written in matrix form as follows

$$\mathbf{X} = \mathbf{SA} + \mathbf{W} \tag{2.2}$$

where $\mathbf{X}$=[$\mathbf{x_1}$,...,$\mathbf{x_N}$], $\mathbf{S}$=[$\mathbf{s}_1$,...,$\mathbf{s}_P$], $\mathbf{A}$=[$\mathbf{a}_1$,...,$\mathbf{a}_N$], and $\mathbf{W}$=[$\mathbf{w}_1$,...,$\mathbf{w}_N$] are a matrix representation for the hyperspectral image, N, and P are the number of pixels, and the number of endmembers , respectively.

Spectral unmixing is the inverse procedure by which given the image $\mathbf{X}$, we want to determine the endmembers matrix $\mathbf{S}$, and the abundance matrix $\mathbf{A}$. Mathematically, this can be expressed in the following approximation problem

$$\hat{\mathbf{S}}, \hat{\mathbf{A}} = \arg \min_{\substack{S_{ij} \geq 0, A_{ij} \geq 0, \\ \mathbf{A}^T \mathbf{1}_p = \mathbf{1}_N}} ||\mathbf{X} - \mathbf{SA}||_F^2 \tag{2.3}$$

where $\|\|_F$ is the Frobenous norm, and $\mathbf{1}_N$ is a vector of 1's of dimension N. If the constraint $\mathbf{A}^T \mathbf{1}_p = \mathbf{1}_N$ was not present, this problem will be similar to the positive matrix factorization problem studied in many areas [10], [11], [12], and [13]. We can think of (2.3) then as a constrained positive matrix factorization (cPMF).

If the endmember matrix $\mathbf{S}$ is known, the problem becomes the abundance estimation problem which is a constrained linear least squares problem. This problem has been solved using different approaches as will be discussed later.

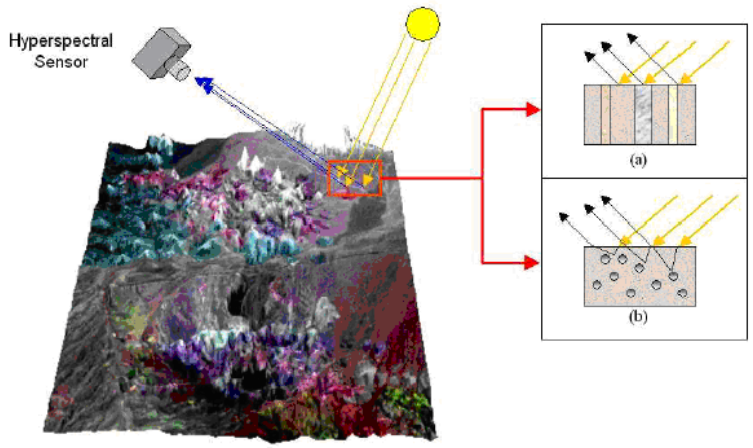


Figure 2–9: Mixed pixel illustration

Here we are interested in determining (i) the endmember matrix **S** and (ii) the corresponding abundances matrix **A**. For example, if both matrices are assumed to be deterministic but unknown then using Equation 2.1 and assuming $\mathbf{w}_j$ to be uncorrelated zero-mean gaussian, we have that Equation 2.3 is the maximum likelihood estimator for **A** and **S**.

2.4 Nonlinear mixing

The linear mixing model discussed in the previous section assumes that the incident radiation bounces only once upon its surface. On the other hand, when the incident radiation can experience reflections with multiple substances, and the aggregate spectrum of reflected radiation may no longer uphold the linear proportions of the constituent substance spectrum, the linear mixing model is inappropriate to describe this interaction, this scenario, is referred to as nonlinear mixing [1]. The nonlinear mixing model performs better than the linear mixing model [50] in some applications. One question arise according to [50]. Why would we consider using a linear mixing model at all to estimate abundance fractions, away from the fact that it is computationally simpler? The answer is related to the scale of spatial resolution. When we work in the laboratory, the scale is millimetric scale; while in remote

sensing application we work with meter scale. According to [50], in the meter scale, the linear mixing model results in a good approximation to the physical reality.

2.5 Solving the Linear Unmixing Problem

Different approaches have been proposed to solve the linear unmixing problem. The standard unmixing approach solves the unmixing in two stages: first estimates endmembers from the image by using one of several approaches described in the literature [6], and then estimates the abundances. Even though most of the approaches use similar mathematical procedures, they differ in the criteria and the assumption on the original model followed by the algorithm. Keshava [1] introduced a taxonomy of the existing approaches. He built the taxonomy upon three main features: output characteristics, input characteristics, and noise modeling. Figure 2–10 shows a three level diagram illustrating the algorithm taxonomy. The first level is the data interpretation. The second level is the data randomness and the third level is the optimization criteria. There are three criteria that categorize existing algorithms [1]: how mixed-pixel spectra is interpreted, how data randomness is incorporated, and the objective function being optimized.

2.5.1 Endmember Determination Approaches

Endmembers determination is very important in hyperspectral data analysis. It is the process to find pure signatures present in image data, which are generally of major interest for image analysis. In [18], they define the endmember as an idealized, pure signature for a class. Therefore, an endmember does not have to be a full pixel in the image. In [18], endmembers are extracted from a sequence of projections generated using projection pursuit (PP). PP is a linear mapping that searches for interesting low-dimensional projections from a high-dimensional data space via a projection index (PI). In particular, it can be designed to characterize

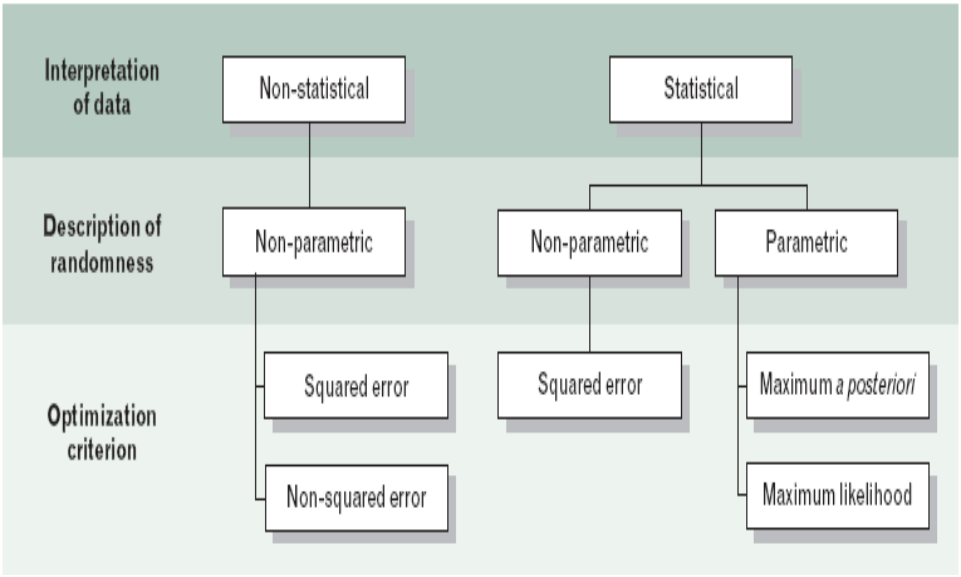


Figure 2–10: Taxonomy of unmixing algorithms [1]

nonlinear structures in projected distributions. For example, if the desired direction of a PI is one pointing to data variance, the PP is reduced to PCA. On the other hand, if the PI measures statistical independence it results in independent component analysis (ICA) [45]. So, PP is a powerful technique in many signal/image-processing applications. Using PP for hyperspectral image classification has been studied previously in [46], who designed a PI based on Bhattacharyya's distance to reduce the dimensionality of feature space and in [47], who used the information divergence (relative entropy) as a PI looking for interesting projections that deviate from Gaussian distributions. the information divergence is used as the PI for the PP approach.

Pixel purity index (PPI) is a well known approach used in endmember determination [19]. PPI finds the most spectrally pure pixels in the image by repeatedly projecting an n-dimensional scatter plot onto random unit vectors. The extreme pixels in each projection are recorded and the total number of times each pixel is marked as extreme is noted. Those Pixels marked as extreme above a certain threshold are declared "pure". To exclude the redundant spectra in the pure pixel list,

the actual endmember spectra are selected by a combination of intelligent review of the spectra themselves and through N-dimensional visualization. This approach requires data preprocessing and post processing before determining the endmembers [27]. PPI is typically run on a principal component transformation or minimum noise fraction (MNF) image excluding noisy bands.

Orasis [27] is another approach for endmember determination. This algorithm does not do dimensionality reduction; instead it thins the data set using Exemplar Selection process in which any redundant spectra are rejected. The rejection is done by calculating the angle between spectral vectors; vectors that are not separated by a threshold angle are thinned from the data. The algorithm then finds a basis set of much lower dimension, than the original data using Gram-Schmidt orthogonalization. The thinned data are then projected onto this basis subspace and a simplex is found through a minimum volume transform [28].

Another set of algorithms is based on a geometric model where pixels are enclosed in a simplex whose corners are the endmembers. The geometrical mixing model provides an intuitive mean to describe the endmember determination problem which is reduced to determine the corners of the simplex that encloses the data cloud [1, 4, 5]. Figure 2–11 shows examples of simplex in the plane.

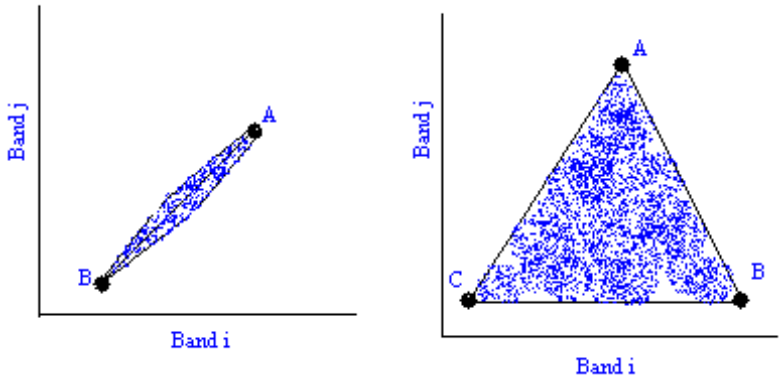


Figure 2–11: Simplex illustration in 2-D

Examples of geometric motivated algorithms are N-FINDR [5] and Vertex Component Analysis (VCA) [6]. N-FINDR [5] attempts to find the simplex of maximum volume that can be inscribed within the hyperspectral data set using a simple non-linear inversion. The non-linear inversion procedure begins with a random set of pixels as endmembers. In order to refine the estimate of endmembers, every pixel in the image must be evaluated as to its likelihood of being a pure or nearly pure pixel. To do this, the volume must be calculated with each pixel in place of each endmember. A trial volume is calculated for every pixel in each endmember position by replacing that endmember and recalculating the volume. If the replacement results in an increase in volume, the pixel replaces the endmember. This is repeated until there are no more replacements. VCA [6] algorithm is unsupervised and exploits two facts: i) the endmembers are the vertices of a simplex and ii) the affine transformation of a simplex is also a simplex. It works with unprojected and with projected data. As PPI and N-FINDR algorithms, 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 found. VCA performs much better than PPI and better than or comparable to N-FINDR; yet it has a computational complexity between one and two orders of magnitude lower than N-FINDR. These algorithms assume that pure pixels are present in the input image and encounter difficulties when no pure pixels exist. In that case, they will find the least mixed pixel (e.g. that most closely approximates the missing endmembers).

Other approaches to endmember determination use clustering to determine the endmembers. For instance, in [7], K-means clustering is used to group the data

and the cluster centroids are used as the endmembers. Notice that the clustering approach does not require the presence of pure pixels in the image.

The mentioned endmember determination approaches discussed in this section are not related to an optimization problem like Equation 2.3. Endmembers are determined separately from their abundances.

2.5.2 Abundances Estimation

If the matrix of endmembers $\mathbf{S}$ is known, the optimization problem developed in (2.3) reduces to the abundance estimation problem (AEP). It can be shown that the cost function in (2.3) is separable in the abundance vectors for each pixel so we can estimate the abundances for each pixel separately. The AEP for a particular pixel is given by the constrained linear least squares problem

$$\hat{\mathbf{a}}_j = \arg \min_{a_{ij} \geq 0, \mathbf{1}_P{}^T \mathbf{a}_j = 1} ||\mathbf{x}_j - \mathbf{S}\mathbf{a}_j||_2^2 \tag{2.4}$$

The objective function in Equation 2.4 and constraints are convex, therefore the solution is unique [3]. For AEP, several solution methods have been proposed in the literature.

The simplest solution presented in the abundance estimation literature is to completely ignore the constraints which lead to an unconstrained linear least squares problem with direct solution given by [3]

$$\hat{\mathbf{a}} = (\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T\mathbf{x} \tag{2.5}$$

The simplest constraint to enforce is the sum to one constraint since a direct solution still exists for this case and is given by [3]

$$\hat{\mathbf{a}}_{STO} = \hat{\mathbf{a}}_{ULS} + (\mathbf{S}^T\mathbf{S})^{-1}\lambda\mathbf{1}_P \tag{2.6}$$

where

$$\lambda = \frac{(1 - \mathbf{1}_P^T \hat{\mathbf{a}}_{ULS})}{\mathbf{1}_P^T (\mathbf{S}^T \mathbf{S})^{-1} \mathbf{1}_P}$$

This is also simple to implement but still has the same disadvantage as ULS in that negative abundances could be obtained.

A direct solution no longer exist when the positive constraints are enforced [3, 4]. The linear least squares problem with positive constraints only is referred in the linear algebra literature as the non-negative linear least squares problem (NNLS) for which Lawson and Hanson's algorithm [3] is the most commonly used algorithm to compute a solution. The NNLS algorithm uses the *Active Set Strategy* presented in [58]. The idea behind the active set strategy is to divide the constraint into two groups: the set acting as active ($\wp$) and the set acting as inactive set (Z) [58]. The set working as inactive, Z, would be ignored to find the solution of LS [58]. The NNLS is an simplification of the Least Square problem with Linear Inequality Constraints (LSI). The NNLS algorithm is described in details by Lawson and Hanson [3]. The convergence of the NNLS algorithm has been shown in [3].

Other iterative algorithms can be used to solve the NNLS. For instance, multiplicative iterative algorithms to solve the NNLS and its application to positively constrained unmixing are described in [4].

Algorithms to solve the fully constrained abundance estimation problem have been proposed in [2, 15]. In [15], an algorithm for the fully constrained abundance estimation problem is proposed where the unmixing problem is transformed to a least distance (LSD) least squares (LS) problem and solved using duality theory as described in [3]. The advantage of this approach is that the resulting dual problem is a NNLS problem for which several algorithms exist for its solution as described previously. In [2], a different approach is used to enforce the constraints by means

of quadratic penalty terms. A nonlinear quadratic penalty term is used in [2] to enforce the nonnegative restriction on the abundances.

2.5.3 Unsupervised Unmixing

We define unsupervised unmixing as an approach to estimate the endmembers and their abundances together by solving an optimization problem (as in (2.3)). A statistical approach is proposed in [8] to solve the unsupervised unmixing. There, abundances and endmembers are modeled as random variables and the maximum a posteriori(MAP) estimator is used. The algorithm in [8] solves the optimization problem

$$\hat{\mathbf{A}}_{MAP}, \hat{\mathbf{S}}_{MAP} = \arg\max f(\mathbf{A}, \mathbf{S}/\mathbf{X}) \tag{2.7}$$

where $f(\mathbf{A}, \mathbf{S}/\mathbf{X})$ is the a posteriori joint distribution for $\mathbf{A}$ and $\mathbf{S}$.

The probabilistic model assumes that the abundances and endmembers are independent random variables. The endmembers are independent for different materials, and the noise term in (2.1) is identically independent normally distributed.

2.6 Summary

This chapter presents a background about hyperspectral imaging. The hyperspectral imaging concept, sensors types, and image formation are introduced. The mixing problem using linear and nonlinear mixing models is presented. Previous work in unmixing is reviewed. The standard two-stage unmixing procedures was discussed in detail. The techniques described in the literature for each stage of the standard unmixing are described.

CHAPTER 3

Unsupervised Unmixing using the Constrained Positive Matrix Factorization

This chapter presents the constrained positive matrix factorization and its application to unsupervised unmixing.

3.1 Positive Matrix Factorization

The Positive matrix factorization (PMF) also known as nonnegative matrix factorization (NNMF), has recently been shown to be a very useful technique in approximating high dimensional data where the data are comprised of positive component. In [9, 11], Lee and Seung proposed the use of the PMF technique to find a set of basis functions to represent image data where the basis functions enable the identification and classification of intrinsic parts that make up the object being imaged by multiple observations. During the past few years, several papers have discussed PMF techniques and successful applications to various fields where the data values are positive [10, 12, 41].

To formulate the Positive Matrix Factorization problem, we need to introduce some notation. The set $R_+ = [0, +\infty)$ is called the set of *positive real numbers*. Let

$Z_+ = \{1, 2, ...\}$ be the set of positive integers, for $n \in Z_+$, let $Z_n = \{1, ..., n\}$. Denote the set of natural numbers $N = \{0, 1, 2, ...\}$. Denote by R_+^k the set of k-tuples of the positive real numbers. The set $R_+^{n \times m}$ is the set of $n \times m$ positive real matrices.

Definition 3.1.1. *Let* $\mathbf{X} \in R_+^{m \times n}$ *for* $m, n \in Z_+$. *If* $\mathbf{X}$=$\mathbf{0}$, *the positive rank of X is defined to be 0. The positive rank of the matrix* $\mathbf{X} \neq 0$ *is defined as the least integer* $P \in Z_+$ *for which there exists a factorization given by Equation 3.1.*

Definition 3.1.2. *(Positive Matrix Factorization): Given a positive matrix* $\mathbf{X} \in R_+^{m \times n}$, *the standard PMF problem is to find two new reduced-dimensional matrices* $\mathbf{S} \in R_+^{m \times P}$ *and* $\mathbf{A} \in R_+^{P \times n}$ *such that*

$$\mathbf{X} = \mathbf{SA} \tag{3.1}$$

The smallest integer P for which such a factorization exists is called the *positive rank* of $\mathbf{X}$ and the corresponding factorization is called a *minimal positive matrix factorization*. The problem is to determine the positive rank of a positive matrix and the corresponding minimal factorization [29]. Finding and exact solution for PMF is an unsolved problem in linear algebra.

Clearly P cannot be smaller than the positive rank of $\mathbf{X}$, but in many cases even larger value of P do not guarantee the existence of an exact PMF. It is therefore, of interest to consider the approximate PMF problem where, given $\mathbf{X}$ and P , one minimizes distance between $\mathbf{X}$ and $\mathbf{SA}$ with respect to positive matrices $\mathbf{S}$ and $\mathbf{A}$. There are different distances that can be used in the minimization process. Later in this chapter, we will discuss two well known distances; the divergence and the Frobenous norm. Mathematically, the approximate positive matrix factorization can be written as

Definition 3.1.3. *(Approximate Positive Matrix Factorization): Given a positive matrix* $\mathbf{X} \in R_+^{m \times n}$*, the approximate PMF problem is to find two new reduced-dimensional matrices* $\mathbf{S} \in R_+^{m \times P}$ *and* $\mathbf{A} \in R_+^{P \times n}$ *such that*

$$\hat{\mathbf{S}}, \hat{\mathbf{A}} = \arg \min_{S_{ij} \geq 0, A_{ij} \geq 0} ||\mathbf{X} - \mathbf{SA}||_F^2 \tag{3.2}$$

This optimization problem always has a solution [36].

3.2 Polyhedral Cones and Positive Matrices

Geometrically, finding the approximate PMF consists of finding the positive rank P and the corresponding P vectors that generate the polyhedral cone in $\mathbf{R}_+^m$ which are the P columns of $\mathbf{S}$ and then the n elements inside it which are the columns of the product $\mathbf{SA}$ that best approximate the n columns of $\mathbf{X}$. To understand the concepts of polyhedral cones and positive matrices, we need the following definitions.

Definition 3.2.1. *Convex set: A set* S *in* R^m *is said to be convex if, for each* $\mathbf{x}_1, \mathbf{x}_2 \in S$*, the line segment* $\lambda \mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2$ *for* $\lambda \in [0, 1]$ *belongs to* S*. Points of the form* $\mathbf{x} = \lambda \mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2$ *for* $\lambda \in [0, 1]$ *are called* convex combinations *of* $\mathbf{x}_1$ *and* $\mathbf{x}_2$*.*

Definition 3.2.2. *Convex polyhedron: is the set of solutions to a finite system of linear inequalities. Hence, a set* C *of vectors in* R *is called a convex polyhedron if:* $C = \{\mathbf{v} | \mathbf{Av} \leq \mathbf{b}\}$ *for some* $n \times m$ *matrix* $\mathbf{A}$ *and a vector* $\mathbf{b}$*.*

Definition 3.2.3. *Polyhedral cone: is the set of solutions to a finite system of linear inequalities:* $C = \{v | Av \leq 0\}$ *for some* $n \times m$ *matrix* $\mathbf{A}$*.*

3.3 Approaches for Computing the Positive Matrix Factorization

PMF can be drawn back to 1994, when Pentti Paatero published his paper [30] introducing PMF. This work was previous to the work of Lee and Seung [9] on Nonnegative Matrix Factorization.

Since the work of Lee and Seung [9], a lot of published and unpublished work has focused on the analysis and application of the PMF in science, engineering, and other fields. The PMF problem had been formulated alternatively based on the application and the corresponding constraints. In [34], PMF was used for document clustering. Nonnegative matrix factorization was used for rapid recovery of constituent spectra in magnetic resonance chemical shift imaging of the brain [13].

The PMF algorithms can be divided into three different general classes [32]. Multiplicative updating rules based algorithms is one class. These algorithms belong to Lee and Seung [10] algorithms. This class uses either the mean squared error or the divergence objective functions. In both cases, the rank of factorization P is assumed known.

The least squares error approximation problem is given by

$$\hat{\mathbf{S}}_{LS}, \hat{\mathbf{A}}_{LS} = \arg \min_{\mathrm{S}\geq 0, \mathrm{A}\geq 0} ||\mathbf{X} - \mathbf{SA}||_F^2 \tag{3.3}$$

An algorithm developed in [10] to solve (3.3) is summarized as in Algorithm 1. where ε is a small positive number added to avoid division by zero.

The divergence objective function problem is given by

$$\hat{\mathbf{S}}, \hat{\mathbf{A}} = \arg \min_{\mathrm{S}\geq 0, \mathrm{A}\geq 0} D(\mathbf{X}||\mathbf{SA}) \tag{3.5}$$

Algorithm 1 :Multiplicative update for solving (3.3)

Input: P, $\mathbf{S}^0$, $\mathbf{A}^0$, ϵ, and set $k = 0$
Output: $\hat{\mathbf{S}}$, $\hat{\mathbf{A}}$: optimized solution

1. Let k=k+1, iterate until the stopping criteria is achieved
2. For all $0 \leq i \leq P$, $0 \leq j \leq N$, and $0 \leq b \leq M$ update the entries of $\mathbf{A}$, and $\mathbf{S}$ using

$$\begin{aligned} \hat{a}_{ij}^{k+1} &= \hat{a}_{ij}^{k}\left(\frac{((\hat{S}^k)^T\mathbf{X})_{ij}}{(((\hat{S}^k)^T((\hat{S})^k\hat{A}^k))_{ij}+\varepsilon)}\right) \\ \hat{s}_{bi}^{k+1} &= \hat{s}_{bi}^{k}\left(\frac{(\mathbf{X}(\hat{A}^{k+1})^T)_{bi}}{((\hat{S}^k\hat{A}^{k+1}(\hat{A}^{k+1})^T)_{bi}+\varepsilon)}\right) \end{aligned} \tag{3.4}$$

where

$$D(\mathbf{X}||\mathbf{SA}) = \sum_{ij}\left(X_{ij}\log\frac{X_{ij}}{(SA)_{ij}} - X_{ij} + (SA)_{ij}\right)$$

The corresponding updating rules are given by an algorithm developed in [10] to solve (3.5) and summarized as in Algorithm 2.

Algorithm 2 :Multiplicative update: minimizing the divergence (3.5)

Input: P, $\mathbf{S}^0$, $\mathbf{A}^0$, ϵ, and set $k = 0$
Output: $\hat{\mathbf{S}}$, $\hat{\mathbf{A}}$: optimized solution

1. Let k=k+1, iterate until the stopping criteria is achieved
2. For all $0 \leq i \leq P$, $0 \leq j \leq N$, and $0 \leq b \leq M$ update the entries of $\mathbf{A}$, and $\mathbf{S}$ using

$$\begin{aligned} \hat{a}_{ij}^{k+1} &= \hat{a}_{ij}^{k}\left(\frac{\sum_{r=1}^{M}\hat{s}_{ri}^{k}x_{rj}/(\hat{S}^k\hat{A}^k)_{rj}}{\sum_{h=1}^{M}\hat{s}_{hi}^{k}}\right) \\ \hat{s}_{bi}^{k+1} &= \hat{s}_{ij}^{k}\left(\frac{\sum_{r=1}^{N}\hat{a}_{jr}^{k+1}x_{ir}/(\hat{S}^k\hat{A}^{k+1})_{ir}}{\sum_{h=1}^{N}\hat{a}_{ih}^{k+1}}\right) \end{aligned} \tag{3.6}$$

Equations (3.4) and (3.6) are called multiplicative updating rules because at each iteration the new parameter value is being updated by multiplying the current value by a factor as shown in both Equations (3.4) and (3.6) for the minimum squared error and divergence objective functions, respectively.

These algorithms have become a baseline against which the newer algorithms are compared. Since the objective is convex in either $\mathbf{S}$ or $\mathbf{A}$ but not in both. The

main contribution of [10] was the development of the updating rules that converge to a local minima.

The second class of NMF algorithms are based on the gradient descent methods. This class of algorithms update the S and A matrices in each iteration as follows:

$$\begin{aligned} A_{ij}^{k+1} &= A_{ij}^{k} - \eta_{A_{ij}} \frac{\partial J}{\partial A_{ij}^{k}} \\ S_{ij}^{k+1} &= S_{ij}^{k} - \eta_{S_{ij}} \frac{\partial J}{\partial S_{ij}^{k}} \end{aligned} \tag{3.7}$$

where J is the objective function, η_A and η_S are the step size parameters.

The step size parameters can be chosen in different ways according to the algorithm. Some algorithms initially set these step size values to 1, then keep changing at each iteration by dividing by 2; this is also known as step halving method [33]. This type of algorithms does not guarantee that **S** and **A** entries to be nonnegative. To overcome the negative value entries a projection step is added [16, 33–35, 60]. The projection step updates the elements of **S** and **A** by setting all the negative elements to 0. Mathematically, this can be written as

$$A_{ij}^{k+1} = \max\{0, A_{ij}^{k} - \eta_{A_{ij}} \frac{\partial J}{\partial A_{ij}}\}$$

This class of algorithms is step size dependent; the convergence also depends on step size selection. On the other hand, projecting the results in each iteration to nonnegative values, makes the analysis even more difficult.

The last class of PMF algorithms are the alternating least squares (ALS) class. This class of algorithms structure alternates between two least squares steps. These algorithms were used first by Paatero [30]. The algorithms take advantage of the

fact that the problem has bilinear structure, that is for fixed $\mathbf{A}$ the problem is linear in $\mathbf{S}$, and for fixed $\mathbf{S}$ the problem is linear in $\mathbf{A}$.

This ALS algorithm structure at each iteration is summarized in the following steps:

1. Solve for $\mathbf{A}$ using $(\mathbf{S}^T\mathbf{S})\mathbf{A} = \mathbf{S}^T\mathbf{X}$.
2. Set all negative elements in $\mathbf{A}$ to 0. (This is a projection step.)
3. Solve for $\mathbf{S}$ using $(\mathbf{A}\mathbf{A}^T)\mathbf{S}^T = \mathbf{A}\mathbf{X}^T$.
4. Set all negative elements in $\mathbf{S}$ to 0. (This is a projection step.)

The algorithms keep alternating between the previous steps until convergence is achieved. These algorithms are different from the multiplicative ones in that they are more flexible in changing the path to the fixed point. The multiplicative algorithms have the drawback that once an element in $\mathbf{S}$ or $\mathbf{A}$ becomes 0, it stays at 0. This means that once the algorithm starts heading down a path towards a fixed point, it must continue even if it is a poor path [31]. The ALS algorithms can be very fast and requires less work than other PMF algorithms.

The ALS algorithm is a coordinate descent Gauss-Seidel type. It is similar to the two-stage Gauss-Seidel algorithm we proposed to solve the unsupervised unmixing problem. The two-stage Gauss-Seidel algorithm will be introduced later.

3.4 Multiplicative Updating Rules Convergence Issues

In this section, we present some convergence issues concerning the multiplicative updating rules that optimize the least squares error objective function problem (3.3) and the divergence objective function problem (3.5).

3.4.1 Least Squares Error Objective Function

The algorithm to optimize the least squares error objective function is shown in Algorithm 1. According to [10], the Euclidean distance is invariant under the given updates if and only if $\mathbf{S}$ and $\mathbf{A}$ are at a stationary point of the distance (see [10] for proof).

Theorem 1. *If* $\mathbf{X}$ *has neither zero column nor row, and* $s_{bi}^0 > 0$ *and* $a_{ij}^0 > 0$ $\forall i, j, b$, *then*

$s_{bi}^0 > 0$ *and* $a_{ij}^0 > 0$ $\forall i, j, b$, $\forall k \geq 1$.

By definition, $(\mathbf{S}, \mathbf{A})$ *is a stationary point of (3.3) if it satisfies the Karush-Kuhn Tucker(KKT) optimality conditions given in [59]as* $s_{ij} \geq 0$ *and* $a_{ij} \geq 0$

$$\begin{aligned} &(\nabla_{\mathbf{S}}(||\mathbf{X} - \mathbf{SA}||_F^2))_{bi} \geq 0, (\nabla_{\mathbf{A}}(||\mathbf{X} - \mathbf{SA}||_F^2))_{ij} \geq 0 \\ &s_{bi}.(\nabla_{\mathbf{S}}(||\mathbf{X} - \mathbf{SA}||_F^2))_{bi} = 0, \mathrm{a_{ij}}.(\nabla_{\mathbf{A}}(||\mathbf{X} - \mathbf{SA}||_F^2))_{ij} = 0, \forall i, j, b \end{aligned} \tag{3.8}$$

where

$$\nabla_{\mathbf{S}}(||\mathbf{X} - \mathbf{SA}||_F^2) = (\mathbf{SA} - \mathbf{X})\mathbf{A}^T$$

and

$$\nabla_{\mathbf{A}}(||\mathbf{X} - \mathbf{SA}||_F^2) = \mathbf{S}^{\mathrm{T}}(\mathbf{SA} - \mathbf{X})$$

are respectively partial derivatives to elements in $\mathbf{S}$ *and* $\mathbf{A}$.

In [10], the following properties are shown.

If $f(\mathbf{S}, \mathbf{A}) = ||\mathbf{X} - \mathbf{SA}||_F^2$ then

1. The function value is non-increasing after every update: $f(\mathbf{S}^k, \mathbf{A}^{k+1}) \leq f(\mathbf{S}^k, \mathbf{A}^k)$ and $f(\mathbf{S}^{k+1}, \mathbf{A}^{k+1}) \leq f(\mathbf{S}^k, \mathbf{A}^{k+1})$

Algorithm (1) intends to have a fixed point update : if $a_{ij}^{k+1} = a_{ij}^{k} = 0$ and $((\mathbf{S}^k)^T\mathbf{S}^k\mathbf{A}^k)_{ij} \neq 0$, then $((\mathbf{S}^k)^T\mathbf{X})_{ij} = ((\mathbf{S}^k)^T\mathbf{S}^k\mathbf{A}^k)$ which implies that $(\nabla_{\mathbf{A}} f(\mathbf{S}^k, \mathbf{A}^k))_{ij} = 0$, which is part of the KKT conditions given in (3.8).

3.5 Using PMF to Solve Unsupervised Unmixing

The PMF was used as a base for the development of a constrained non-negative matrix factorization (cNMF) algorithm for recovering constituent spectra in [13]. There, the multiplicative method of [9] was extended to include a constraint on the minimum amplitude of the recovered spectra to enable the algorithm to deal with observations having negative values by assuming they arise from noise distribution. This algorithm does not enforce the sum-to-one on the abundances as required by the linear mixing model. Our research work focused on developing an algorithm that enforces the positive constraints on the enmembers and on the abundances and the sum to one constraint on the abundances.

Factor analysis (FA) and principal component analysis (PCA) are other methods for data representation in lower rank, which are, in principle, also capable of doing low dimensional data representation similar to PMF, when applied to a matrix of spectra. The results of these methods are, however, often ambiguous and difficult to interpret, since the basic factors and principal components may have negative entries. The physical meaningful representation can be found only after a series of transformations which are called rotations.

By requiring non-negativity for both the endmembers and the weights, PMF is able to produce results which should have physical interpretation. Another important aspect of PMF is the optimal use of error estimates. PMF computes the solution by minimizing the least squares error of the fit weighted with the error estimates. Positive rank of a matrix is the minimum number for which the matrix

factorization does exist [29]. The number of endmembers can be considered the same as the positive rank.

3.6 Estimating Number of Endmembers

An important issue in endmember determination is how to select the desired number of endmembers. If the number is too small, not all endmembers present in the image can be extracted. On the other hand, if the number is too large, some extracted endmembers may not be pure pixels.

Estimating the correct number of endmembers is an important issue in the solution of the unmixing problem. In the case of the PMF, the number of endmembers determines the dimension of the factorization. Different approaches can be used to estimate this number. In the literature, the positive rank of a matrix is defined as the minimum dimension for which the positive matrix factorization does exist [29]. If we consider the positive rank to be equal to the number of endmembers, then we need to determine its values. In this section, we present some approaches to estimate the number of endmembers.

3.6.1 Dimensionality Estimation

An approach to estimate the number of endmembers is to use the rank of the image matrix. Dimensionality estimation approaches can be divided into linear and nonlinear approaches. In [65], the dimensionality estimation approaches were presented. Examples of methods for linear dimensionality estimation are PCA [17], correlation PCA, Scree test, and continuous significant linear dimensionality method (CSD). Examples of nonlinear dimensionality estimation approaches are the near neighbor method (Pettis), Fukunaga and Olsen, and fractal dimension estimation

method . The linear and nonlinear dimensionality estimation approaches mentioned here, are presented in more details in [65].

3.6.2 Mean Squared Error Curve

Another approach to estimate the number of endmembers is by plotting the modeling error as a function of number of endmembers, we can estimate the number of endmembers as the after which the error does not further decrease significantly.

3.6.3 Virtual Dimensionality

In [20], the virtual dimensionality (VD) criterion is used to estimate the number of endmembers. In this method, the eigenvalues from both sample correlation matrix and sample covariance matrix are calculated, referred to as correlation-eigenvalues and covariance-eigenvalues for each of spectral bands. A binary composite hypothesis testing problem can be formulated with the null hypothesis representing the case that correlation-eigenvalue is equal to its corresponding covariance-eigenvalue and the alternative hypothesis corresponding to the case that the correlation-eigenvalue is greater than its corresponding convariance-eigenvalues. By specifying a false alarm probability, P_F, a Neyman-Pearson detector can be further derived to determine whether or not a distinct signature is present in each of spectral bands. How many times the alternative hypothesis is true is an estimate of how many endmembers are there.

In this research work, we seek to study linear and nonlinear dimensionality estimation approaches, and the mean squared error curve to estimate the number of endmembers.

3.7 Optimality Conditions for the CPMF

The PMF and cPMF are constrained optimization problems. The method of Lagrange multipliers is a powerful tool to analyze constrained optimization problems [38]. The most important results used in this work are summarized next. Suppose that we have the following constrained optimization problem:

$$\begin{aligned} &\min f(x) \\ &subject \quad to \\ &g(x) = 0 \\ &h(x) \geq 0 \end{aligned} \tag{3.9}$$

then to find the solution of the constrained optimization function, we need to minimize the related Lagrangian function given by

$$L(x, \lambda, \mu) = f(x) + \lambda^T g(x) - \mu^T h(x) \tag{3.10}$$

where $\lambda \geq 0$ and $\mu \geq 0$ are called Lagrange multipliers.

Theorem 2. *(Karush-Kuhn-Tucker(KKT) Theorem). Let* $\mathbf{x}^*$ *be a local minimum for the constrained optimization problem defined in (3.9), then, there exist* λ^* *and* μ^* *such that*

1. $\mu^* \geq 0$;
2. $\nabla(f(\mathbf{x}^*)) + \lambda^{*T}\nabla g(\mathbf{x}^*) - \mu^{*T}\nabla h(\mathbf{x}^*) = \mathbf{0}^T$;
3. $\mu_j^* h(\mathbf{x}_j^*) = 0$.

The KKT conditions for optimality are given by [38]

1. $\mu_j^* \geq 0$;
2. $\nabla(f(\mathbf{x}^*)) + \lambda^{*T}\nabla g(\mathbf{x}^*) - \mu^{*T}\nabla h(\mathbf{x}^*) = \mathbf{0}^T$;

3. $\mu_j^* h_j(\mathbf{x}^*) = 0$;
4. $g(\mathbf{x}^*) = \mathbf{0}$;
5. $h(\mathbf{x}^*) \geq 0$.

Proof: See [38].

Using the vec operator and the kronecker product notation, (see Appendix A) we can re-arrange the PMF and cPMF problems in the framework of (3.9). Let us define $\bar{\mathbf{x}} = vec(\mathbf{X})$, $\bar{\mathbf{s}} = vec(\mathbf{S})$, and $\bar{\mathbf{a}} = vec(\mathbf{A})$, then, the cPMF Problem 2.3 can be written as

$$\begin{aligned} f(\mathbf{S}, \mathbf{A}) &= ||\bar{\mathbf{x}} - \mathbf{M}\bar{\mathbf{a}}||_2^2 = ||\bar{\mathbf{x}} - \mathbf{H}\bar{\mathbf{s}}||_2^2 \\ g_1(\mathbf{A}) &= \mathbf{B}\bar{\mathbf{a}} - \mathbf{1}_N = \mathbf{0} \\ g_2(\mathbf{A}) &= \bar{\mathbf{a}} \geq 0 \\ g_3(\mathbf{S}) &= \bar{\mathbf{s}} \geq 0 \end{aligned} \tag{3.11}$$

where

$$\begin{aligned} \mathbf{M} &= \mathbf{I}_N \otimes \mathbf{S} \\ \mathbf{B} &= \mathbf{I}_N \otimes \mathbf{1}_P^T \\ \mathbf{H} &= \mathbf{A}^T \otimes \mathbf{I}_m \end{aligned} \tag{3.12}$$

The Lagrangian function for the cPMF is given by

$$L(\mathbf{S}, \mathbf{A}, \lambda, \xi, \gamma) = ||\bar{\mathbf{x}} - \mathbf{M}\bar{\mathbf{a}}||_2^2 + \lambda^T(\mathbf{B}\bar{\mathbf{a}} - \mathbf{1}_N) - \xi^T\bar{\mathbf{a}} - \gamma^T\bar{\mathbf{s}} \tag{3.13}$$

The KKT optimality conditions for the cPMF are given by

$$\mathbf{H}^T(\mathbf{H}\bar{\mathbf{s}} - \bar{\mathbf{x}}) = \gamma \tag{3.14a}$$

$$\mathbf{M}^T(\bar{\mathbf{x}} - \mathbf{M}\bar{\mathbf{a}}) = \mathbf{B}^T\lambda - \xi \tag{3.14b}$$

$$\lambda^T(\mathbf{B}\bar{\mathbf{a}} - \mathbf{1}_N) = \mathbf{0} \tag{3.14c}$$

$$\bar{\mathbf{a}}_i \geq 0 \tag{3.14d}$$

$$\bar{\mathbf{s}}_j \geq 0 \tag{3.14e}$$

$$\mathbf{B}\bar{\mathbf{a}} = \mathbf{1}_N \tag{3.14f}$$

$$\xi_i\bar{\mathbf{a}}_i = 0 \tag{3.14g}$$

$$\gamma_j\bar{\mathbf{s}}_j = 0 \tag{3.14h}$$

The KKT optimality conditions for the positive matrix factorization were presented by [16] and given by

$$\begin{aligned}(\mathbf{X} - \mathbf{SA})\mathbf{A}^\mathbf{T} &\leq 0 \\ \mathbf{S}^\mathbf{T}(\mathbf{X} - \mathbf{SA}) &\leq 0\end{aligned} \tag{3.15}$$

3.8 Summary

This chapter presents the Positive Matrix Factorization. We discuss the geometrical interpretation of the PMF and it relation with convex cones. We present how the PMF is modified to solve the unsupervised unmixing problem. The KKT conditions for the cPMF are presented. In addition, we present prior work on the determination of the number of endmembers.

CHAPTER 4

Computing PMF

4.1 Introduction

Our research developed a solution to the unsupervised unmixing problem using the constrained Positive Matrix Factorization (cPMF). This is an optimization problem with equality and inequality constraints. Different approaches can be used to solve this optimization problem.

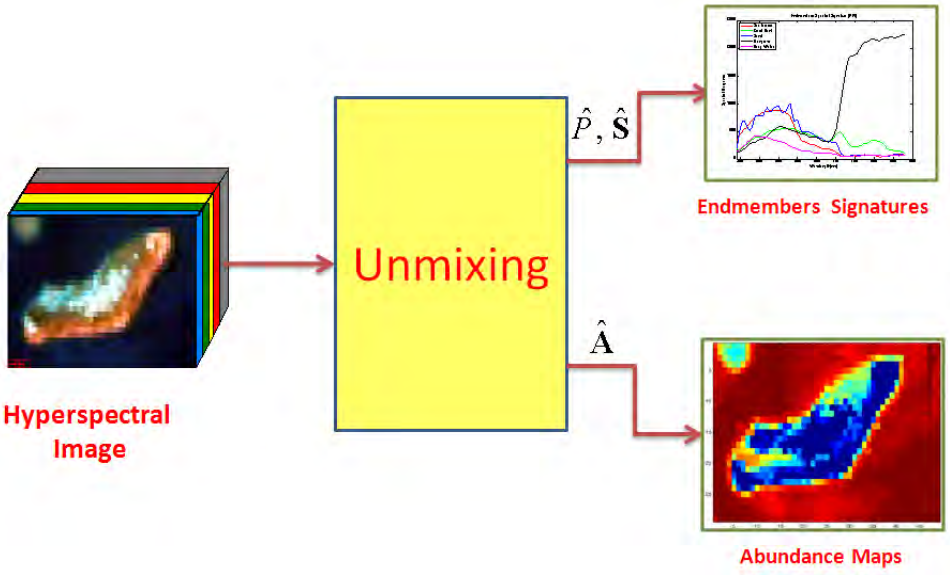


Figure 4–1: Unmixing process

In this chapter, we present different approaches used in this work to compute the PMF to solve the unsupervised unmixing problem. The basic unsupervised unmixing concept is illustrated in Figure 4–1

4.2 Algorithms

Different algorithms can be used to solve the CPMF problem. As a nonlinear least squares problem, we can solve it using the Gauss-Newton method [22] however, because of its structure and large size of data sets being processed other approaches such as Gauss-Seidel, multiplicative iterative methods, and penalty approaches are easier to implement.

4.2.1 Gauss-Seidel Method

This method solves the optimization problem by decomposing the estimation problem and solving it into stages. Let (α, β) be a partitioning of the parameter vector and let

$$(\hat{\alpha}, \hat{\beta}) = \arg\min \mathbf{V}(\alpha, \beta) \tag{4.1}$$

be the optimization problem of interest.

Define the component mappings

$$\hat{\alpha} = \arg\min_{\alpha} \mathbf{V}(\alpha,\beta) = \mathbf{V}_1(\beta) \tag{4.2a}$$

$$\hat{\beta} = \arg\min_{\beta} \mathbf{V}(\alpha,\beta) = \mathbf{V}_2(\alpha) \tag{4.2b}$$

The bilinear structure of (3.3) defines a natural partitioning of the parameter vector with $\alpha = vec(\mathbf{S})$, $\beta = vec(\mathbf{A})$. The Gauss-Seidel algorithm alternate iteratively between two steps. In the first step (4.2a), the abundance matrix will be fixed and the endmember matrix is estimated. Notice that this is a nonnegatively constrained linear least squares problem, NNLS described previously. Multiplicative iterative rules [10] are used for endmembers update. In the second step (4.2b), the endmember matrix will be fixed and the corresponding abundances will be estimated using the Non-Negative Sum-To-One Least Squares (NNSTOLS) proposed in [3]. NNSTOLS is a fully constrained abundance estimation algorithm. The method alternates between the two steps until convergence achieved.

4.2.2 Penalty Approach

The penalty approach is a well known approach to solve optimization problem with equality and inequality constraints [38]. This method approximates the constrained optimization problem (3.3) by a positively constrained optimization problem. The positively constrained optimization problem is solved and the solution is used as an approximation to the solution of the original problem. The idea behind the penalty approach is that, for each constraint in the original problem, a penalty parameter and penalty function are added to the cost function. The penalty function should satisfies the following conditions:

1. It is continuous.
2. It is nonnegative.
3. It is zero when the argument is feasible, i.e. when the argument satisfies the constraints.

The role of the penalty function is to penalize deviations from the feasible set. Therefore, it is natural that the penalty function be defined in terms of the constraint

Algorithm 3 :

Gauss-Seidel Method

- Input P, $\mathbf{S}^0$, $\mathbf{A}^0$, ϵ, and set $k = 0$
- Let k=k+1, iterate until the stopping criteria is achieved
  1. Estimate $\mathbf{A}$

     For i=1: Number of pixels

     $$\hat{a}_i^k = \underset{\substack{a_{ij}\geq 0 \\ \mathbf{a}_i^T \mathbf{1}_p = 1}}{\arg\min} ||\mathbf{X}_i - \hat{\mathbf{S}}^{(k-1)} \hat{\mathbf{a}}_i||_2^2$$

  2. Update

     $$\hat{\mathbf{S}}^{(k)} = \text{argmin} \left\| \mathbf{X} - \hat{\mathbf{S}} \hat{\mathbf{A}}^{(\text{k})} \right\|_F^2$$

     we can use multiplicative updating rules of Lee and Seung [9]

     For i=1:number of bands

     For j=1:P

     $$\hat{S}_{ij}^{(k)} = \hat{S}_{ij}^{(k-1)} \frac{(X \hat{A}^{T(k)})_{ij}}{(\hat{S}^{(k-1)} \hat{A}^{(k)} \hat{A}^{T(k)})_{ij}}$$

     End

     End

  3. If the stopping criteria is satisfied then stop and return $\hat{\mathbf{S}}^{(k)}$, $\hat{\mathbf{A}}^{(k)}$.

functions. Consequently, to enforce the constraints introduced in Equation 3.3, the objective function for our optimization problem should include a penalty terms in addition to Equation 3.3. The modified cost function becomes

$$J(S, A) = ||\mathbf{X} - \mathbf{SA}||_F^2 + \lambda_1 J_1(\mathbf{A}) + \lambda_2 J_2(\mathbf{S}) \tag{4.3}$$

where $J_1(\mathbf{A})$ and $J_2(\mathbf{S})$ are penalty terms used to enforce the constraints on the solution of Equation 3.3, and their corresponding weights are λ_1 and λ_2 used to weight the degree of compliance against model fit. Different penalty terms may be used depending upon the desired effects on the computed solution. An approximate solution to the original constrained problem is obtained by minimizing (4.3). The degree of model fitting and constraint compliance is dependent on the values of λ_1 and λ_2. The gradient of J with respect to the entries of $\mathbf{A}$ and $\mathbf{S}$ are given by,

$$\begin{aligned} \frac{\partial J}{\partial S_{ij}} &= \frac{\partial ||X - SA||_F^2}{\partial S_{ij}} + \lambda_2 \frac{\partial J_2(S)}{\partial S_{ij}} \\ \frac{\partial J}{\partial A_{ij}} &= \frac{\partial ||X - SA||_F^2}{\partial A_{ij}} + \lambda_1 \frac{\partial J_1(A)}{\partial A_{ij}} \end{aligned} \tag{4.4}$$

An approximate solution to the cPMF problem is obtained by solving the positively constrained problem

$$\hat{A}_{PA}, \hat{S}_{PA} = \arg \min_{A \geq 0, S \geq 0} J(S, A) \tag{4.5}$$

where

$$J(\mathbf{S}, \mathbf{A}) = ||\mathbf{X} - \mathbf{SA}||_F^2 + \lambda_1^2 \left\| \mathbf{A}^{\mathrm{T}} \mathbf{1}_P - \mathbf{1}_N \right\|_2^2 = \left\| \begin{bmatrix} \mathbf{X} \\ \lambda_1 \mathbf{1}_N^{\mathbf{T}} \end{bmatrix} - \begin{bmatrix} \mathbf{S} \\ \lambda_1 \mathbf{1}_P^{\mathbf{T}} \end{bmatrix} \mathbf{A} \right\|_F^2 \tag{4.6}$$

and $\mathbf{1}_N$ is an N-dimensional vector of 1's, $\left\|\mathbf{A}^{\mathrm{T}}\mathbf{1}_P - \mathbf{1}_N\right\|_2^2$ is a penalty term that penalizes deviation from sum-to-one constraint, and λ_1, is a penalty parameter.

For a given λ_1, Problem 4.5 is a positively constrained nonlinear least squares problem. The larger λ_1 is the more weight to meet the sum-to-one constraint over minimizing the fitting error.

4.2.3 Gauss-Newton Approach

Since (3.3) is a nonlinear least squares problem, we can solve it using the Gauss-Newton approach. To solve our optimization problem using the Gauss-Newton approach, we need to rewrite the problem in vector form using the properties of the vec operator (see Appendix A). The resulting equations are of large size due to the Kronecker product (see Appendix A). The memory management and matrix size limitation in MATLAB makes it difficult to implement the Gauss-Newton approach.

4.3 Initialization

As with any iterative algorithm, the PMF algorithms are sensitive to the initialization of $\mathbf{S}$ and $\mathbf{A}$ [39]. Different algorithms to solve PMF are presented in the literature [40, 44]. Some algorithms require initialization of both factors $\mathbf{S}$ and $\mathbf{A}$. Other algorithms required the initialization of $\mathbf{S}$ only; using the initial $\mathbf{S}$, $\mathbf{A}$ can be found using known least squares algorithms. Good initialization has a direct impact on the speed and accuracy of the algorithms.

For initialization of cPMF we used the work of Langville et al. [39] which presents procedures for the initialization PMF and applicable as well. We also introduced a new approach developed in this work based on the SVD-subset selection method of [43]. These procedures are summarized next.

1. Random initialization is the procedure by which **S** and **A** are initialized as dense matrices of random number between 0 and 1. Both **S** and **A** should be in the feasible region; i.e. **S** and **A** satisfy the constraints given in Equation 2.3. One drawback of this procedure is that zero entries could be obtained in the initial matrices. For the multiplicative updating rules presented in [9, 10], entries of **S** initialized with zero will stay equal zero even if no minima is achieved.
2. A second initialization procedure is the centroid initialization. In this procedure, **S** is chosen to be the centroids that resulted from clustering the data matrix **X**. This procedure is an expensive initialization step.
3. The third initialization procedure is the singular value decomposition (SVD) - centroid initialization. In this approach , the SVD for the image **X** is computed

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \tag{4.7}$$

where, **U** and **V** are orthonormal matrices and $\mathbf{\Sigma}$ is the diagonal matrix of singular values [42]. Let k be the number of the most significant principal components. The SVD-centroid initialization procedure initializes **S** with a centroid decomposition of the first k columns of the right singular vectors matrix **V**, $\mathbf{V}_k$. This procedure is less time-consuming than the centroid decomposition $\mathbf{X} \in R_+^{m \times N}$
4. Selection of the most P- independent signatures of the image. Notice that spectral signatures are the columns of **X**. A procedure based on the SVD to find such subset is described in [42]. This has been applied to hyperspectral band subset selection in [43] we modified to select the most independent signatures. We will refer to this as SVDSS initialization.
5. Another initialization procedure is Random Acol [39] which initializes each column of **S** by averaging r columns of **X** selected at random. This procedure lies between random and centroid initializations procedures in terms of performance.

6. The last two initialization procedures are the random C initialization and the co-occurrence initialization. The random C initialization is similar to the random Acol method, except it chooses r columns at random from the longest (in the 2-norm) columns of **X**. The idea behind this is that these columns might be more likely to be closer to the centroids.

In Chapter 5, we will present a comparison between the initialization methods using simulated data.

4.4 Computational Complexity

In this section, we present the computational complexity of the developed algorithms. The worst case computational complexity of the two-stage algorithm is $O(MNP)$ per iteration, where M is the number of spectral bands, N is number of pixels, and P is the number of endmembers, respectively.

The penalty approach has the same structure as the general Positive Matrix Factorization described in [59]. The worst case computational complexity of the penalty algorithm is $O((M+1)NP)$ per iteration, where M is the number of spectral bands, N is number of pixels, and P is the number of endmembers, respectively.

4.5 Stopping Criteria

It is common in iterative algorithms to use a stopping criteria to terminate your algorithm. Setting a maximum number of iterations is well known stopping criteria. It is also important to monitor a metric in each iteration for stopping purpose if the solution has progressed far enough to be satisfactory. This process makes it possible to stop before reaching the maximum iteration. Another metric for stopping criteria is to measure the solution difference in two successive iterations

and stop the procedure if the difference is smaller than some threshold. Let us assume that the solution obtained at iterations k and k+1 are given by $\mathbf{S}^k$ and $\mathbf{S}^{k+1}$, respectively. The stopping metric is given by

$$\begin{aligned} \varepsilon_S &= ||\mathbf{S}^{k+1} - \mathbf{S}^k||_F \\ \varepsilon_A &= ||\mathbf{A}^{k+1} - \mathbf{A}^k||_F \end{aligned} \tag{4.8}$$

According to Schau [48], achieving this stopping criteria does not necessarily mean that we are close to the solution.

$$\epsilon = ||\mathbf{X} - \mathbf{S}^k\mathbf{A}^k||_F \tag{4.9}$$

achieves convergence early but it does not mean that we are approaching the actual solution. For any square matrix $\mathbf{T}$, the following is always true

$$(\mathbf{S}^k\mathbf{T})(\mathbf{T}^{-1}\mathbf{A}^k) = \mathbf{S}^k\mathbf{A}^k$$

Due to this effect, the product $\mathbf{S}^k\mathbf{A}^k$ could be close to original matrix $\mathbf{X}$, but individual parameters $\mathbf{S}^k$ and $\mathbf{A}^k$ could be far from convergence. This presence of this matrix makes it difficult to get a unique solution.

Different stopping metrics have been tried in the literature. A stopping metric used in independent component analysis (ICA) [48], is the smoothness factor of the solution during the iterations. For each column of $\mathbf{S}$ and $\mathbf{A}$, the smoothness is defined factor can be written as

$$\begin{aligned} m_{\mathbf{s}_i} &= \ln(17.09\sigma_{\mathbf{s}_i}) \\ m_{\mathbf{a}_i} &= \ln(17.09\sigma_{\mathbf{a}_i}) \end{aligned} \tag{4.10}$$

where $\sigma_{\mathbf{s}_i}$ and $\sigma_{\mathbf{a}_i}$ are the standard deviation of the elements of the i-th columns of $\mathbf{S}$ and $\mathbf{A}$, $\mathbf{s}_i$ and $\mathbf{a}_i$, at each iteration, respectively. A typical plot of the solution

smoothness factor for a given column is shown in Figure 4–2. If the solution smoothness factor of each column $\mathbf{s}_i$ and $\mathbf{a}_i$ is constant, then convergence is achieved. The red line shown in Figure 4–2 indicates the iteration number after which the solution smoothness factor reaches a plateau.

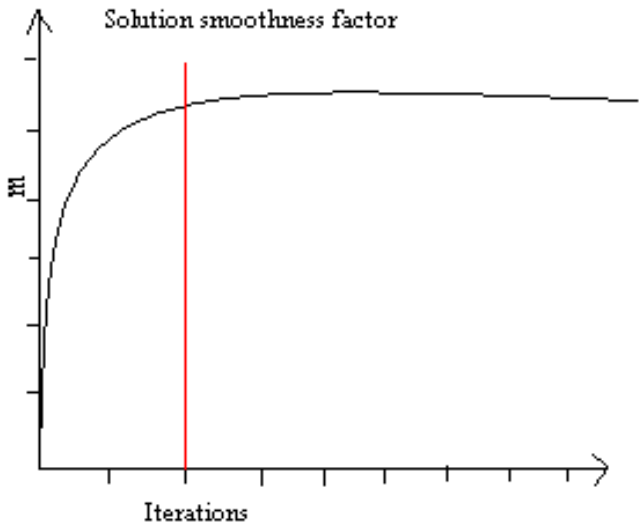


Figure 4–2: Solution smoothness factor typical plot

4.6 Convergence Analysis

Multiplicative updating rules are steepest descent approaches. At each iteration k, the algorithm generates the point $(\hat{S}^k, \hat{A}^k)$ which is calculated based on the point preceding it. The method is descent because as each point is generated by the algorithm, the corresponding value of the objective function decreases in value. We say that an iterative algorithms is *globally convergent* if for any arbitrary starting point the algorithm is guaranteed to generate a sequence of points converging to a point that satisfies the *first order necessary condition for a minimizer*. When the algorithm is not globally convergent, it may still generate a sequence that converges to local point [38]. Our problem is a compound problem of two parameters S and A. Also our problem is not convex in S or A but not in both S and A. This means that no global minima is not guaranteed, however, local minima could be reached.

A related issue of interest pertaining to a given locally or globally convergent algorithm is the *rate of convergence*; that is how fast the algorithm converges to a solution point.

Theorem 3. *Rate of Convergence. Let* $\mathbf{x}^{(k)}$ *be a sequence that convergence to* $\mathbf{x}^*$. *If*

$$\lim_{k\to\infty} \frac{||\mathbf{x}^{(k+1)} - \mathbf{x}^*||}{||\mathbf{x}^{(k)} - \mathbf{x}^*||^{\rho}} = C \tag{4.11}$$

then the order of convergence is at least ρ .

Proof: See [38].

There are three types of convergence.

- Linear: if $\rho = 1$ and $0 < C < 1$.
- Quadratic: if $\rho = 2$.
- Superlinear: $\rho = 1$ and $C = 0$.

Finding the rate of convergence for our problem is not a trivial task. Rate of convergence of unconstrained two-stage Gauss-Seidel algorithm was found in [70] to be linear. If the convergence point is an interior point of the feasible region the unconstrained results apply for rate of convergence as described in [70]. Since the unconstrained Gauss-Seidel convergence rate is linear, then we can conclude that the convergence rate of our constrained Gauss-seidel algorithm is also linear.

4.7 Analysis of Estimation Quality

For algorithms performance evaluation, we need to evaluate the quality of estimated endmembers and the model fitness. In simulated data, we have access to the

endmembers that generate the image data. So we can directly assess the quality of the estimates.

4.7.1 Endmember Estimation Quality

Two metrics are used to measure how close the estimated endmember signature $\hat{\mathbf{s}}_i$ and the true signature $\mathbf{s}_i$. The Euclidean distance and the angular distance.

The Euclidean distance between the estimated and actual endmember signatures, $d(s_i, \hat{s}_i)$ is given by

$$d(\mathbf{s}_i, \hat{\mathbf{s}}_i) = ||\mathbf{s}_i - \hat{\mathbf{s}}_i|| = \sqrt{\sum_j (s_{ji} - \hat{s}_{ji})^2} \tag{4.12}$$

We suggest to use the relative error instead of using the absolute error measured by the Euclidean distance. The relative error between the estimated endmember signature $\hat{\mathbf{s}}_i$ and the true signature $\mathbf{s}_i$ is given by

$$error(\mathbf{s}_i, \hat{\mathbf{s}}_i) = \frac{||\hat{\mathbf{s}}_i - \mathbf{s}_i||}{||\mathbf{s}_i||} \tag{4.13}$$

The angle between the estimated endmember signature $\hat{\mathbf{s}}_i$ and the true signature $\mathbf{s}_i$ is given by

$$\theta(\mathbf{s}_i, \hat{\mathbf{s}}_i) = \cos^{-1}\left(\frac{\hat{\mathbf{s}}_\mathbf{i}^\mathbf{T}\mathbf{s}_\mathbf{i}}{||\hat{\mathbf{s}}_\mathbf{i}||.||\mathbf{s}_\mathbf{i}||}\right) \tag{4.14}$$

This is proper as a measure of spectral differences in hyperspectral image analysis in [1, 64]. These two metrics could be used in the case of simulated data or real data

with ground truth. In Chapter 5, we will use these metrics to evaluate the quality of endmember estimation using simulated data.

4.7.2 Model Fitness

When evaluating the goodness of a model, the model should fit well the training data used to derive its parameters. Here, we look at two important metrics used for the measure of fit: $\mathbf{R}^2$ value, and root-mean-square of the residuals.

4.7.2.1 $\mathbf{R}^2$ Value

Let $\hat{\mathbf{x}}_j = \hat{\mathbf{S}}\hat{\mathbf{a}}_j$ be the estimated spectral signature of the j-pixel, and let $\mathbf{x}_j$ be the measured signature for the j-th pixel. Define the residual vector of the j-pixel to be

$$\mathbf{r}_j = \mathbf{x}_j - \hat{\mathbf{x}}_j$$

In [61], the $\mathbf{R}^2$ value for our linear model is given by

$$R_j^2 = 1 - \frac{\sum_{i=1}^{M}(r_{ij})^2}{\sum_{i=1}^{M} x_{ij}^2} \tag{4.15}$$

The $\mathbf{R}^2$ is the square of the sample correlation between $\mathbf{x}_j$ and $\hat{\mathbf{x}}_j$ [61]. It gives information about the fitness of the estimated model to the data points. The closer $\mathbf{R}^2$ value to unity, indicates good fitness of the model to the measured data.

4.7.2.2 Root-Mean-Square (RMS) Residuals

The RMS residual is a convenient measure of the model fit. The RMS residuals can be displayed as images. High residual values call attention to pixels that are not modeled well using the given set of endmembers. The RMS residual is given by

$$RMS_j = \sqrt{\frac{1}{M}\sum_{i=1}^{M} r_{ij}^2} \tag{4.16}$$

We will use (4.15) and (4.16) to evaluate the quality of the obtained results.

4.8 Summary

This Chapter presents some approaches to compute the constrained PMF. Since the developed algorithms are iterative, we describe initialization schemes.

The stopping criteria of the algorithms is presented as well as the solution smoothing factor as a stopping criteria. We present some metrics for estimation quality. We focus on the quality of estimated endmembers and the model fitness. For the estimated endmembers quality, we propose the percentage error and the spectral angle. We also present metrics to measure the fitness of the estimated endmembers to the measured data. We use $\mathbf{R}^2$ value, and root-mean-square residual.

CHAPTER 5

Experimental Results using Simulated Data

5.1 Introduction

This chapter presents an evaluation of the unsupervised algorithms using simulated data. We performed different experiments using simulated data to evaluate initialization criteria, model fitness, and algorithms performance. We study the sensitivity of the algorithms to these parameters.

5.2 Simulated Data Generation

In order to test the effectiveness of the developed algorithms to solve the unmixing problem, we used two different synthetic data sets. The first data set consists of 1000 pixels generated using the linear mixing model as in (2.1), with five known endmembers selected from the ground truth of the AVIRIS Cuprite image. This is an image taken over the Cuprite mining district, 2 km north of Nevada. The spectral response of the endmembers were used to generate the data is shown in Figure 5–1. The data was generated with different signal-to-noise ratio to fit different of experiments we conducted.

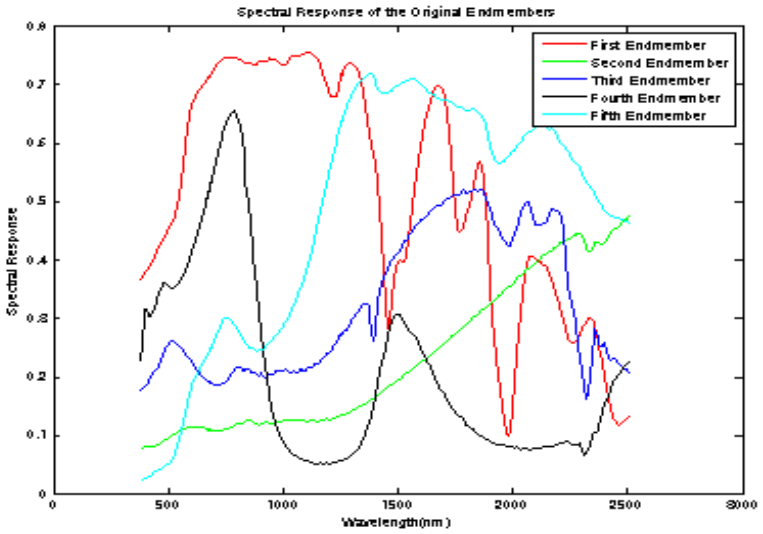


Figure 5–1: Spectral response of the original endmembers

An other simulated data set was generated in the form of a cube with endmembers are located in its corners. The endmembers are used to mix along the diagonals of the cube with different weights based on its position with respect to the cube corners. The spectral signatures of the cube endmembers are shown in Figure 5–2. The corresponding abundances maps are shown in Figure 5–3

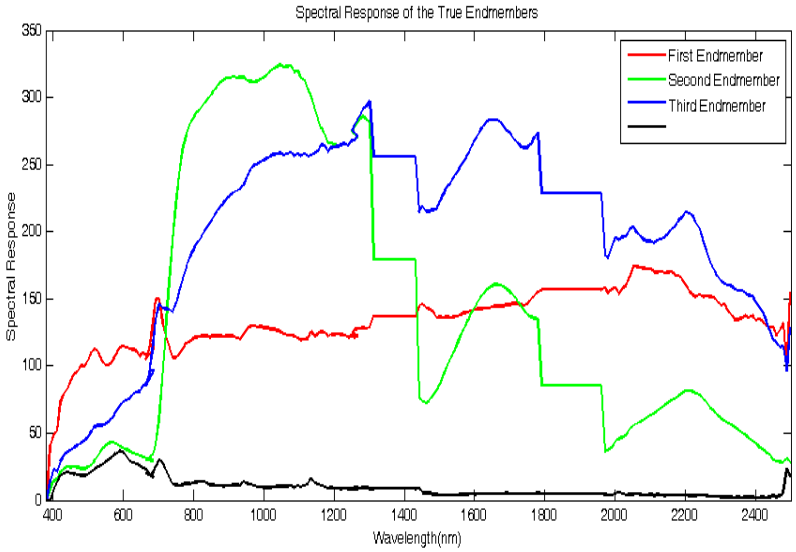
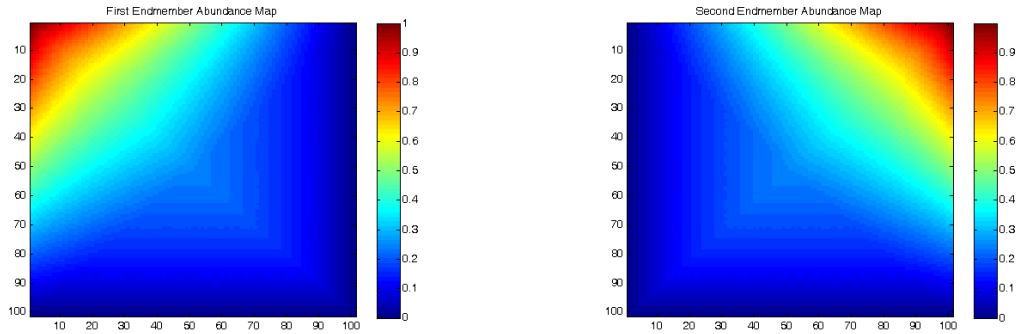
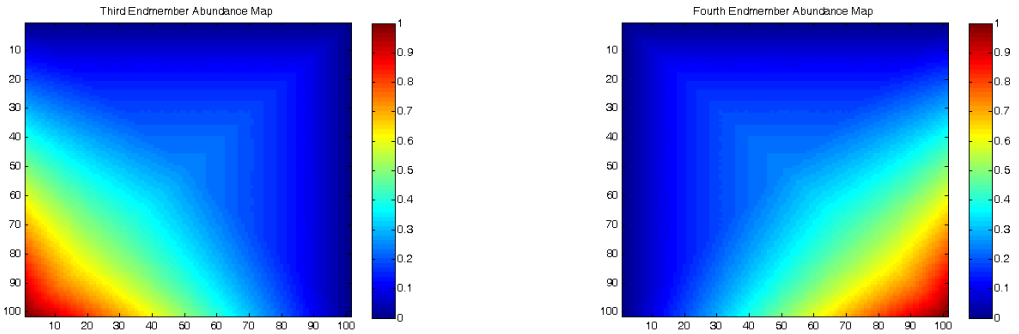


Figure 5–2: Spectral response of cube original endmembers



(a) (b)



(c) (d)

Figure 5–3: Cube true abundance maps

Since the true endmembers are known in both simulated data sets, then, these data sets are useful to evaluate the developed algorithms. In this Chapter, we will use the simulated data set to evaluate the performance of the developed algorithms

5.3 Estimating Number of Ednemembers

In this section, we estimate the number of endmembers using the linear dimensionality of hyperspectral data. We used the linear approaches for estimating the hyperspectral image dimension. The estimated dimension using linear and nonlinear approaches of the simulated data is shown in Table 5–1. It can be seen that the SVD approach with percentage variance equal to 99% gives the correct number of endmembers. The rest of approaches estimated the same data dimension.

Table 5–1: Simulated data dimension estimate

Approach	SVD	KG	CSD	Fukunaga
Dimension	5	4	4	4

5.4 Effect of Initialization Schemes

To evaluate the effect of initialization schemes on the results, we conducted an experiment using Cuprite simulated data. Simulated pixels were used to estimate the endmembers and abundances using the Gauss-Seidel algorithm with different initialization methods described in Chapter 4. Four initialization approaches, random initialization, random Acol initialization, random C initialization, and singular value decomposition subset selection (SVDSS) approach were used. We used the angle, and the percentage difference of Euclidean distance between the estimated and the true endmembers to evaluate the resulting endmember estimates. Table

5–2 shows the angle between the estimated and the true endmembers for four initialization procedures for the two-stage algorithm. Table 5–3 shows the percentage difference of Euclidean distance between the estimated and true endmembers. It can be seen from Table 5–2 and Table 5–3 that for the estimated first endmember, shown in Figure 5–4, the Random Acol initialization criteria performed the best while the random initialization has the worst performance. Figure 5–5 together with Table 5–2 and Table 5–3 show that the SVDSS initialization criteria gives the best overall performance. Figures 5–6, to 5–8 show the spectral response of the third, fourth, and fifth endmembers respectively. Table 5–2 and Table 5–3 show that the SVDSS initialization criteria performed better than the other initialization criteria. To minimize randomness effects on the results, we repeated the experiment four times; we use the average of the four trials to compare between the initialization procedures. Table 5–4 shows the initial and the final values of the objective function for each initialization method; it also shows the execution time of the algorithm for initialization method.

Table 5–2: Angle between the estimated and the true endmembers (degrees)

	Random	Random Acol	Random C	SVDSS
$\theta(s_1, \hat{s}_1)$	14.27	1.24	2.60	1.43
$\theta(s_2, \hat{s}_2)$	14.12	3.14	4.38	2.28
$\theta(s_3, \hat{s}_3)$	12.95	9.35	8.20	6.65
$\theta(s_4, \hat{s}_4)$	8.31	6.38	5.89	6.41
$\theta(s_5, \hat{s}_5)$	25.24	6.47	5.63	0.12

Table 5–3: Percentage error of the estimated endmembers (%)

	Random	Random Acol	Random C	SVDSS
$(s_1, \hat{s}_1)$	3.85	0.45	0.81	0.71
$(s_2, \hat{s}_2)$	6.64	1.42	2.01	1.30
$(s_3, \hat{s}_3)$	6.72	7.88	7.81	2.66
$(s_4, \hat{s}_4)$	3.59	2.96	2.55	2.80
$(s_5, \hat{s}_5)$	6.66	1.61	1.61	0.03

Table 5–4: Initial and final values of objective function and execution time

	Random	Random Acol	Random C	SVDSS
E_0	532.734	186.935	360.814	6.205
E_f	0.733	0.0464	0.0689	0.0084
Time(sec)	576.75	608.22	607.68	612.81

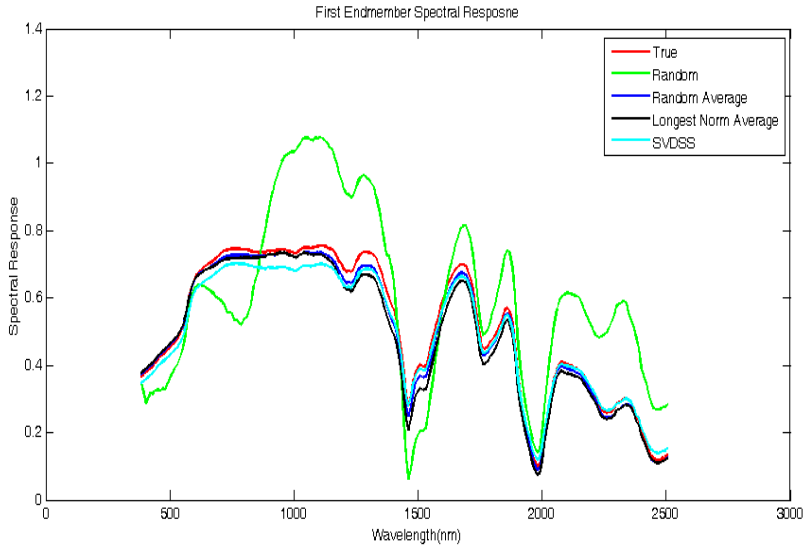


Figure 5–4: True and estimated spectral response of the first endmember

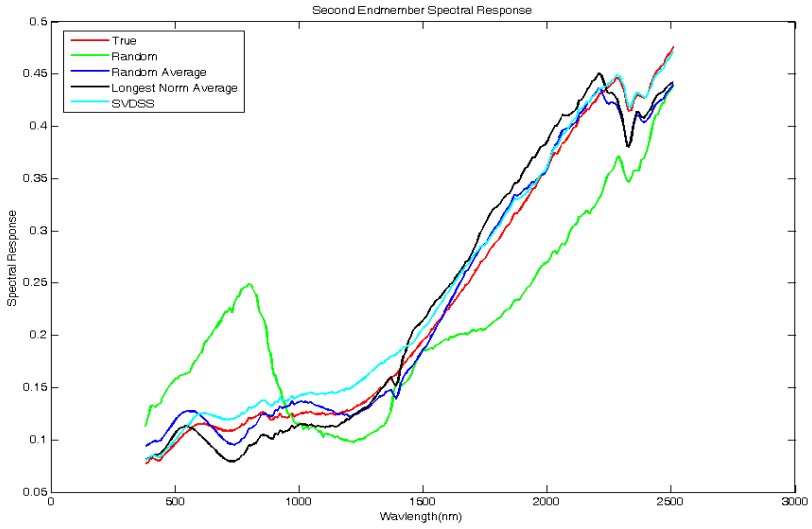


Figure 5–5: True and estimated spectral response of the second endmember

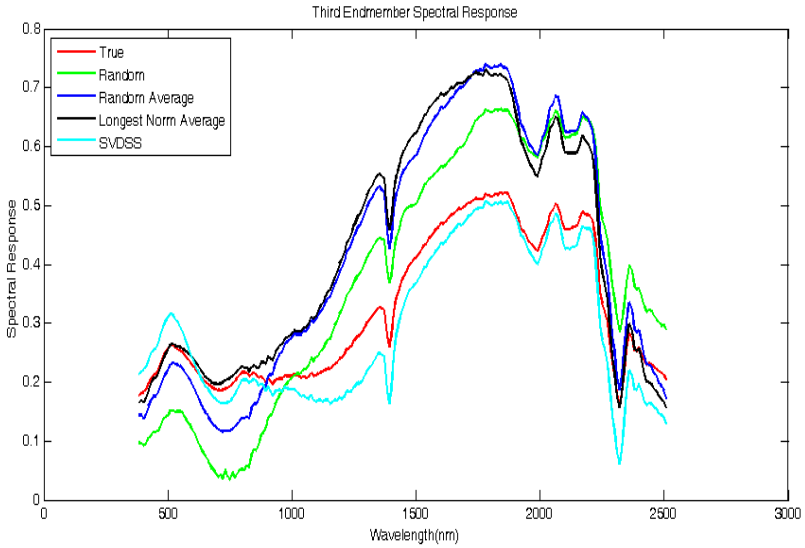


Figure 5–6: True and estimated spectral response of the third endmember

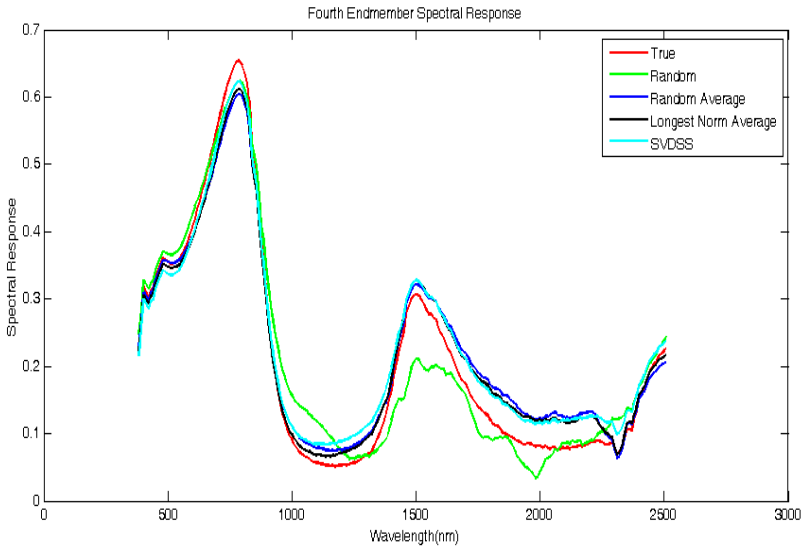


Figure 5–7: True and estimated spectral response of the fourth endmember

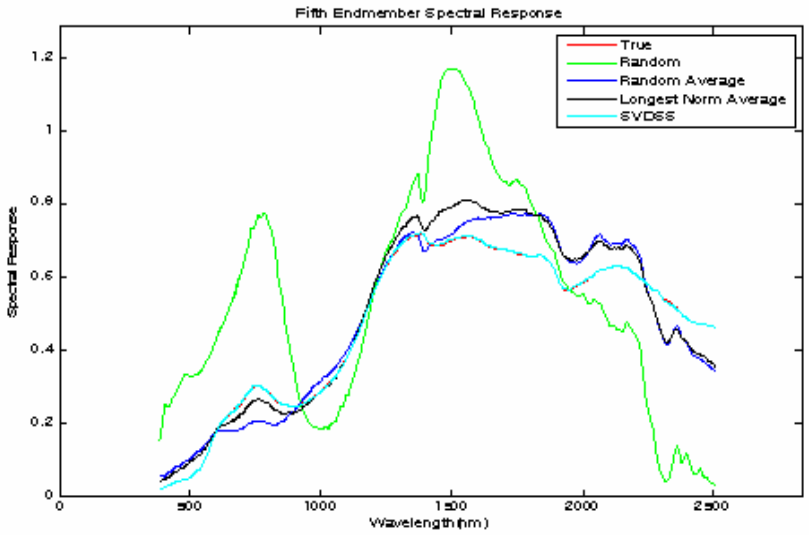


Figure 5–8: True and estimated spectral response of the fifth endmember

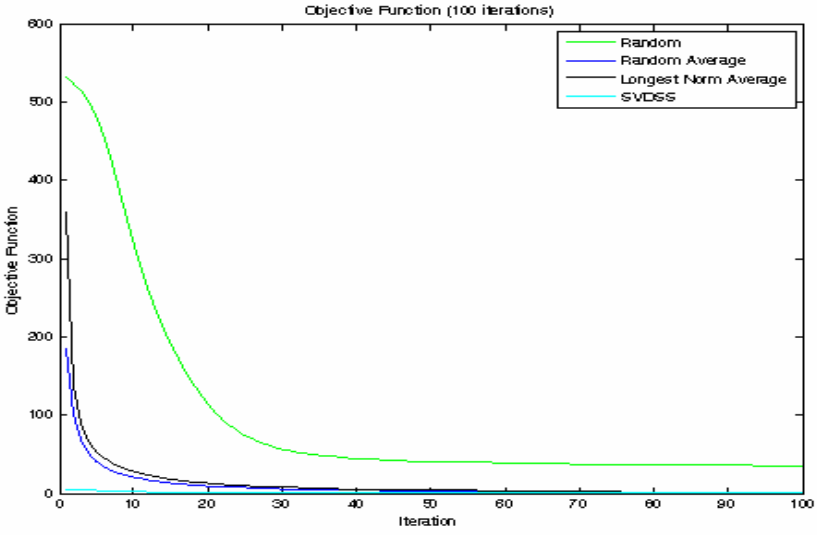


Figure 5–9: Objective function

The value of the objective function at each iteration is shown in Figure 5–9 for the four initialization methods. It can be seen that for SVDSS initialization method both the initial and the final values of the objective function are much smaller than for the other initialization methods. If a threshold value of the objective function is used as a stopping criteria, then this results in faster stopping. No useful information can be taken from the execution time since we the maximum number of iteration as a stopping criteria.

5.5 Evaluation of the Effect of Number of Spectral Bands

We studied the effect of the number of spectral bands on the performance of the algorithms. For this purpose, we used the same simulated data set used for initialization criteria evaluation. In this experiment, we used SVDSS -band selection [43] to select different bands subsets. The two-stage approach was used to study the effect of number of bands on the algorithm performance.

Table 5–5: RMS residuals result as a function of number of spectral bands (Two-stage)

Num. spectral bands	Min rms value($\times 10^{-6}$)	Max rms value	Average rms($\times 10^{-5}$)
10	1.08	0.0042	4.194
50	1.58	0.0045	6.092
100	1.40	0.005	6.49
150	1.26	0.0049	5.96
200	1.20	0.0049	5.47
224	21.2	0.0234	95.6

Table 5–6: RMS residuals result as a function of number of spectral bands (Penalty)

Num. spectral bands	Min rms value($\times 10^{-4}$)	Max rms value	Average rms($\times 10^{-3}$)
10	4.44	0.045	6.0
50	3.37	0.0387	5.4
100	3.55	0.0388	5.3
150	4.02	0.0382	5.4
200	4.88	0.0377	5.4
224	22	0.1337	30.7

Tables 5–5, and 5–6 show the minimum, maximum, and average values of the rms residuals for different number of spectral bands obtained using the SVDSS approach for the two-stage approach, and the penalty approach, respectively. It can be seen from Tables 5–5 and 5–6, that when the number of bands is equal to 224, the rms residual value are the worst. In the 10 spectral bands case, the rms residuals are the best.

Figure 5–10 shows the plot of the rms residuals average as a function of the number of spectral bands for the two-stage and penalty approaches. It can be shown from Figure 5–10 that between 50 bands and 200 bands the variation in the rms residuals average is so small for both approaches. On the other hand, the variation in rms residuals average between 200 bands and 224 bands is noticeable, also it is true for both approaches. Figure 5–10 shows that the rms average values are smaller for the two-stage approach than that for penalty approach.

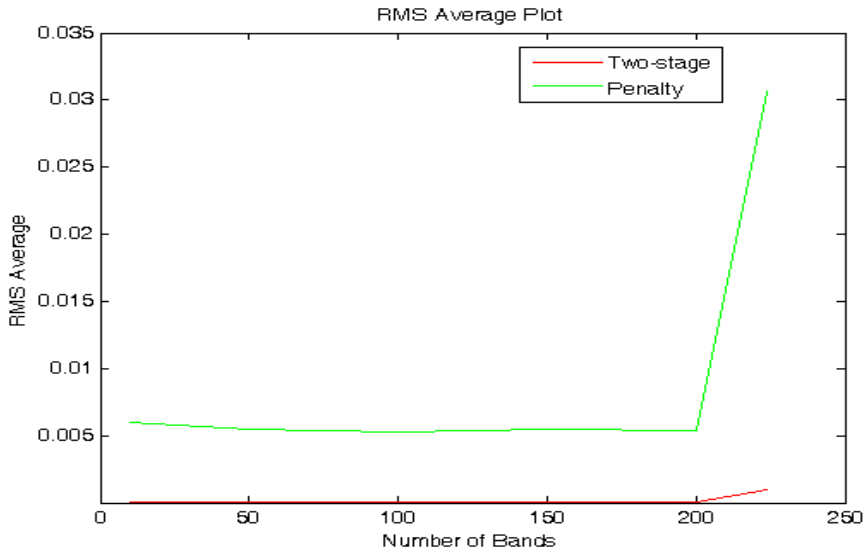


Figure 5–10: Plot of RMS residuals average plot as a function of number of spectral bands

5.6 Evaluation of the Effect of Number of Endmembers

We conducted an experiment using the cube simulated data and various number of endmembers to study the effect of the number of endmembers on the results. The cube data has four endmembers located on its four corners. Figure 5–2 shows the spectral signature of the true four endmembers of the cube data, and Figure 5–3 shows their corresponding abundance maps.

Table 5–7 shows the $\mathbf{R}^2$ average, RMS residual average, and their standard deviation for three, four, five, six, and eight endmembers experiments using the cube simulated data. The $\mathbf{R}^2$ value average, and the RMS residual average shown in Table 5–7, indicate that four endmembers is the best number which is expected since the structure of the cube data is already known. Figure 5–11 shows that when the number of endmembers is greater than four, the mean squared error does not change. This is also an indication that four endmembers is the best choice for number of endmembers for the cube data.

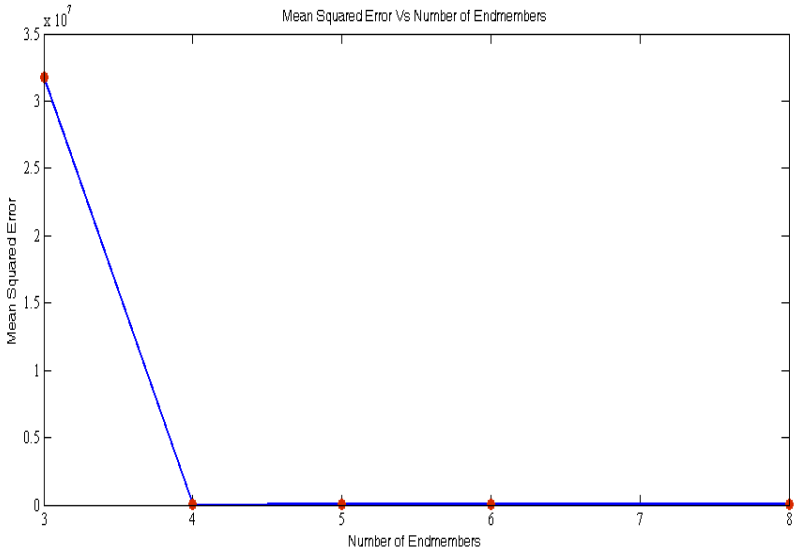


Figure 5–11: Cube mean squared error

To show the effect of the number of endmembers on the results, we chose number of endmembers less than four (three endmembers), four endmembers, and greater than four (eight endmembers). The estimated spectral signatures for the three endmembers, four endmembers, and eight endmembers are shown in Figures 5–12, 5–14, and 5–16, respectively. The corresponding abundance maps are shown in Figures 5–13, 5–15, and 5–17, respectively.

Table 5–7: $\mathbf{R}^2$ and RMS results for different number of endmembers

Num. of Endmembers	$\mathbf{R}^2$ Average	$\mathbf{R}^2$ Stdev	RMS Average	RMS Stdev
3	0.9953	0.0467	3.8118	3.6646
4	1	0	3.1677×10^{-6}	7.8225×10^{-7}
5	1	1.488×10^{-5}	0.0260	0.1516
6	1	4.927×10^{-6}	0.0137	0.0780
8	1	4.368×10^{-6}	0.0098	0.0636

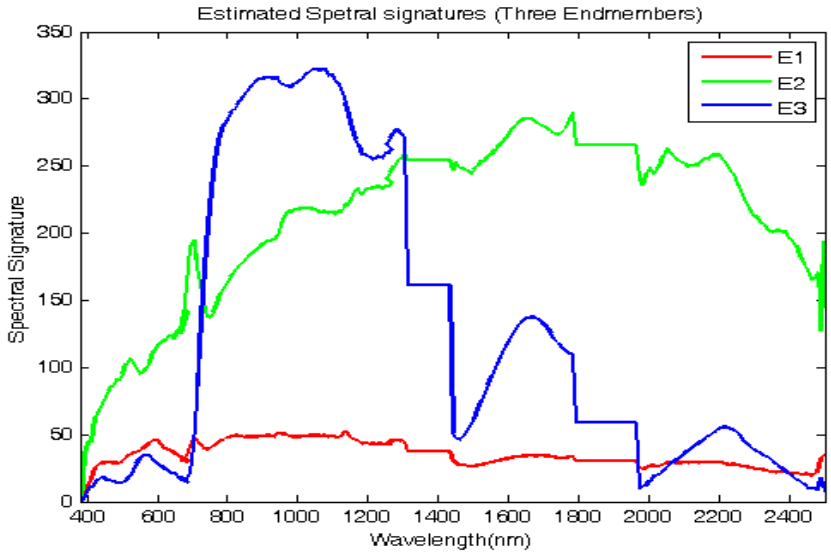
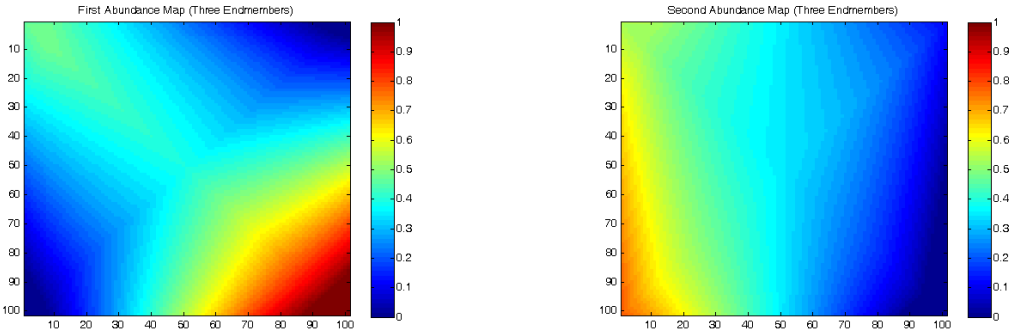
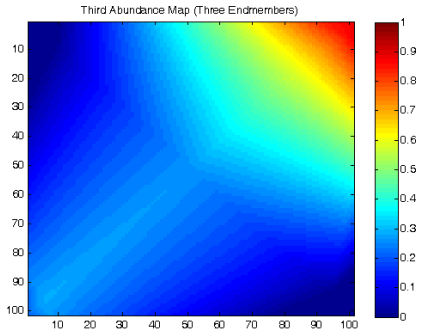


Figure 5–12: Three endmembers estimated Spectral signatures

It can be seen from Figure 5–12, Figure 5–13, and Figure 5–2 that the first estimated endmember (E1) is close to the fourth true endmember and the third estimated endmembers (E3) is close to the second true endmember. The second estimated (E2) is a linear combination of the first and the third true endmembers. Comparing Figure 5–14 and 5–2, we can see that the four estimated endmembers agree to the true endmembers.



(a) (b)



(c)

Figure 5–13: Cube abundance maps for three endmembers

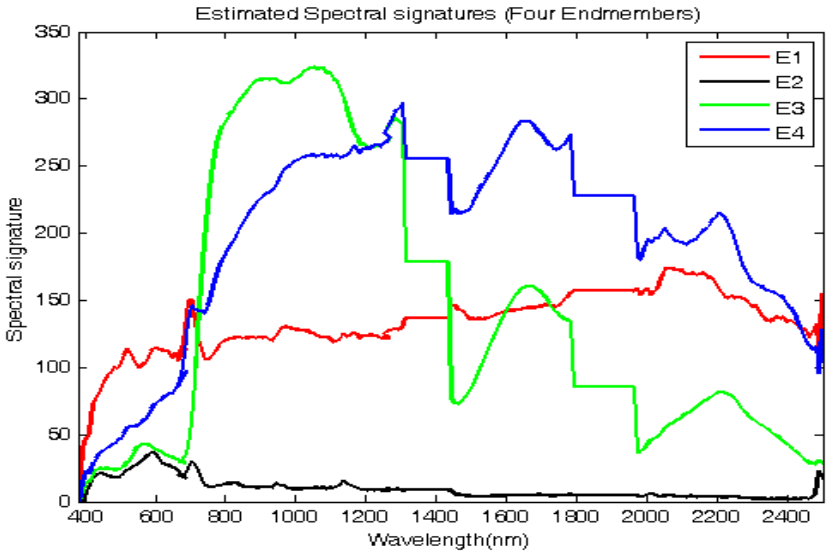


Figure 5–14: Four endmembers estimated Spectral signatures

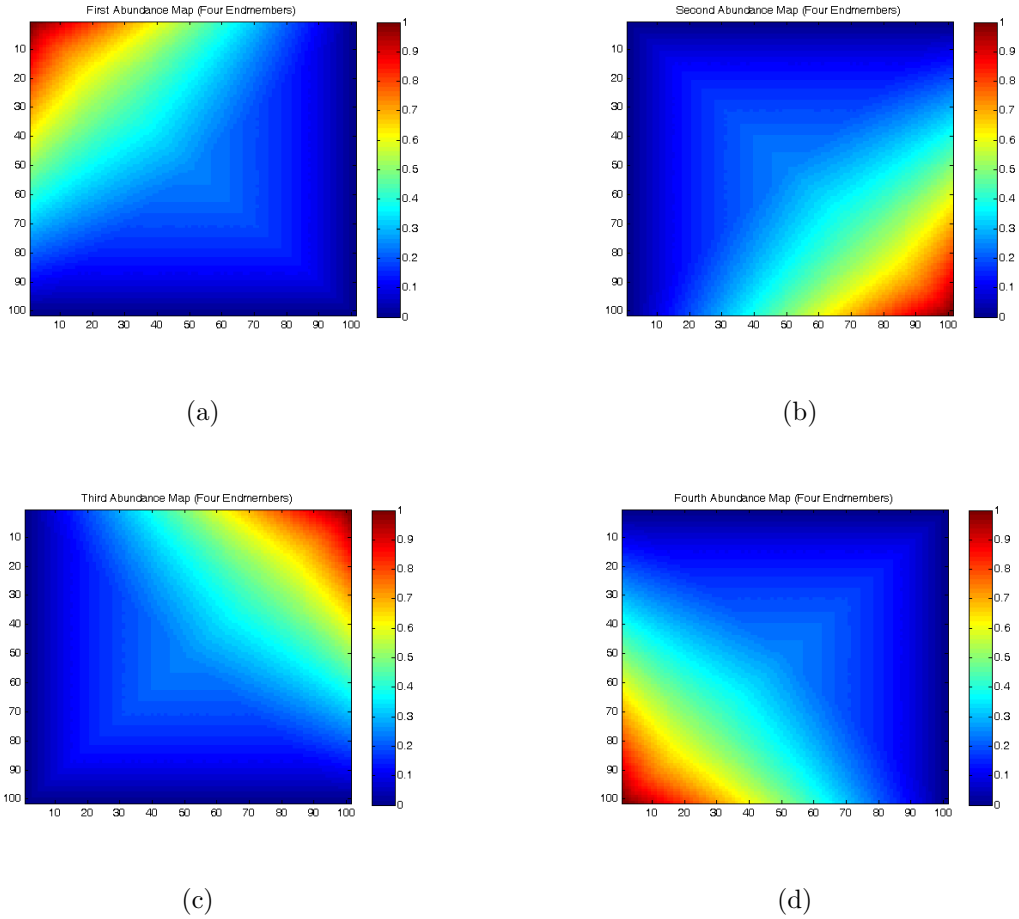


Figure 5–15: Cube abundance maps for four endmembers

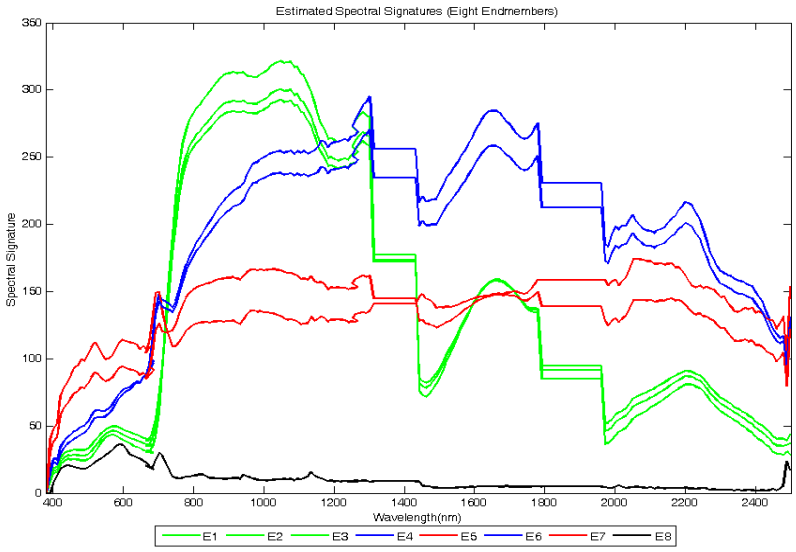
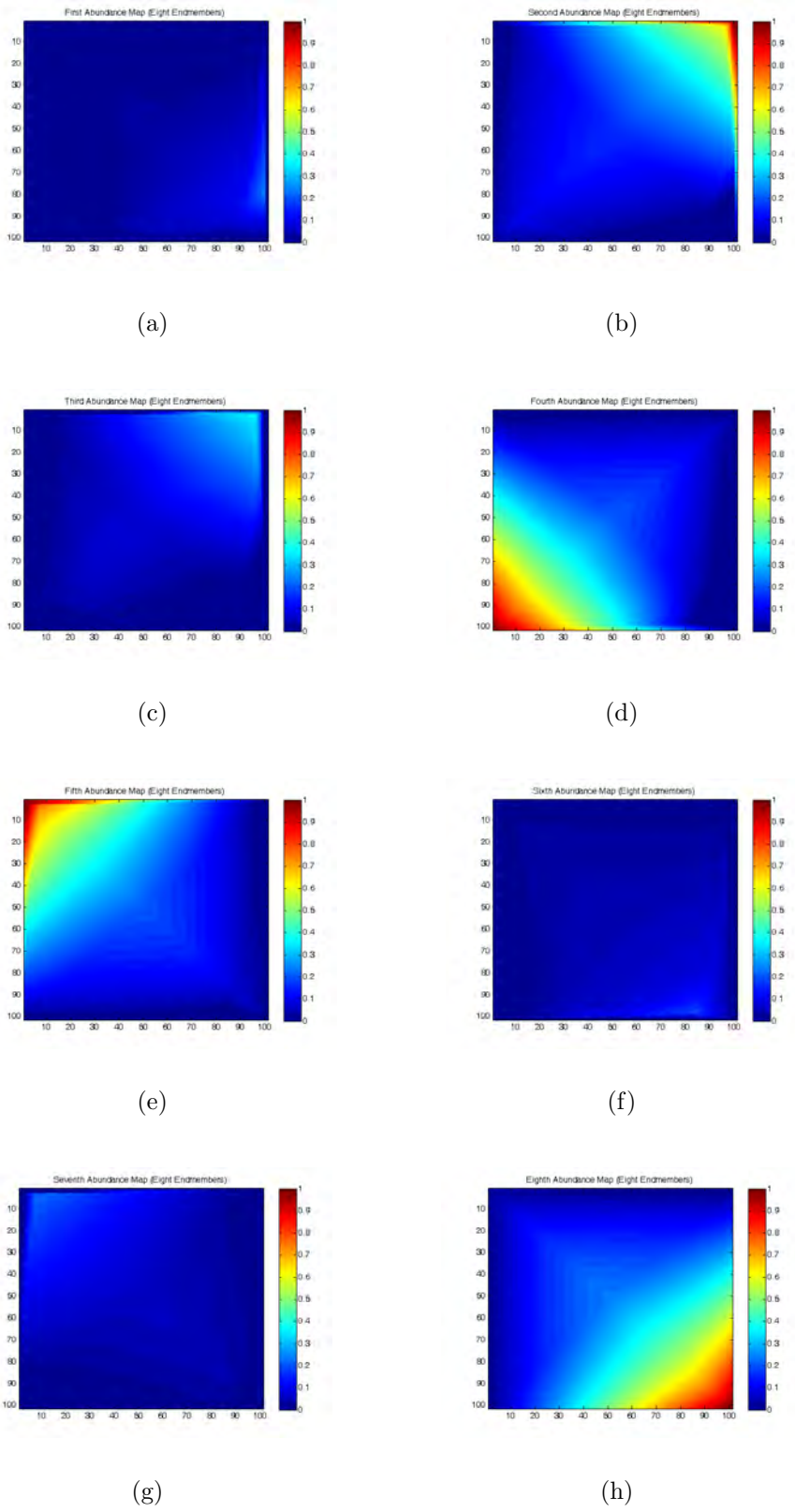


Figure 5–16: Eight endmembers estimated Spectral signatures



(a) (b)

(c) (d)

(e) (f)

(g) (h)

Figure 5–17: Cube abundance maps for eight endmembers

The estimated spectral signature for the eight endmembers and their corresponding abundance maps are shown in Figures 5–16, and 5–17, respectively. We can see from Figure 5–16 that

- The first (E1), the second (E2), and the third (E3) estimated endmembers are similar and close to the second true endmember.
- The fourth (E4) and the sixth (E6) estimated endmembers are similar and close to the third true endmember.
- The fifth (E5), and the seventh (E7) estimated endmembers are similar and close to the first true endmember.
- The eighth (E8) estimated endmember is close to the fourth true endmember.

From the previous observations, we can conclude that for the cube simulated data, when the number of endmembers (eight endmembers) is greater than the true number of endmembers (four endmembers), some of the estimated endmembers are of similar shape. The whole estimated endmembers can be divided into number of sets equal to the number of true endmembers. The estimated endmembers in each set have similar spectral signatures and close to one spectral signature of true endmember. The sum of the abundance fractions corresponding to each estimated endmember in each set, is the same as the abundance fractions of the closest true endmember in the original image.

5.7 Evaluation of the Effect of Noise

In nature, signals are mixed with some noise due to different factors. We studied the effect of noise on the results of the developed algorithms. For this purpose, we used the 1000 pixel generated from the five endmembers taken from the Cuprite image. We added noise to these pixels such that the specified SNR is obtained. The

added noise is a zero-mean white noise. According to [63], the signal-to-noise ratio for the zero-mean white noise is given by

$$SNR = \frac{E[\mathbf{X}^T\mathbf{X}]}{M\sigma^2} \approx \frac{\|\mathbf{X}\|_F^2}{NM\sigma^2} \tag{5.1}$$

where N is the number of pixels in the image, M is the number of spectral bands, and σ^2 is the variance of the white noise.

Table 5–8: $\mathbf{R}^2$ value average and RMS residuals average results for different SNR

SNR	$\mathbf{R}^2$ Average	Stdev	RMS Average	Stdev
5	0.8381	0.0343	0.1557	0.0071
10	0.9083	0.0218	0.1116	0.0053
20	0.9509	0.0126	0.0796	0.0038
50	0.9796	0.0056	0.0504	0.0023
100	0.9897	0.0028	0.0356	0.0017
1000	0.9990	2.8406×10^{-6}	0.0112	5.185×10^{-4}

Table 5–8 shows the $\mathbf{R}^2$ average value and the RMS residual average and the corresponding standard deviations for different SNR values.

Figures 5–19, 5–20, 5–21, and 5–22, show the estimated spectral signature of the second, third, fourth, and fifth endmembers, respectively, for different SNR values. The estimated spectral signature of the second endmember is noisy for SNR=5, and SNR=10. In the third endmembers case, we notice that for SNR less than 100, the estimated spectral signatures were noisy and do not in amplitude to the true spectral signature.

Figure 5–18 shows the estimated spectral signature of the first endmember for different SNR values. It can be seen that for SNR=5, and SNR=10 the estimated spectral signature is very noisy and not close to the true spectral signature. For SNR equal or greater than 20, the estimated spectral signatures were clean and agree to

the true spectral signature. For the four endmembers and five endmembers cases, it can be seen that for SNR greater than 20, the estimated spectral signatures agree to the true spectral signature.

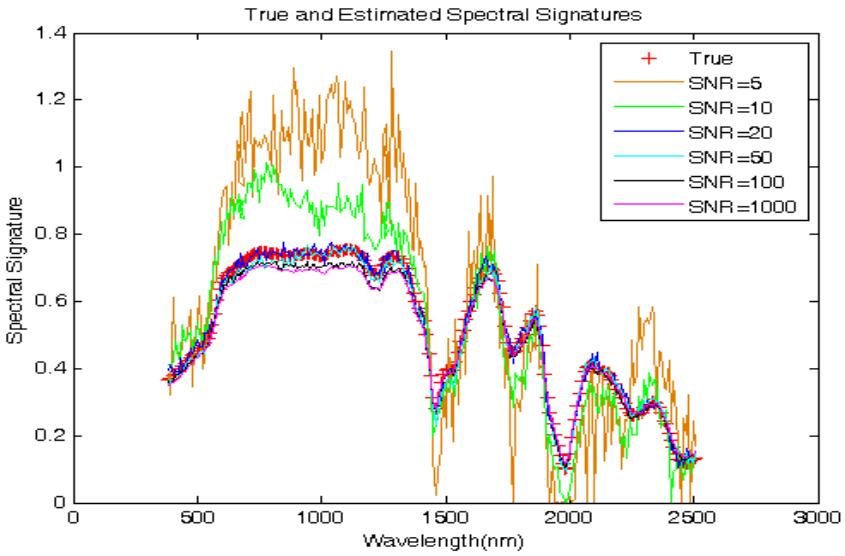


Figure 5–18: First endmember estimated Spectral signatures

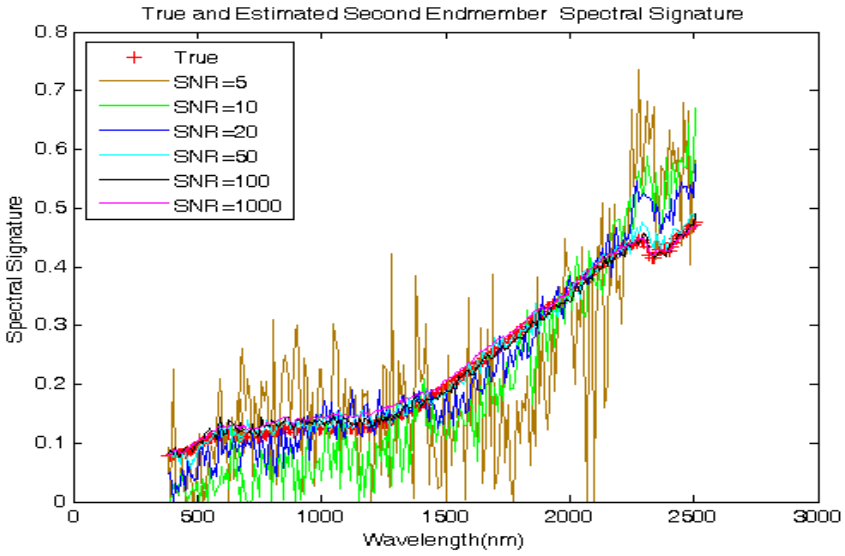


Figure 5–19: Second endmember estimated Spectral signatures

Previous results show that the two-stage approach deals with data where signal-to-noise ratio is low or high. In the case of low SNR, the estimated spectral signatures

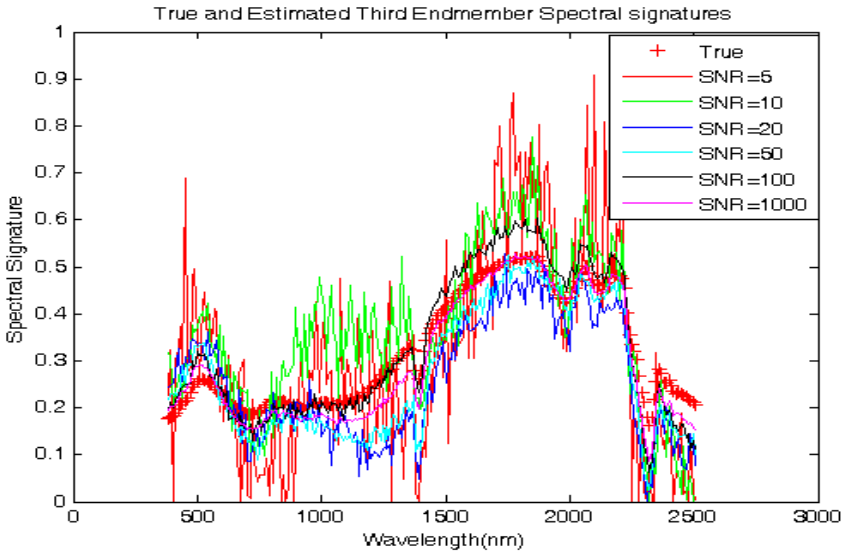


Figure 5–20: Third endmember estimated Spectral signatures

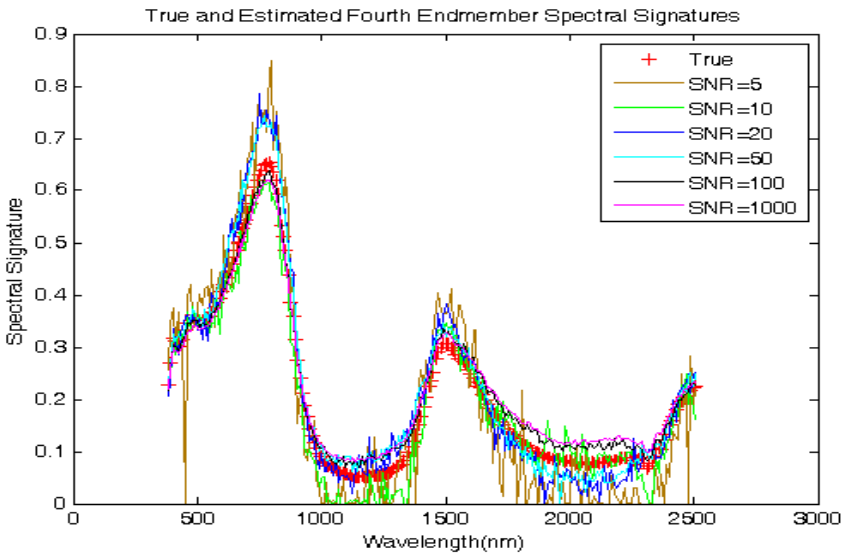


Figure 5–21: Fourth endmember estimated Spectral signatures

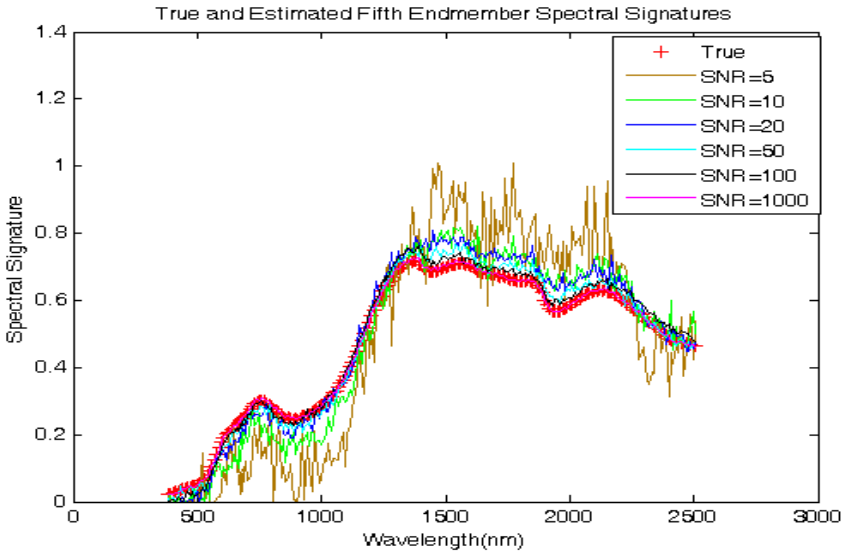


Figure 5–22: Fifth endmember estimated Spectral signatures

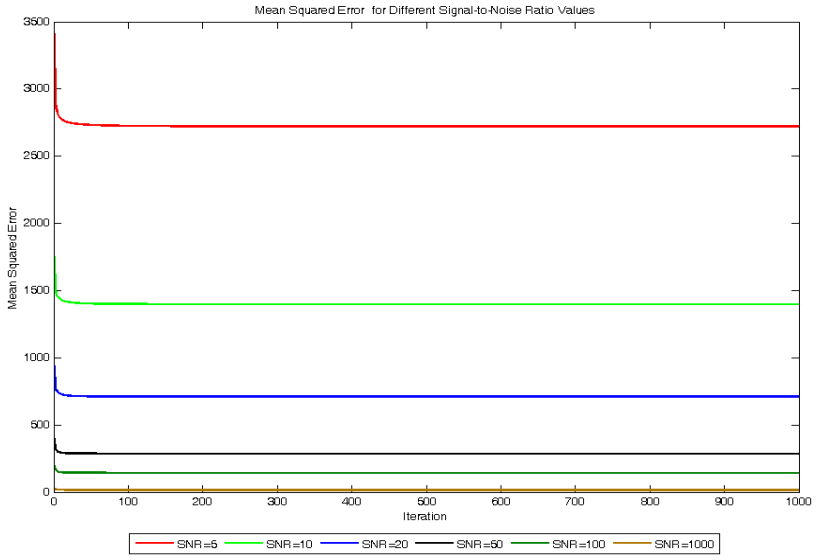


Figure 5–23: Mean squared error for different SNR values

were close in shape with some noise. When the SNR is high, in our case for SNR equal or greater than 20, the algorithm estimates the spectral signatures correctly.

5.8 Algorithm Evaluation

We evaluated the developed algorithms by conducting two experiments. The purpose of these experiment was to measure the capability of the developed algorithms for estimating the endmembers spectra using simulated data. Another purpose was to compare between the developed algorithms based on ability to estimate the endmembers from simulated data.

5.8.1 Experiment I: Cube simulated Data

To evaluate the performance of the algorithms, we conducted an experiment using simulated data arranged in a cube of 101 rows, 101 columns, and 223 spectral bands. This cube is constructed with the four endmembers placed at the corners of the cube. The spectral responses of the endmembers used in mixing the cube pixels and their corresponding abundance maps are shown in Figure 5–2, and 5–3.

The estimated cube data dimension is shown in Table 5–9. It can be seen from Table 5–9 that the SVD approach estimates the correct number of endmembers. In this experiment, we will use the estimated data dimension obtained using the SVD approach as our number of endmembers. Hence, the number of endmembers is four.

Table 5–9: Cube data dimension estimate

Approach	SVD	KG	CSD	Fukunaga
Dimension	4	3	3	3

Figure 5–24 shows the spectral responses of true and the estimated first endmember. The estimated spectral response from both the penalty and the two-stage

approaches agree completely with the true spectral response. The abundance map of the first endmember obtained from the two-stage and the penalty approaches are shown in Figure 5–25(a) and Figure 5–25(b), respectively. It is clear from the abundance maps that both approaches performed well in estimating the first endmember spectral response and abundance map.

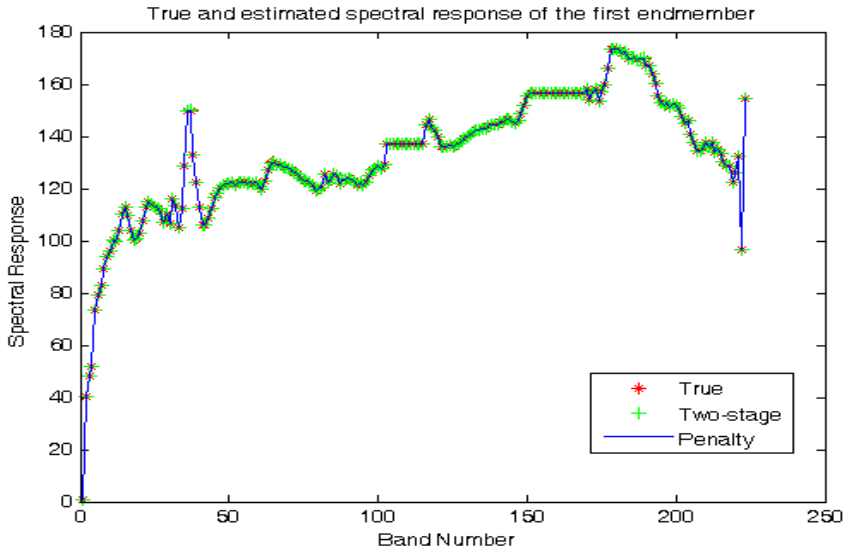
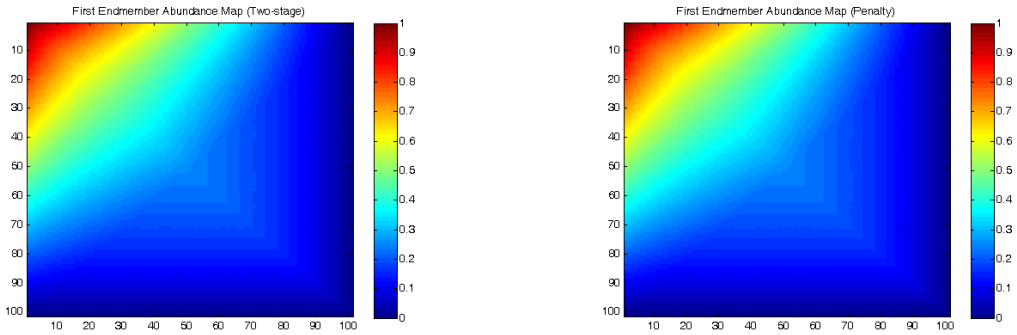


Figure 5–24: True and estimated spectral response of the first endmember



(a) (b)

Figure 5–25: Abundance map of the first endmember using (a)Two-stage approach (b)Penalty approach

Figure 5–26 shows the spectral response of the true and estimated second endmember. A complete agreement between the true and the estimated spectral response can be seen from the figure.

The abundance map of the second endmember obtained from the two-stage and the penalty approaches are shown in Figure 5–27(a) and Figure 5–27(b), respectively. As we know previously from the structure of the simulated data, the high concentration of the second endmember exists in the upper right corner of the cube and starts to reduce along the diagonal; this is clear in the abundance maps obtained from both approaches.

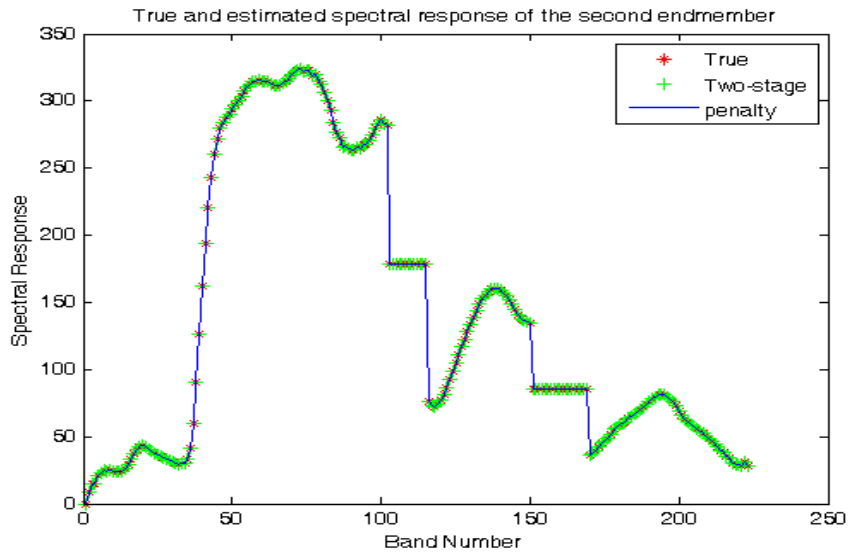
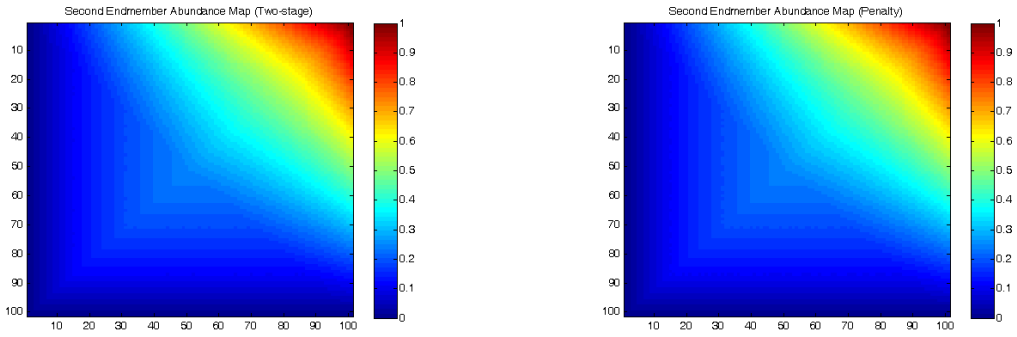


Figure 5–26: True and estimated spectral response of the second endmember

Figure 5–28 shows the spectral response of the true and estimated third endmember. The percentage error, and the spectral angle between the estimated and the true endmembers were found to be 0 which shows, together with Figure 5–28 that both algorithms estimated the endmembers correctly.

The abundance map of the third endmember obtained from the two-stage and the penalty approaches are shown in Figure 5–29(a) and Figure 5–29(b), respectively.



(a) (b)

Figure 5–27: Abundance map of the second endmember using (a)Two-stage approach (b)Penalty approach

Again, the concentration of the third endmember is high in the lower left corner of the cube and start to reduce along the diagonal toward the intersection of the diagonals of the cube face. This also agree with the structure of the simulated data.

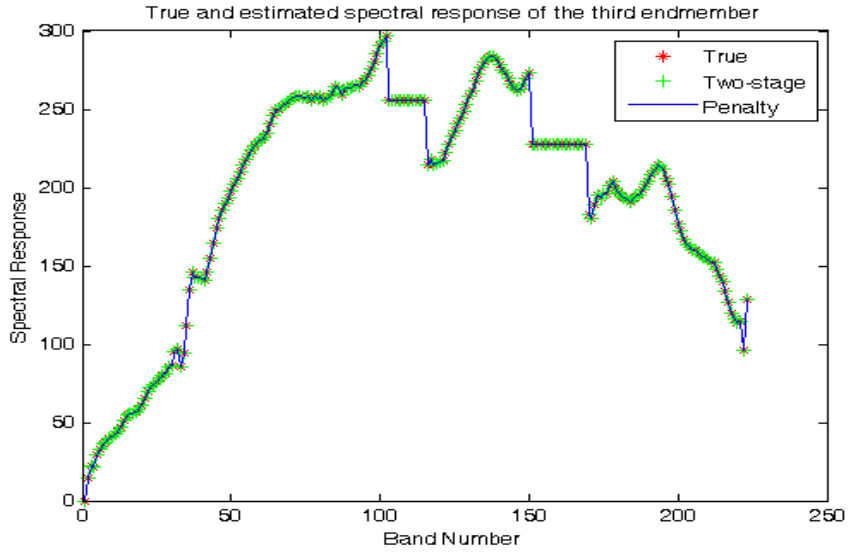
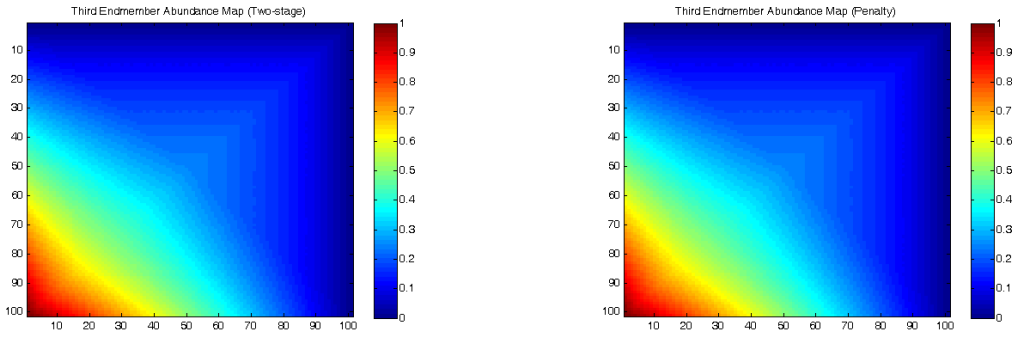


Figure 5–28: True and estimated spectral response of the third endmember

Figure 5–30 shows the spectral response of the true and estimated fourth endmember. The estimated spectral response agrees completely with the true spectral response.



(a) (b)

Figure 5–29: Abundance map of the third endmember using (a)Two-stage approach (b)Penalty approach

The abundance map of the fourth endmember obtained from the two-stage and the penalty approaches are shown in Figure 5–31(a) and Figure 5–31(b), respectively. The concentration of the fourth endmember is high in the lower right corner of the cube and start to reduce along the diagonal toward the intersection of the diagonals of the cube face. This also agree with the structure of the simulated data.

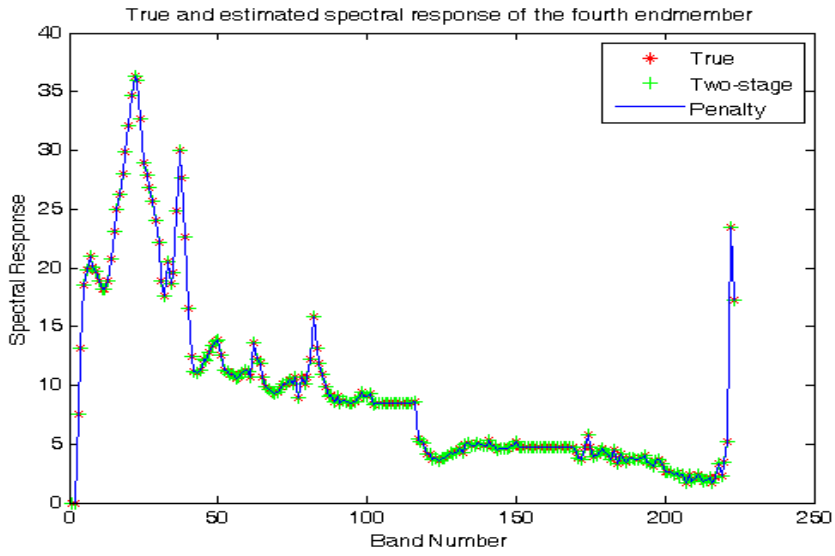
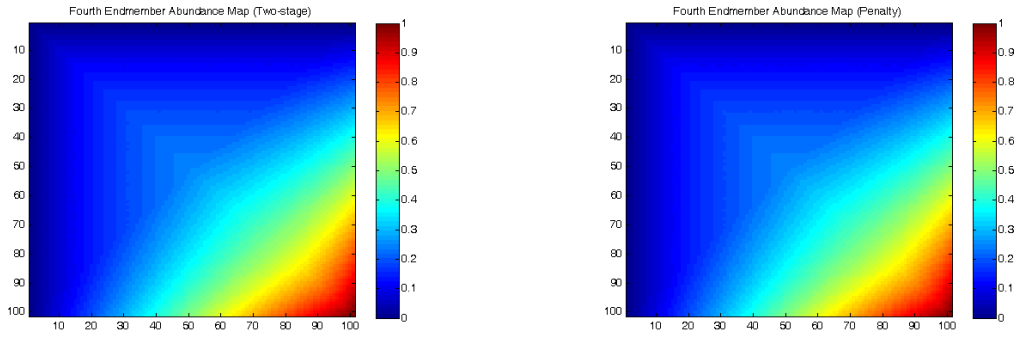


Figure 5–30: True and estimated spectral response of the fourth endmember

Figure 5–32, and Figure 5–33 show the $\mathbf{R}^2$ plot for the two-stage approach, and the penalty approach, respectively. The $\mathbf{R}^2$ plot measures the fitness of the estimated



(a) (b)

Figure 5–31: Abundance map of the fourth endmember using (a)Two-stage approach (b)Penalty approach

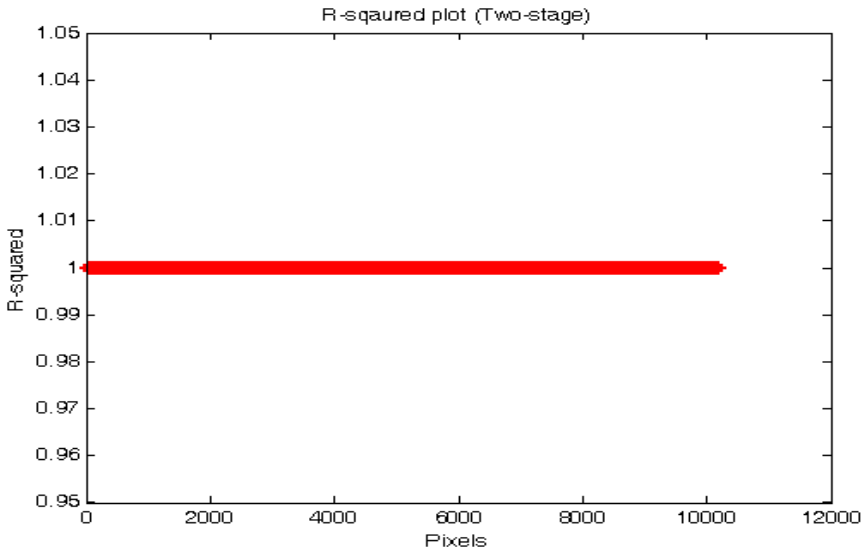


Figure 5–32: R-squared plot (Two-stage)

endmembers to simulated data cube. The obtained $\mathbf{R}^2$ plots from both approaches were close to unity, hence the estimated endmembers fit well the simulated data. This is expected since the spectral response of the four endmembers were correctly estimated.

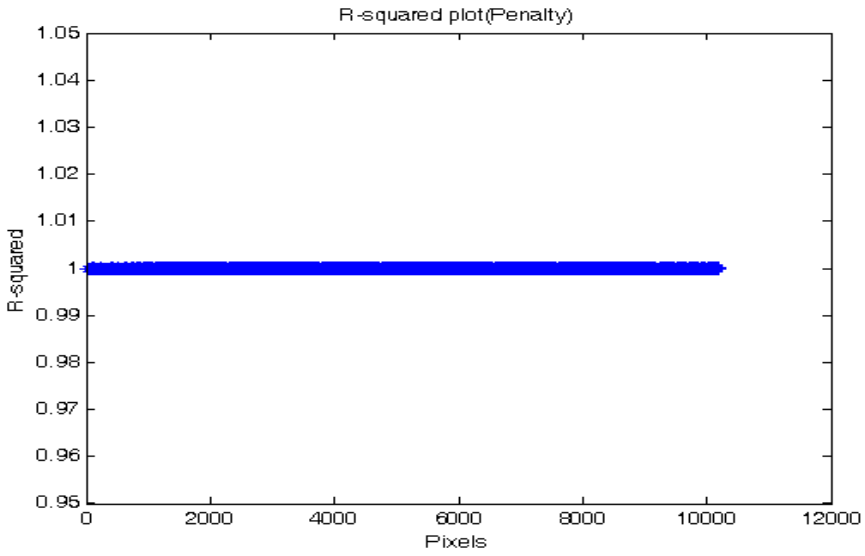


Figure 5–33: R-squared plot (Penalty)

Figure 5–34 shows the mean squared error for the two-stage approach, and the penalty approach. It can be seen from the mean squared error plot that the two-stage approach goes to lower value than the penalty approach.

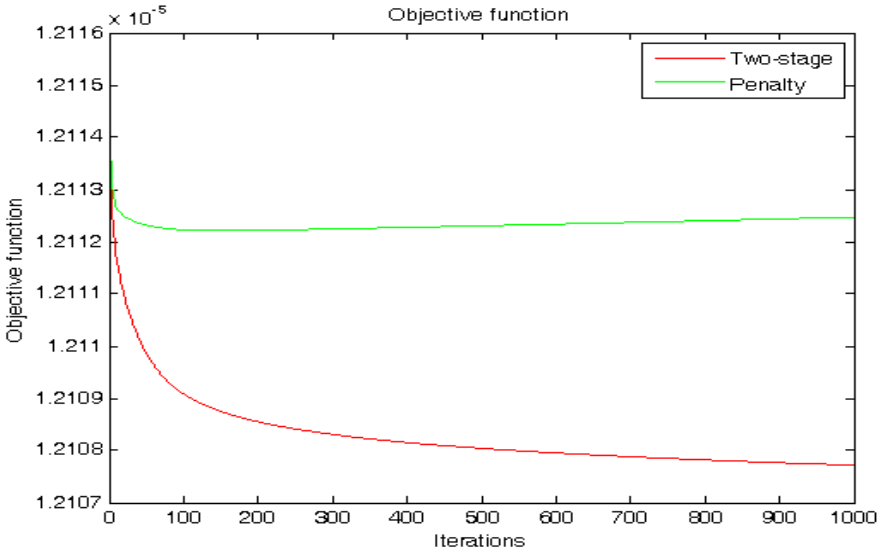


Figure 5–34: Cube data mean squared error curve

In addition to the spectral signatures plot (Figures 5–24, 5–26, 5–28, and 5–30) and their abundance maps (Figures 5–25, 5–27, 5–29, and 5–31) for the estimated

four endmembers of the cube simulated data, the percentage error and the spectral angle between the estimated and the true spectral signatures were found to be 0, hence both algorithms estimated correctly the four endmembers.

5.8.2 Experiment II: Simulated Data using Five Endmembers from Cuprite Image

This experiment was conducted using the same simulated data as in Section 5.4. In addition to evaluating the performance of the developed algorithms, we compare between the Gauss-Seidel approach and the penalty approach discussed in Section 4.2. We compare the algorithms based on how close are the estimated spectral signatures from the true spectral signatures obtained from ground truth of the Cuprite image. The SVDSS initialization scheme will be used in this experiment.

Tables 5–10 and 5–11 show the angle and the percentage error between the true and estimated signatures. We can see from Tables 5–10 and 5–11 that the two-stage approach results in smaller spectral angles and smaller percentage error between the true and estimated signatures than the penalty approach.

Table 5–10: Angle between the estimated and the true endmembers for Two-stage and Penalty approaches(degrees)

	Two-stage Approach	Penalty Approach
$\theta(s_1, \hat{s}_1)$	1.43	3.96
$\theta(s_2, \hat{s}_2)$	2.28	9.26
$\theta(s_3, \hat{s}_3)$	6.65	9.01
$\theta(s_4, \hat{s}_4)$	6.41	11.83
$\theta(s_5, \hat{s}_5)$	0.12	1.99

Table 5–12 shows the initial and the final values of the objective function for both approaches. It also shows the execution times of processing 1000 simulated pixels using the two-stage and the penalty approaches. In both approaches, we

Table 5–11: Percentage error of the estimated endmembers for Two-stage and Penalty approaches(%)

	Two-stage Approach	Penalty Approach
$(s_1, \hat{s}_1)$	5.99	15.45
$(s_2, \hat{s}_2)$	5.06	21.60
$(s_3, \hat{s}_3)$	13.56	18.42
$(s_4, \hat{s}_4)$	11.28	24.54
$(s_5, \hat{s}_5)$	0.24	7.23

Table 5–12: Initial and final values of objective function and execution time for Two-stage and Penalty approaches

	Two-stage Approach	Penalty Approach
E_0	6.2055	6.2055
E_f	0.0084	5.123
Time(sec)	650.55	125.15

started with the same initialization. The final value of the objective function is lower for the two-stage approach. Although the execution time for the penalty approach is smaller than that of the two-stage approach, the spectral angle and the percentage error show that the two-stage performs better than penalty approach.

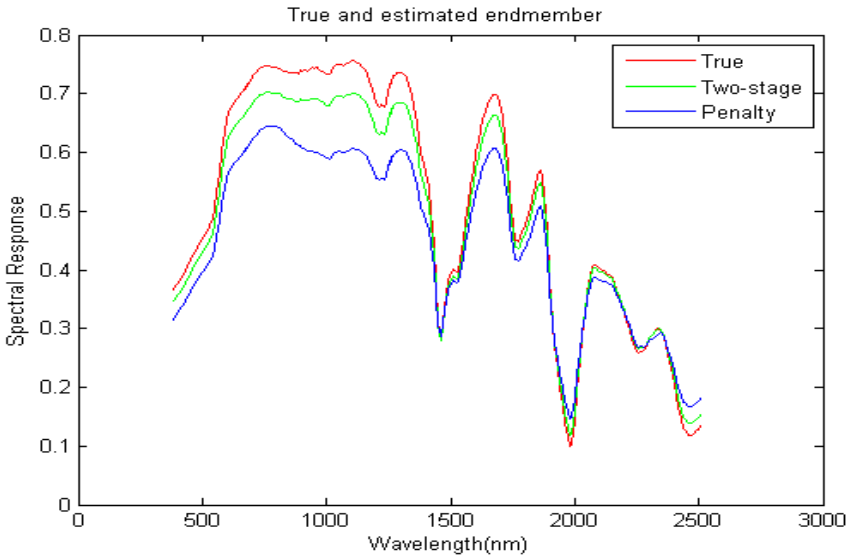


Figure 5–35: True and estimated spectral response of the first endmember

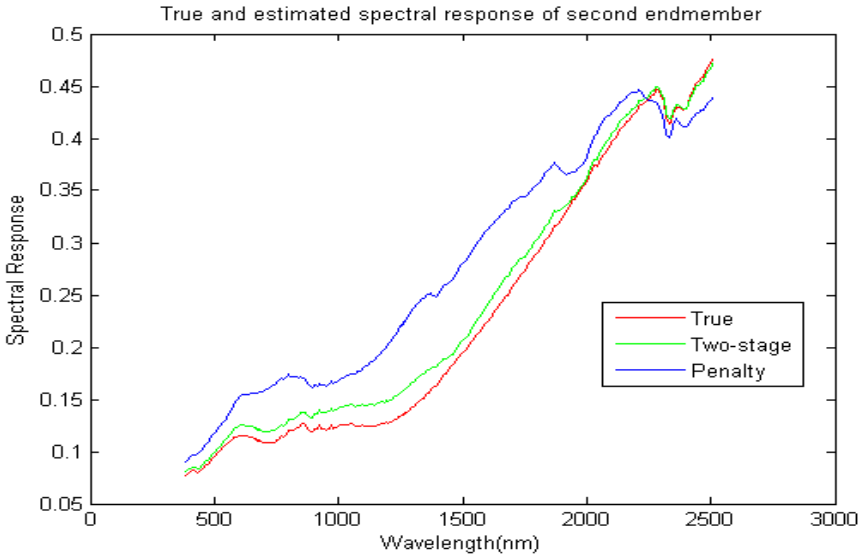


Figure 5–36: True and estimated spectral response of the second endmember

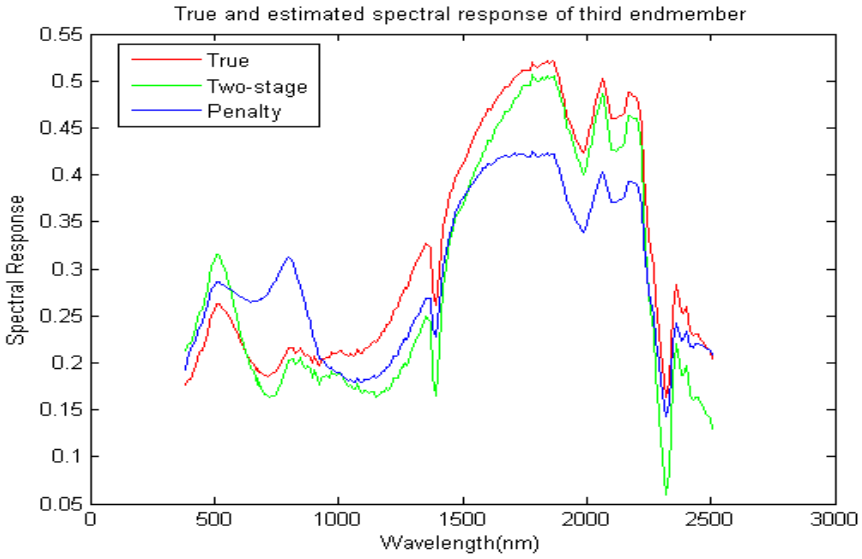


Figure 5–37: True and estimated spectral response of the third endmember

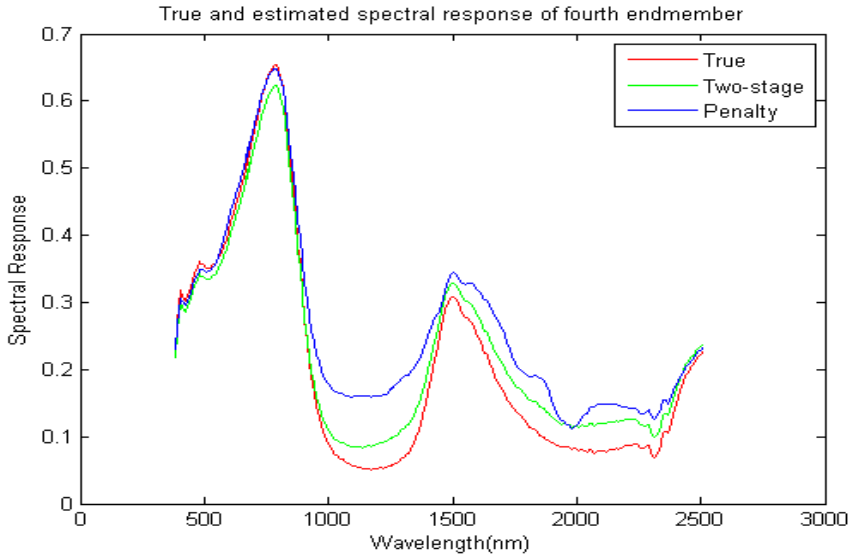


Figure 5–38: True and estimated spectral response of the fourth endmember

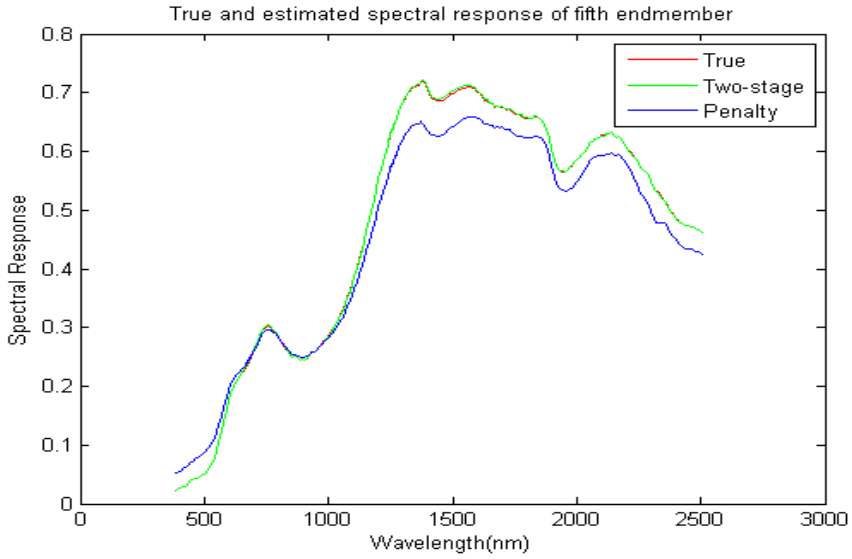


Figure 5–39: True and estimated spectral response of the fifth endmember

Figure 5–35 through Figure 5–39 show the true and estimated spectral response of the five endmembers used in this experiment. It can be seen from these figures and the tables that the two-stage estimate of the spectral signatures is closer than those obtained using the penalty approach.

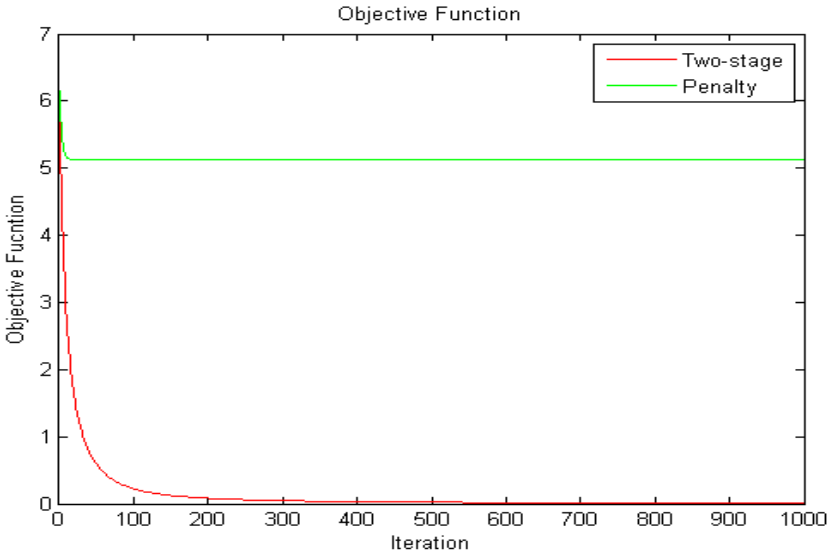


Figure 5–40: objective function

Figure 5–40 shows the objective function for both the two-stage and the penalty approaches. This figure shows that the two-stage goes to lower value of the objective function than the penalty approach.

5.9 Summary

In this Chapter, we presented results using simulated data. We conducted experiments using simulated data to make comparison between different initialization schemes. We used two metrics to compare between results. We conducted experiment to compare the two-stage approach and the penalty approach. The results showed that the SVDSS performs better than other used initialization schemes. In the comparison between algorithms, we can see that the two-stage approach performs better than the penalty approach. For the rest of our experiments using

real data, we will use the SVDSS as our initialization scheme. We estimated the simulated data dimension using different approaches. We also presented results to evaluate the effect of the number of spectral bands, the number endmembers, and the noise on the results.

CHAPTER 6

Experimental Results using Real Data

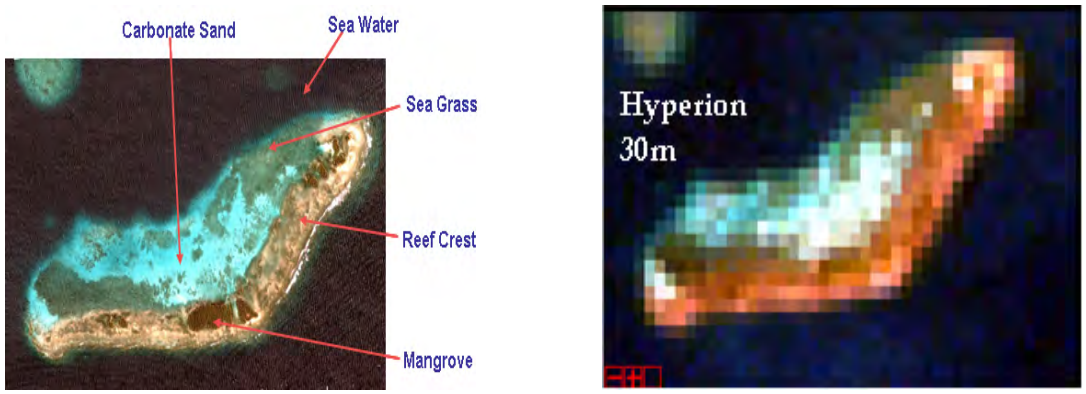
6.1 Introduction

In this chapter, we apply the developed algorithm to real data. We conducted different experiments using different real data sets. Based on the results of Chapter 5, we decide to use the SVDSS initialization scheme in our experiments for the rest of this chapter.

In this chapter, we use real data set to validate the developed algorithms. We use the Enrique Reef Hyperion image, Cuprite AVIRIS image, A.P. Hill AVIRIS image, and Biochemical data sets.

6.2 Enrique Reef Data

Figure 6–1(b) shows a small subset (29x49x90) of a HYPERION image of the Enrique Reef in southwestern Puerto Rico used in the experimental work shown here. The HYPERION image is 30m resolution. The low resolution motivates the spectral unmixing. The image contains areas of deep water, sea grass, carbonate sand, and reef crest as shown in the reference image Figure 6–1(a).



(a) (b)

Figure 6–1: Images of Enrique Reef acquired by: (a)IKONOS 1m resolution multispectral sensor, and (b) HYPERION 30m resolution hyperspectral sensor

6.2.1 Determination of the Number of Endmembers

One important issue in unsupervised unmixing is to determine the number of endmembers. In Chapter 2, we discussed some approaches to determine the number of endmembers. Among these approaches was by plotting the mean squared error for different number of endmembers. The obtained curve has a plateau; the number of endmembers is chosen from the curve as the number where the plateau starts. We plot the mean squared error curve for different number of endmembers using the Enrique Reef HYPERION image; the obtained mean squared curve is shown in Figure 6–2.

It can be seen from Figure 6–2, that the plateau starts when the number of endmembers is five. We also used different approaches to estimate the dimensionality of hyperspectral data. The dimension estimates of the used approaches are shown in Table 6–1. It can be seen from Table 6–1 that all approaches estimated the dimension to be five except the Kaiser-Guttman method which estimated the dimension to be 4. Since the number of endmember could be the same as or one greater than the data dimension we used the number of endmembers to be five. The spatial distribution

in the reference image (Figure 6–1(a)) makes the estimation of five endmembers to be reasonable.

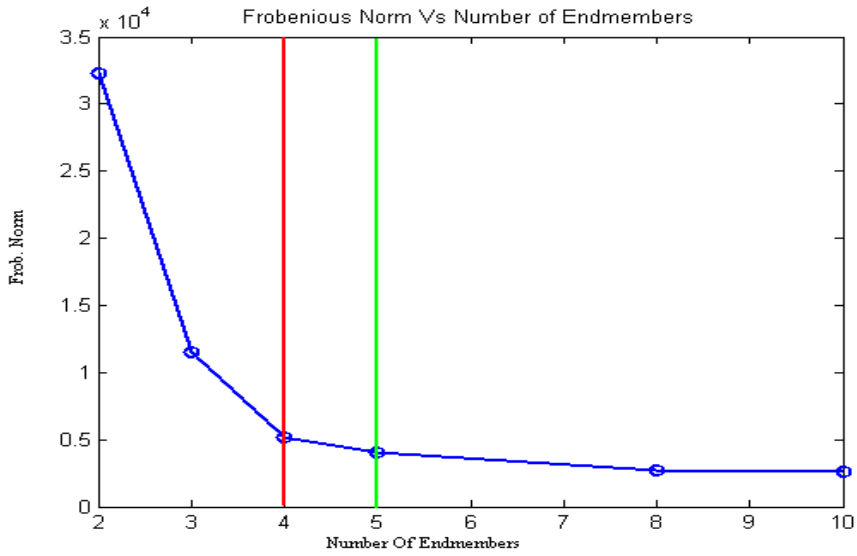


Figure 6–2: Mean squared error curve for Enrique Reef HYPERION image

Table 6–1: Enrique Reef data dimension estimate

Approach	SVD	KG	CSD	Fukunaga
Dimension	5	4	5	5

Herein, we present the results obtained by processing the HYPERION image shown in Figure 6–1(b) using the two-stage approach, and the Penalty approach. For comparison purposes, we present the results obtained by processing the same image using PPI, and N-FINDR approaches. The two-stage approach and penalty approach are discussed in details in Section 4.2. We present the spectral response estimated using the mentioned approaches and the corresponding abundance maps.

Figure 6–3 shows the sea grass spectral response estimated using PPI, N-FINDR, two-stage, and penalty approaches. Although the estimated spectral response from all approaches agree in their shape, we can clearly see the difference

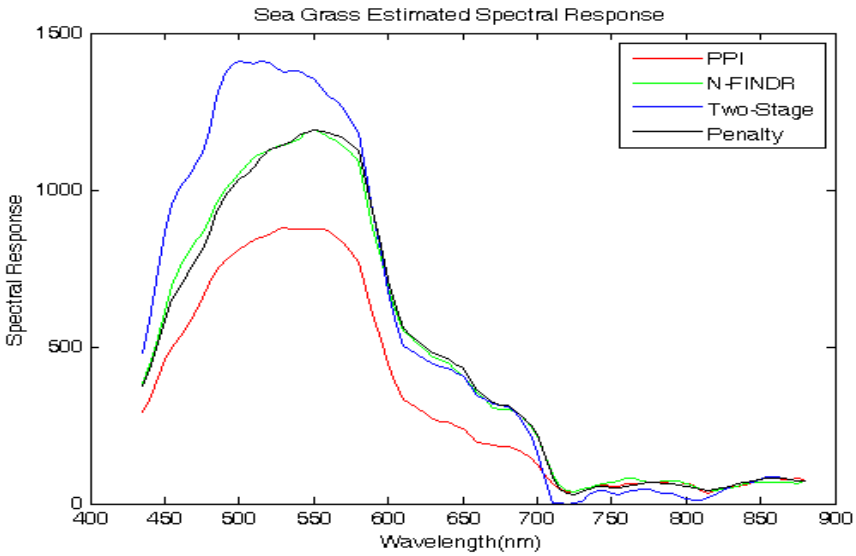
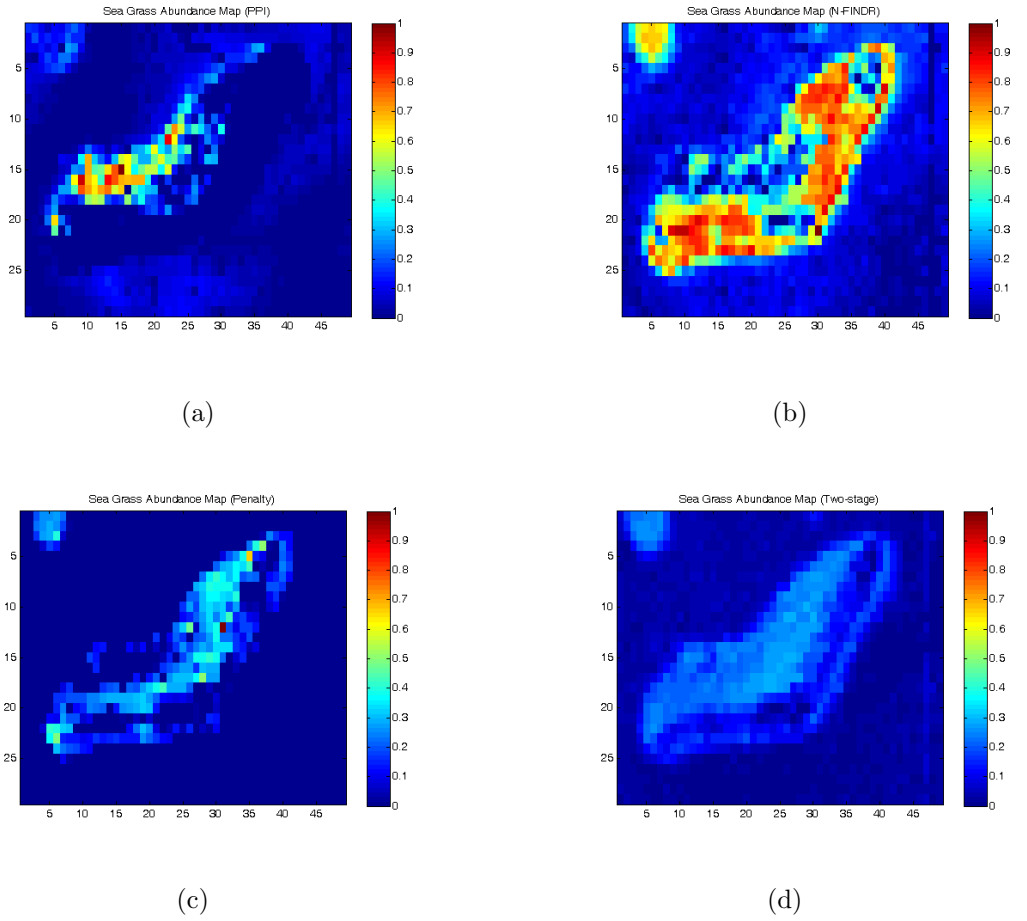


Figure 6–3: Sea grass estimated spectral response

in amplitude; this is clear in the spectral region (380-600)nm. We can see that the penalty approach and the N-FINDR approach agree completely in both shape and amplitude. The PPI approach produces the lower amplitude where the two-stage approach produces the maximum amplitude.

The sea grass abundance maps are shown in Figure 6–4. It can be seen from Figure 6–4 that the grass abundance maps obtained using penalty approach agree with those obtained using PPI, and N-FINDR. The sea grass abundance map obtained using the two-stage approach is different from those obtained using PPI, N-FINDR, and penalty approaches. As discussed in the previous section, the magnitude of the grass spectral response is higher for the two-stage case which results in lower abundance fractions. The abundance maps obtained from does not completely agree with the reference IKONOS image (Figure 6–1(a)). Abundance maps obtained from the PPI approach produced the worst spatial distribution of the grass compared to the reference image. The abundance maps obtained using the N-FINDR, the two-stage, and the penalty approaches are more reasonable because the sea grass region



(a) (b)

(c) (d)

Figure 6–4: Grass abundance maps obtained using (a)PPI (b)N-FINDR (c) Penalty (d) Two-stage

shown in the IKONOS reference image were recovered. Also the sand region were recovered.

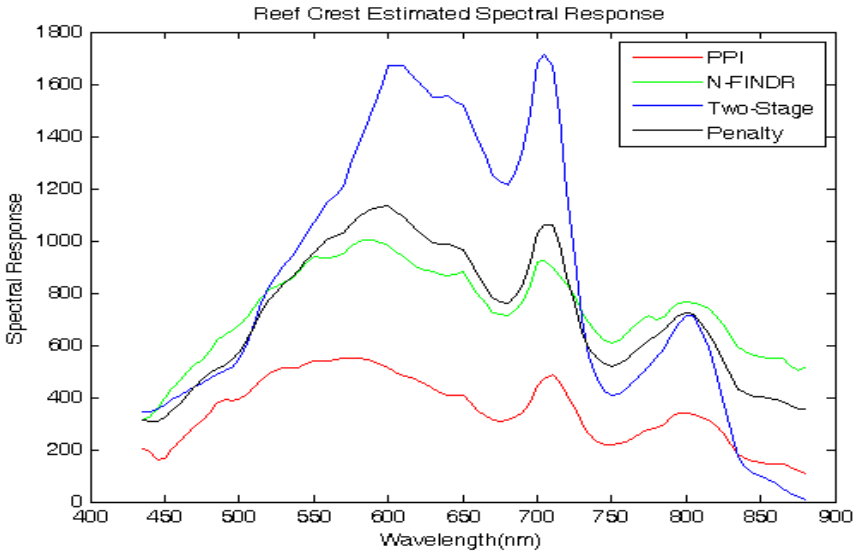


Figure 6–5: Crest reef estimated spectral response

Figure 6–5 shows the crest reef spectral response estimated using PPI, N-FINDR, two-stage, and penalty approaches. It can be seen that the estimated spectral response from all approaches agree in their shape, we can clearly see the difference in amplitude; this is clear in the spectral region (380-700)nm. We can see that the penalty approach and the N-FINDR approach agree completely in shape and close in amplitude. The PPI approach produces the lower amplitude where the two-stage approach produces the maximum amplitude.

Figure 6–6 shows the crest reef abundance maps obtained using the PPI, N-FINDR, two-stage, and penalty approaches. Comparing the contained abundance maps with the IKONOS reference image (Figure 6–1(a)), we can see that all the approaches performed well in recovering the spatial distribution of the crest reef. Although the concentration scale is smaller in the two-stage case; this is due to the fact that the amplitude of the spectral response obtained using the two-stage approach is higher than the other approaches.

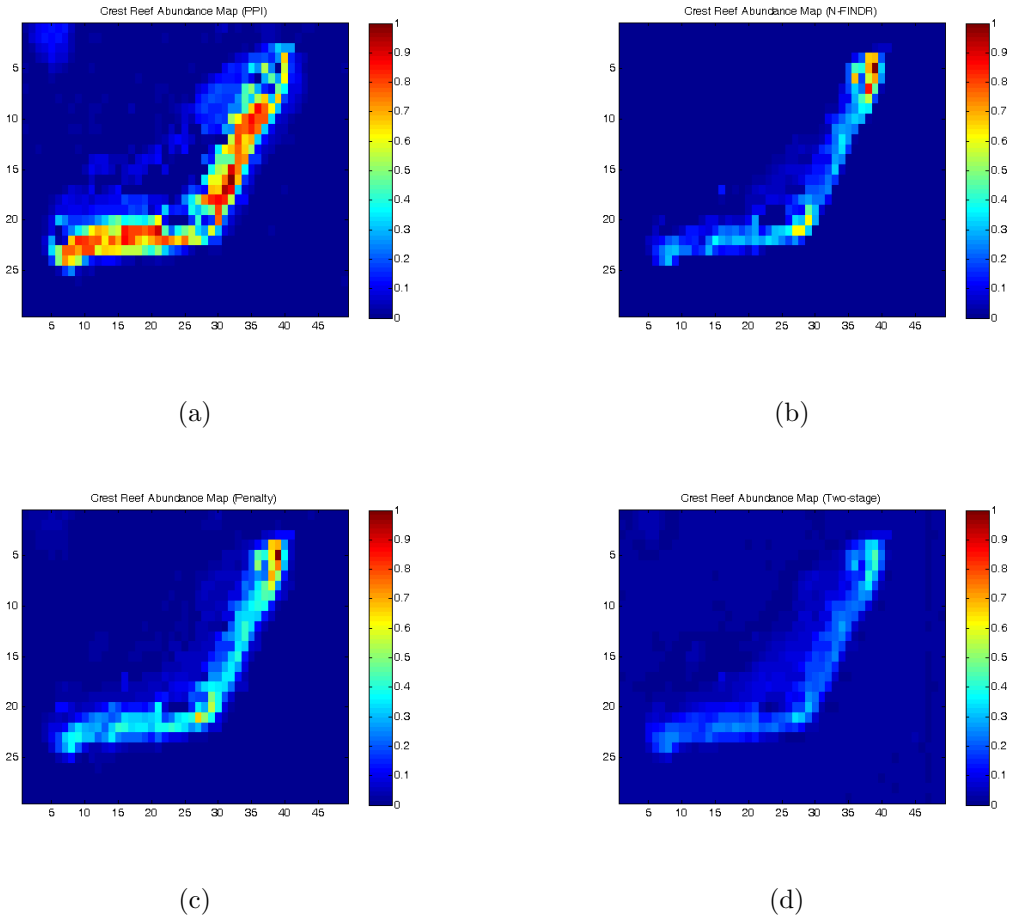


Figure 6–6: Crest reef abundance maps obtained using (a)PPI (b)N-FINDR (c) Penalty (d) Two-stage

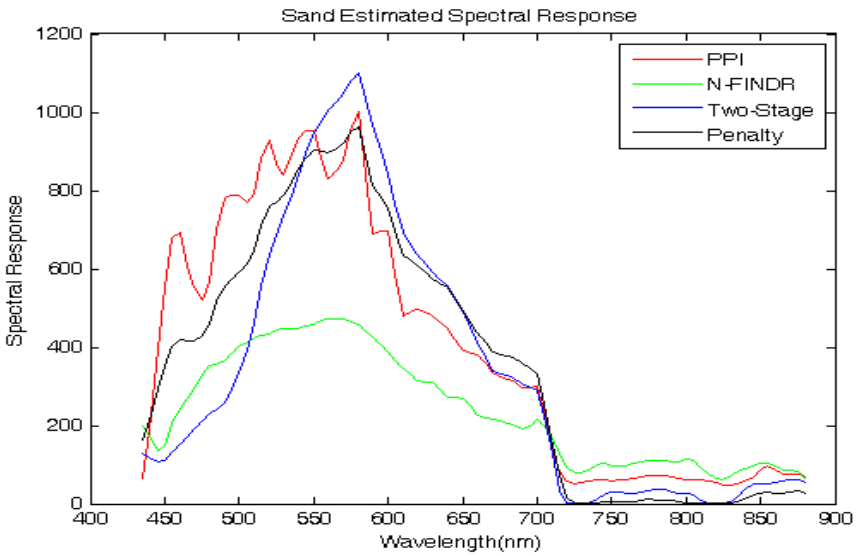
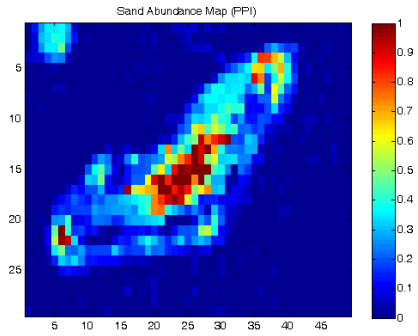
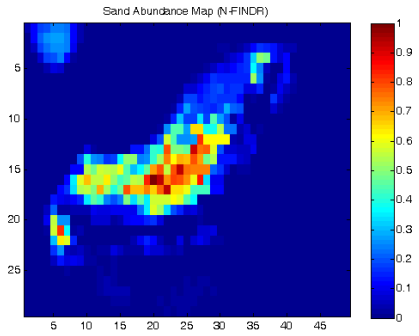


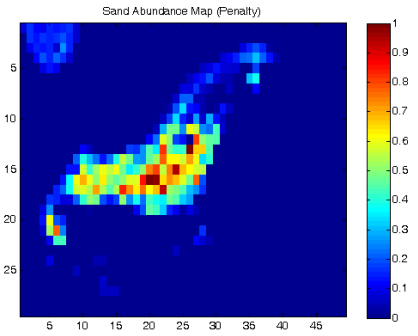
Figure 6–7: Sand estimated spectral response



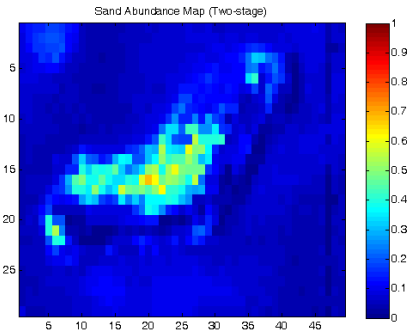
(a)



(b)



(c)



(d)

Figure 6–8: Sand abundance maps obtained using (a)PPI (b)N-FINDR (c) Penalty (d) Two-stage

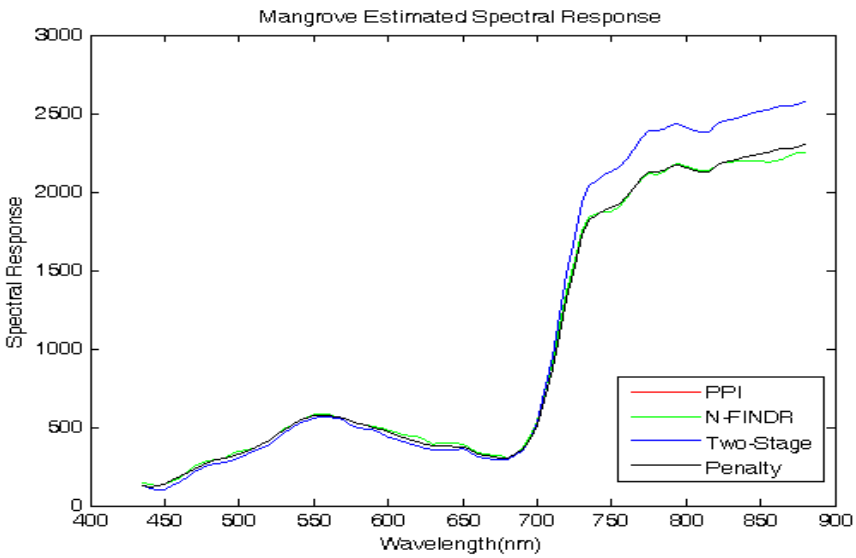
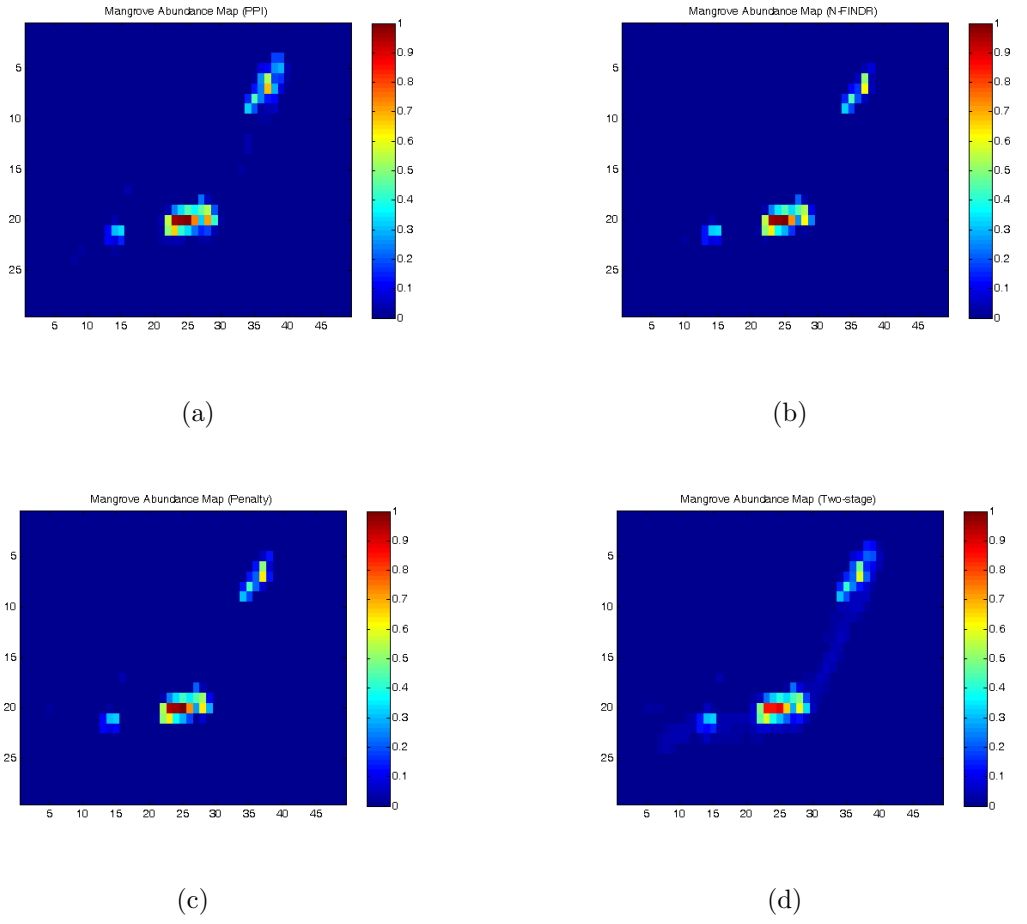


Figure 6–9: Mangrove estimated spectral response

Figure 6–7 shows the sand spectral response estimated using PPI, N-FINDR, two-stage, and penalty approaches. In this case, the spectral response looks different for each approach. We can see that the N-FINDR approach produced a spectral response with the lowest amplitude.

Figure 6–8 shows the sand abundance maps obtained using the PPI, N-FINDR, two-stage, and penalty approaches. Comparing the obtained abundance maps with the IKONOS reference image (Figure 6–1(a)), we can see that also for the sand, all the approaches performed well in recovering the spatial distribution of the sand. We can see from the abundance maps that there is an overlap between the sand and the sea grass.

Figure 6–9 shows the mangrove spectral response estimated using PPI, N-FINDR, two-stage, and penalty approaches. It can be seen that the estimated spectral response from all approaches coincides, except in the two-stage which is higher in amplitude in the spectral region greater than 700nm.



(a) (b)

(c) (d)

Figure 6–10: Mangrove abundance maps obtained using (a)PPI (b)N-FINDR (c) Penalty (d) Two-stage

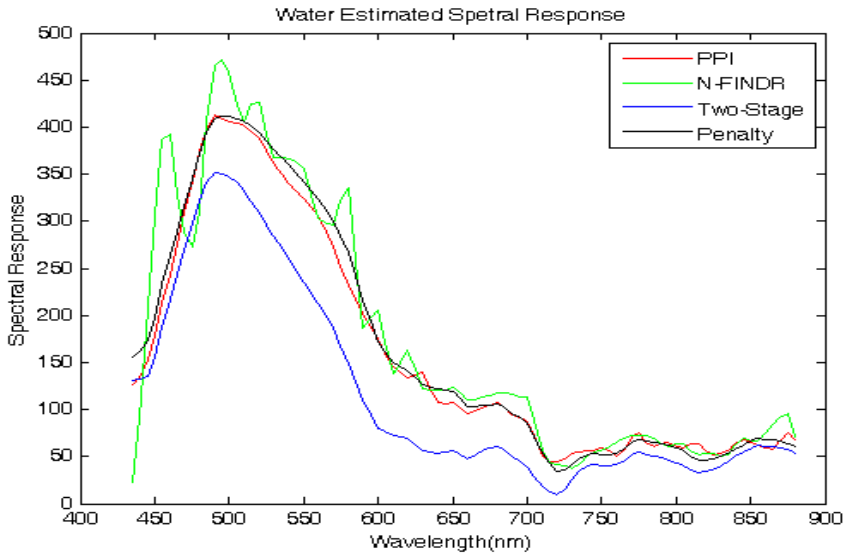
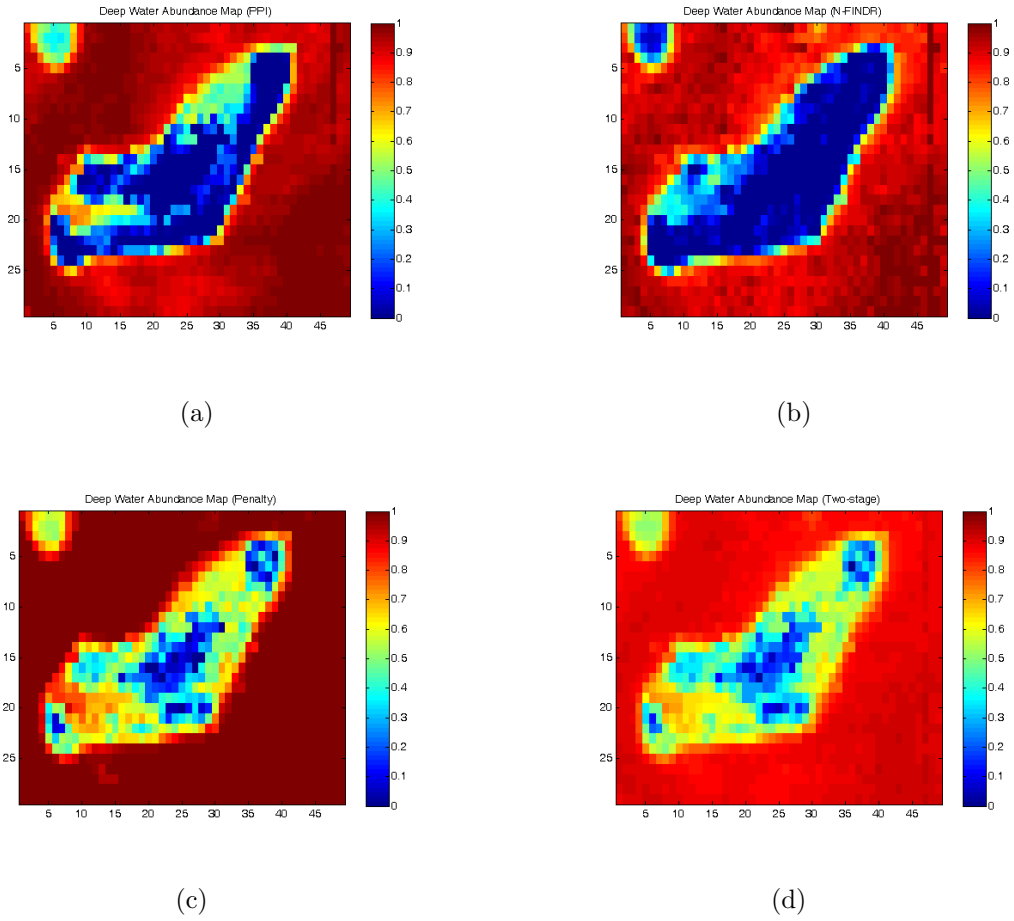


Figure 6–11: Water estimated spectral response

Figure 6–10 shows the mangrove abundance maps obtained using the PPI, N-FINDR, two-stage, and penalty approaches. Comparing the contained abundance maps with the IKONOS reference image (Figure 6–1(a)), we can see that all the approaches performed well in recovering the spatial distribution of the mangrove.

Figure 6–11 shows the deep water spectral response estimated using PPI, N-FINDR, two-stage, and penalty approaches. It can be seen that the estimated spectral response from the N-FINDR approach has a different shape from all other approaches. The spectral response estimated using the PPI approach and the penalty approach agree completely with a very small difference in amplitude.

Figure 6–12 shows the deep water abundance maps obtained using the PPI, N-FINDR, two-stage, and penalty approaches. Comparing the abundance maps with the IKONOS reference image (Figure 6–1(a)), we can see that all the approaches succeeded in recovering the spatial distribution.



(a) (b)

(c) (d)

Figure 6–12: Water abundance maps obtained using (a)PPI (b)N-FINDR (c) Penalty (d) Two-stage

6.2.2 Measure of Fitness

In this section, we use the $\mathbf{R}^2$ plots, and RMS residual plots to measure how good are the estimated endmembers to fit the measured data.

Table 6–2 shows the minimum value, maximum value, average value, and the standard deviation of $\mathbf{R}^2$ values for PPI, N-FINDR, two-stage, and penalty approaches. Figure 6–13 to 6–16 show the $\mathbf{R}^2$ plot obtained using PPI, N-FINDR, two-stage, and penalty approaches, respectively. It can be seen from Table 6–2, and the figures that the two-stage $\mathbf{R}^2$ has and advantage over the other approaches in that, it has the highest average and the lowest standard deviation. On the other hand, the N-FINDR has the lowest average and the highest standard deviation. This indicates that in the two-stage case, most of the pixels were modeled well. The low $\mathbf{R}^2$ value means that the estimated endmembers along with their abundance fractions do not fit the corresponding measured pixel well.

Table 6–2: $\mathbf{R}^2$ value (Enrique Reef Image)

Approach	Min $\mathbf{R}^2$ value	Max $\mathbf{R}^2$ value	Average $\mathbf{R}^2$ value	Standard deviation
PPI	0.8820	1.000	0.9938	0.0086
N-FINDR	0.8590	1.000	0.9780	0.0170
Two-stage	0.9773	0.9999	0.9983	0.0016
Penalty	0.9409	0.9999	0.9954	0.0054

The $\mathbf{R}^2$ maps for the PPI, N-FINDR, two-stage, and penalty approaches are shown in Figure 6–17. Pixels with low $\mathbf{R}^2$ values indicate that these pixels are not modeled well by the estimated endmembers. It can be seen that Figure 6–17(a) to 6–17(c) still have some spatial structure while Figure 6–17(d) does not. That signals, in my opinion, that the retrieved model using the two-stage approach fits well the data.

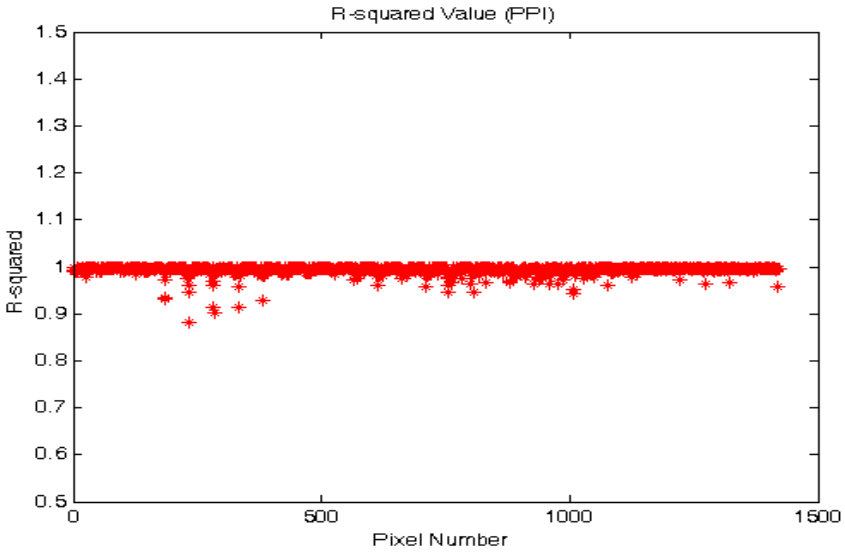


Figure 6–13: R^2 value for PPI approach

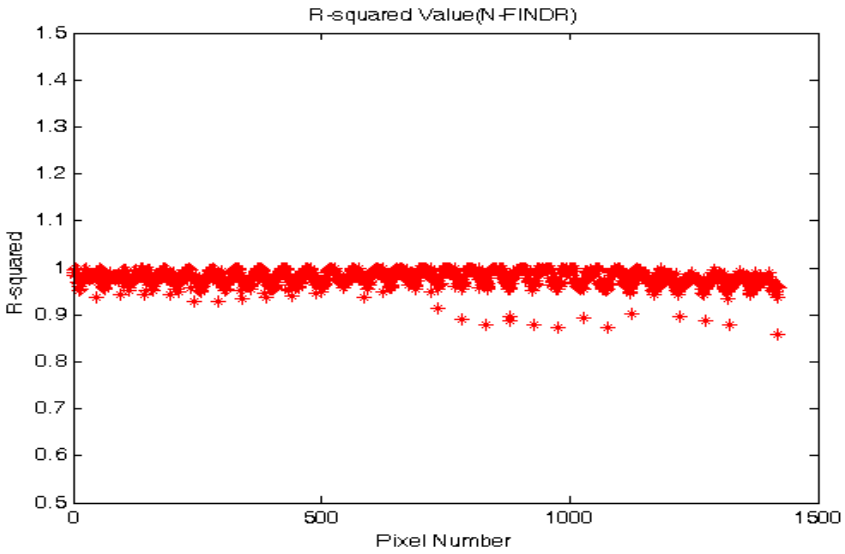


Figure 6–14: R^2 value for N-FINDR approach

Table 6–3: RMS residuals (Enrique Reef Image)

Approach	Min rms value	Max rms value	Average rms	Standard deviation
PPI	0	259.05	21.06	21.25
N-FINDR	0	78.03	32.49	9.17
Two-stage	4.69	49.73	10.10	4.35
Penalty	5.85	100.42	15.92	7.40

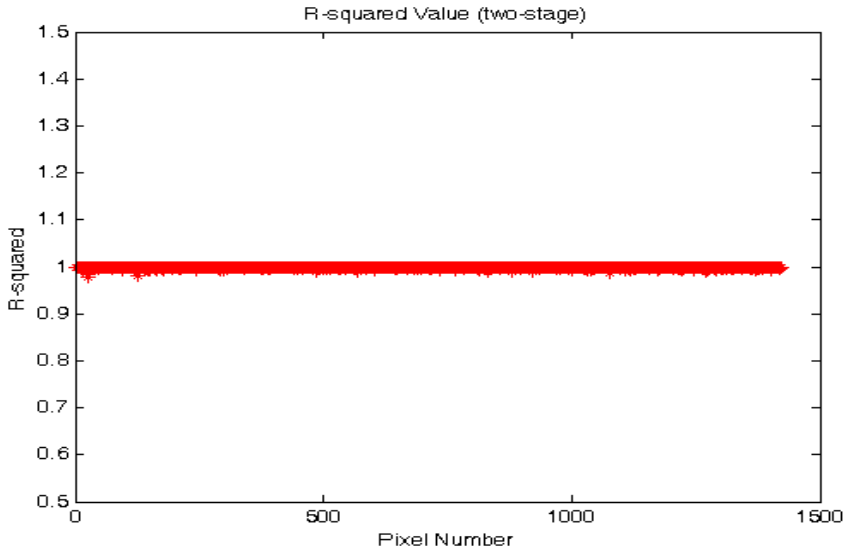


Figure 6–15: R^2 value for the two-stage approach

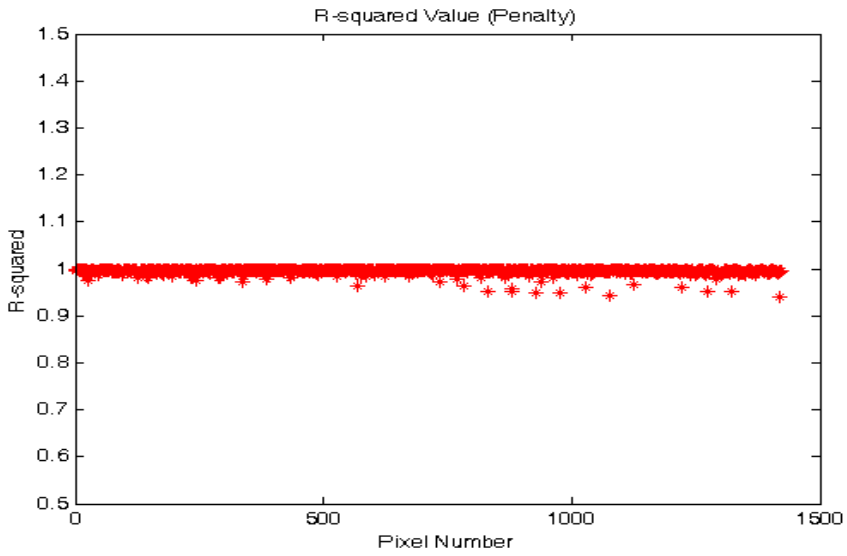
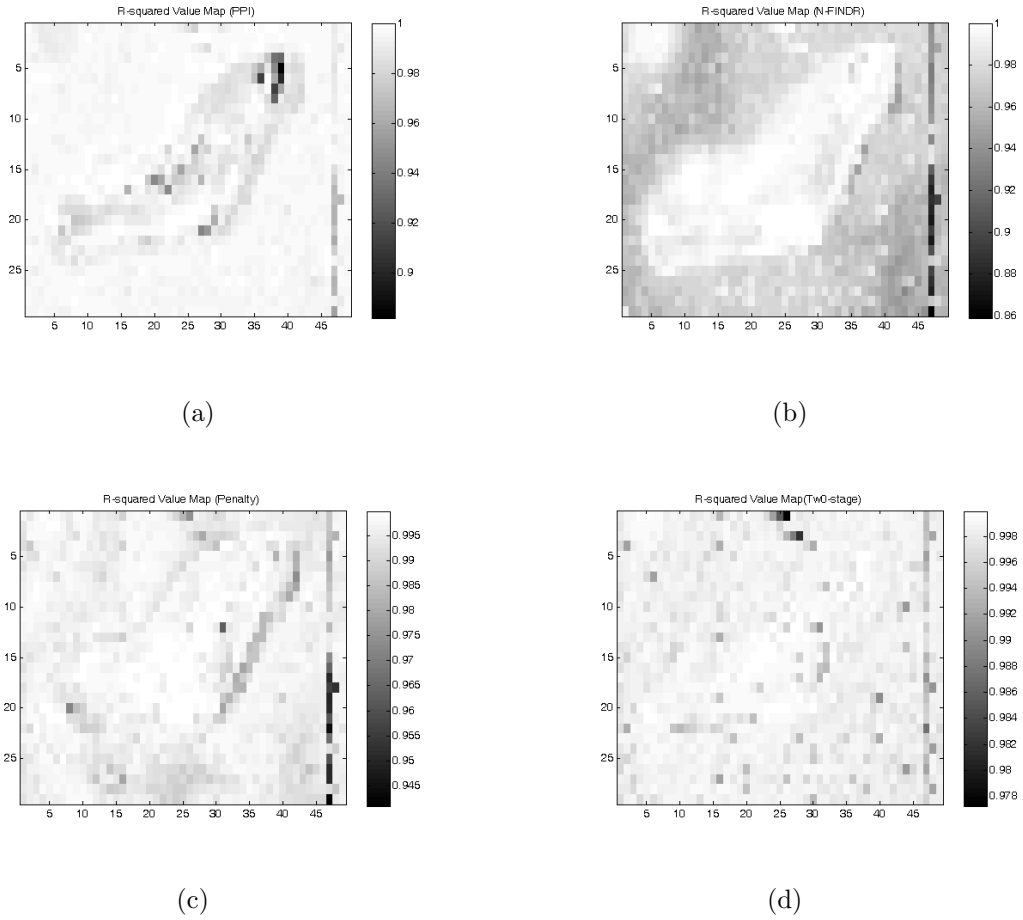


Figure 6–16: R^2 value for the penalty approach



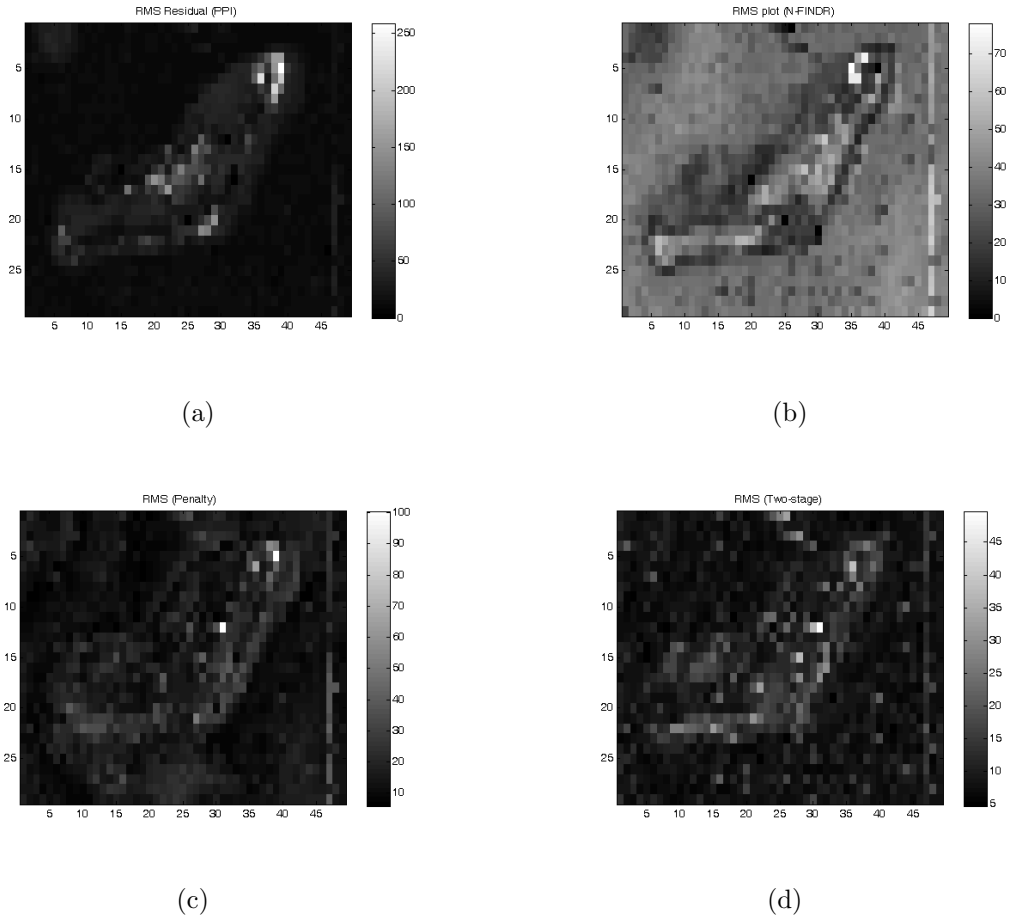
(a) (b)

(c) (d)

Figure 6–17: R^2 plot obtained using (a)PPI (b)N-FINDR (c) Penalty (d) Two-stage

Table 6–3 shows the minimum value, maximum value, average value, and the standard deviation of RMS residuals for PPI, N-FINDR, two-stage, and penalty approaches. Figure 6–18 was obtained using the mentioned approaches. We can see from Table 6–3, and Figure 6–18 that the two-stage approach has the lowest average and the lowest standard deviation, which indicates that the two-stage approach produced a model that fits the measured data better than the other approaches. Table 6–3 shows zero minimum RMS value for PPI, and N-FINDR approaches since pixels from the image are used as endmembers.

Figure 6–19 shows how the objective function value decreases during the iterations, it shows also the value of the objective function for the PPI, and the N-FINDR



(a) (b)

(c) (d)

Figure 6–18: RMS residual plot obtained using (a)PPI (b)N-FINDR (c) Penalty (d) Two-stage

approaches for comparison purposes. It is evident that the two-stage approach has lower objective value than other approaches.

6.3 Cuprite Data

Figure 6–20(a) shows the Cuprite image segment used in this experiment. We use hyperspectral data from the Airborne Visible and Infrared Imaging Spectrometer (AVIRIS) which has been developed and is operated by the Jet Propulsion Laboratory (JPL). The AVIRIS system consists of four spectrometers that produce a

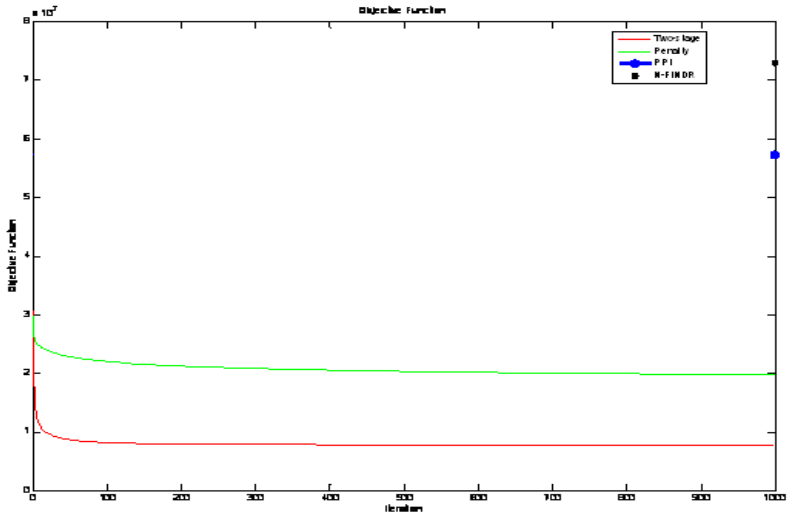
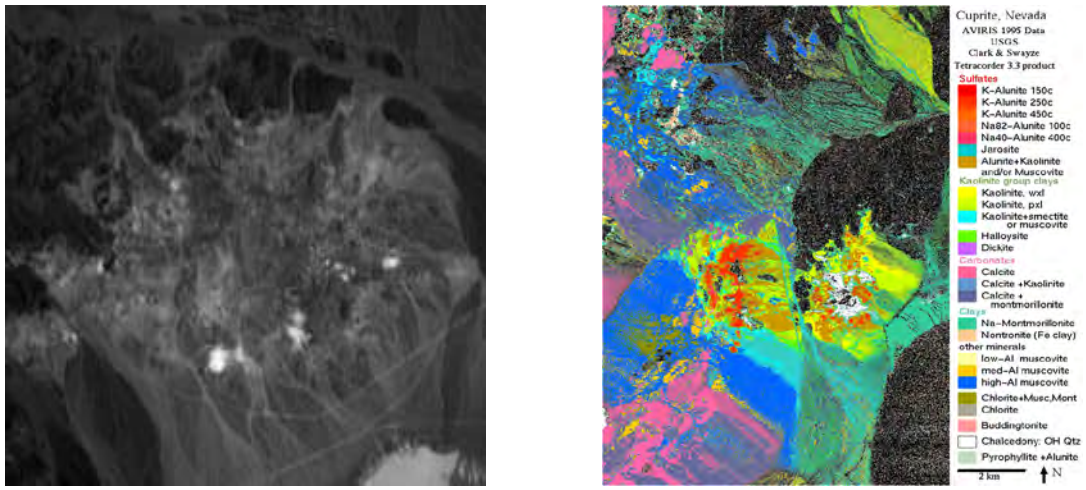


Figure 6–19: Objective function value at each iterations (Enrique Reef image)

hyperspectral image in the visible to short wave infrared band (320 to 2500 nanometers) recorded on 224 channels in whisk broom mode. The Cuprite image was taken over the mining district, 2 km north of Cuprite, Nevada. This image contains has a total of 640x2378 pixels and 224 bands in the 370-2500 nm range. We selected a portion of the Cuprite image, of size 191x250 pixels that corresponds to part of the mineral mapping in the Cuprite mining district [62]. We selected from this image 188 bands that corresponds to the 390-2470 nm spectral range. We use the USGS images as ground truth [62]. This image is usually used in unmixing studies. Different numbers of endmembers are identified. In [63], the number of endmembers identified for the Cuprite image segment shown in Figure 6–20(a), was 15, and 21 endmembers using two different approaches (see [63] for more details about the used approaches).

Our two-stage algorithm was applied to process the selected portion of the cuprite image. Table 6–4 shows the dimension estimates of the cuprite image dimension. We will use the number of endemembers equal to the SVD dimension estimate. Hence, the number of endmembers equal to five.



(a) (b)

Figure 6–20: Cuprite image:(a)Band 30 of the subimage of AVIRIS cuprite Nevada data set; (b)classification map

Table 6–4: Cuprite data dimension estimate

Approach	SVD	KG	CSD	Fukunaga
Dimension	5	4	6	3

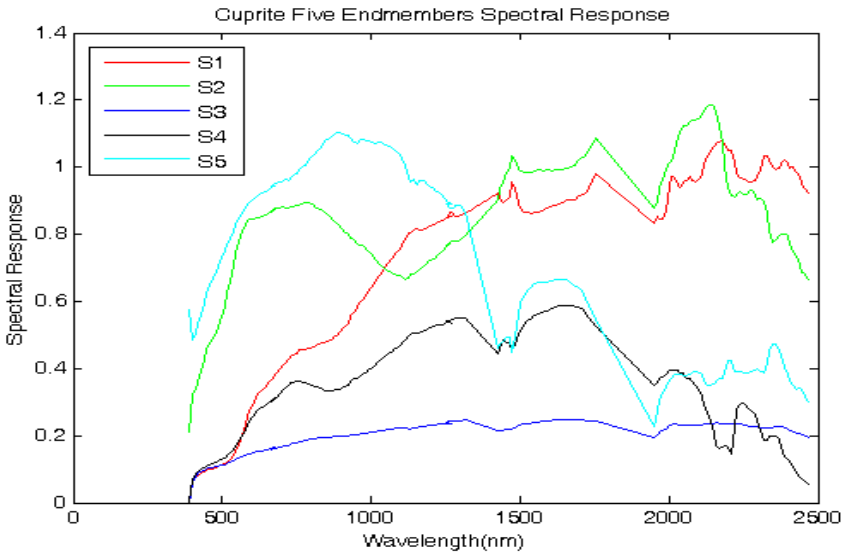
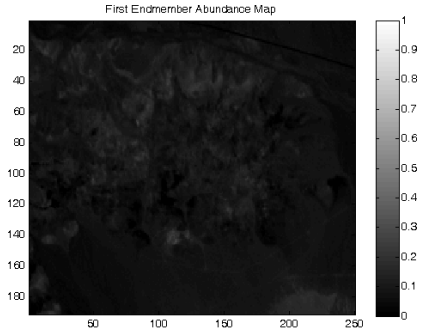
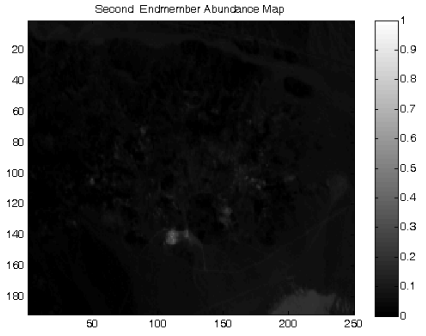


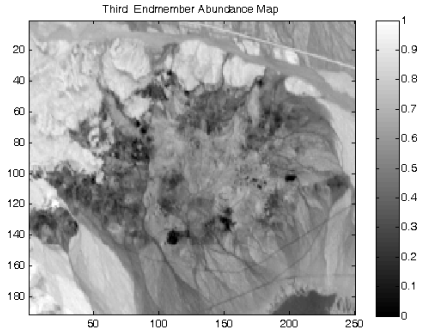
Figure 6–21: Cuprite estimated endmembers spectral response.



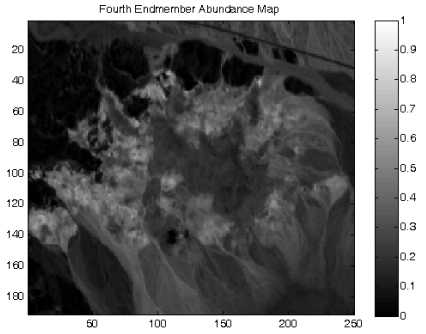
(a)



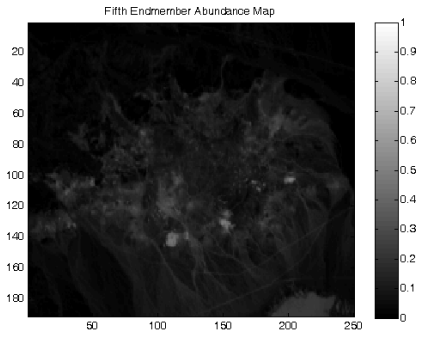
(b)



(c)



(d)

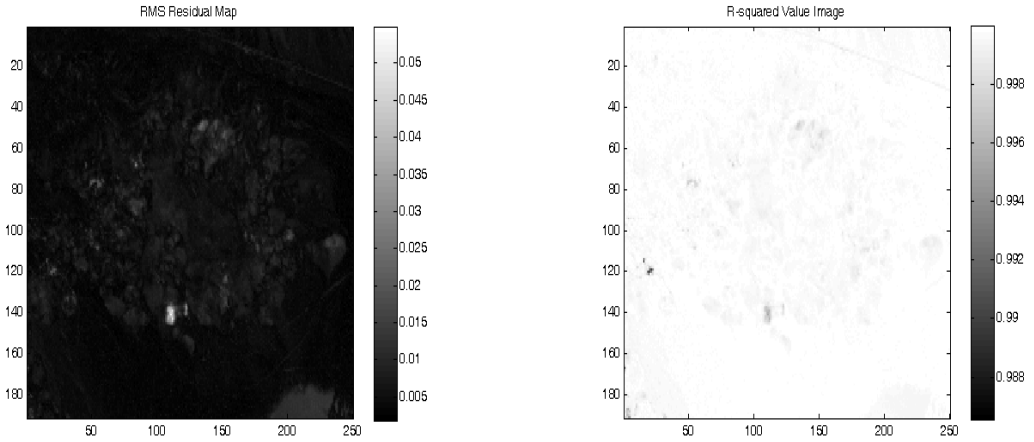


(e)

Figure 6–22: Cuprite abundance maps of (a)first (b)second (c) third (d) fourth (e) fifth, endmembers

Table 6–5: Cuprite RMS residuals (Cuprite Image)

Approach	Min rms value	Max rms value	Average rms
Two-stage	0.0018	0.055	0.0052



(a) (b)

Figure 6–23: Cuprite (a)RMS residual image (b) $\mathbf{R}^2$ image

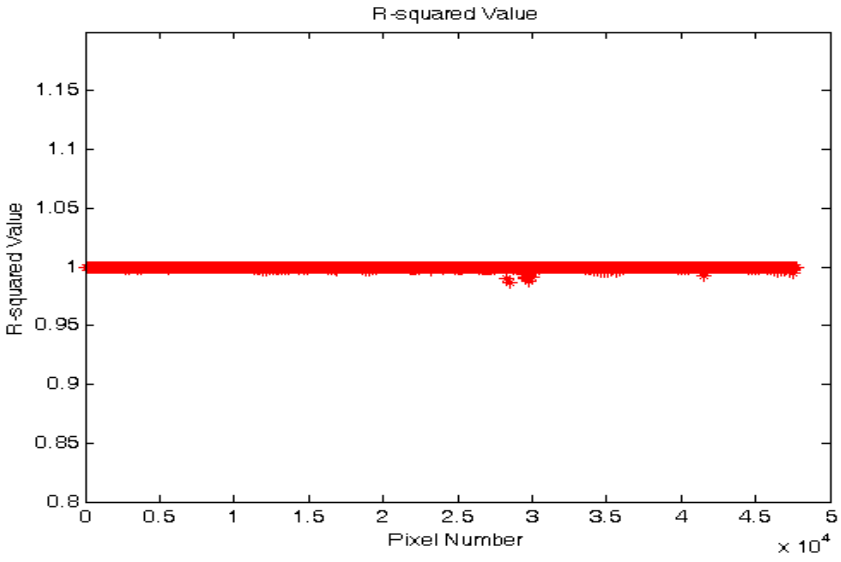


Figure 6–24: Cuprite Rsquared plot.

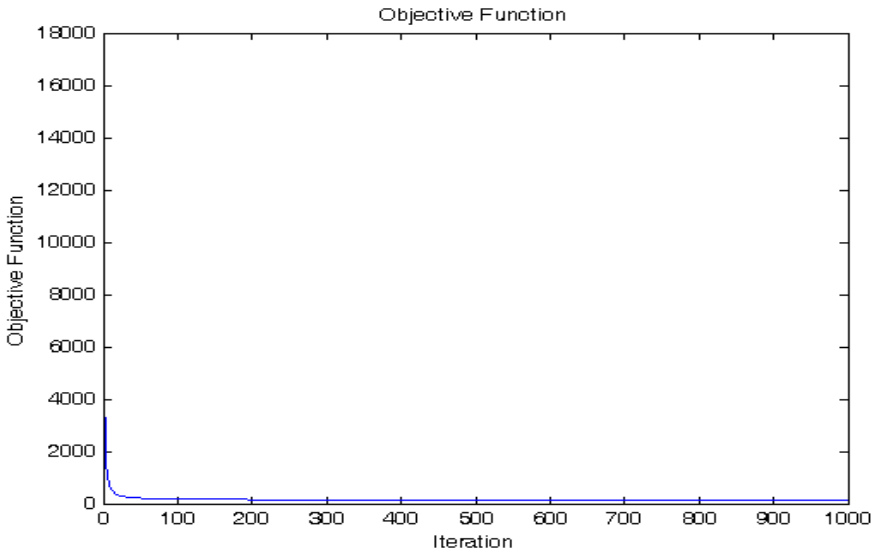


Figure 6–25: Cuprite mean squared error.

6.4 A.P. Hill Data

The AVIRIS data used in this experiment was taken over Fort A. P.Hill, Virginia. A color composite of the Fort A.P.Hill image is shown in Figure 6–26.

Figure 6–26: Fort A. P. Hill AVIRIS data cube [53].

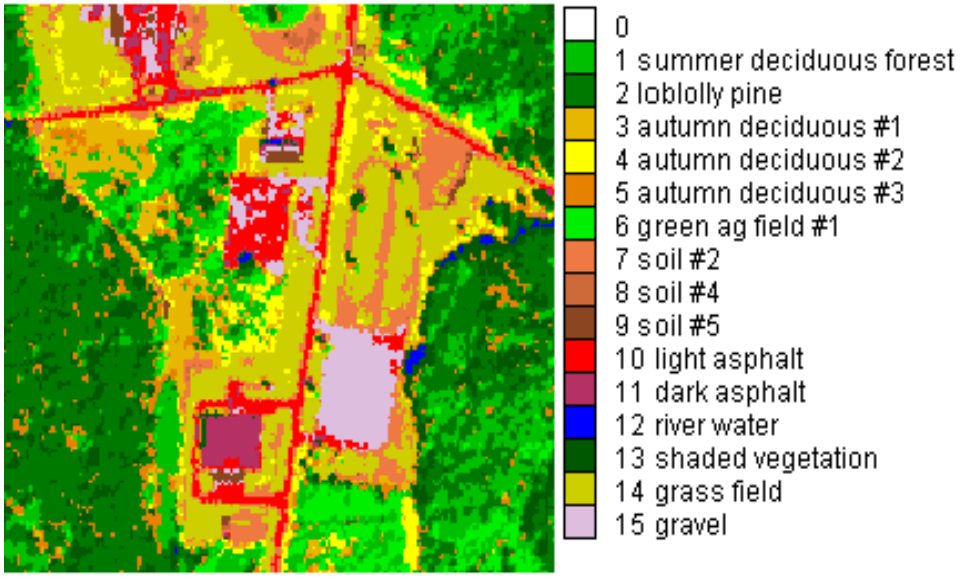


Figure 6–27: Classified Segment of Fort A.P. Hill [53].

Figure 6–27 shows a classified segment taken from Fort A.P.Hill image, obtained from [53]. The segment was classified using the spectral library developed in [53]. We need to compare the results obtained from our developed unmixing algorithms.

To estimate the number of endmembers of the A.P. Hill data we estimated the data dimension using different approaches. The dimension estimates are shown in Table 6–6. The mean squared error curve is shown in Figure 6–28. If we assume the number of endmembers to be equal to the data dimension, then using the the SVD dimension estimate shown in Table 6–6 we can estimate the number of endmembers to be twelve.

Table 6–6: A. P. Hill data dimension estimate

Approach	SVD	KG	CSD	Fukunaga
Dimension	12	4	7	5

Table 6–7 shows the $\mathbf{R}^2$ value average, RMS residual average, and their standard deviations. The $\mathbf{R}^2$ average value are close to one, which indicates that all choices are working well. The effect of number of endmember is clear if look at the RMS residual average in Table 6–7. It can be seen that after eight endmembers the RMS

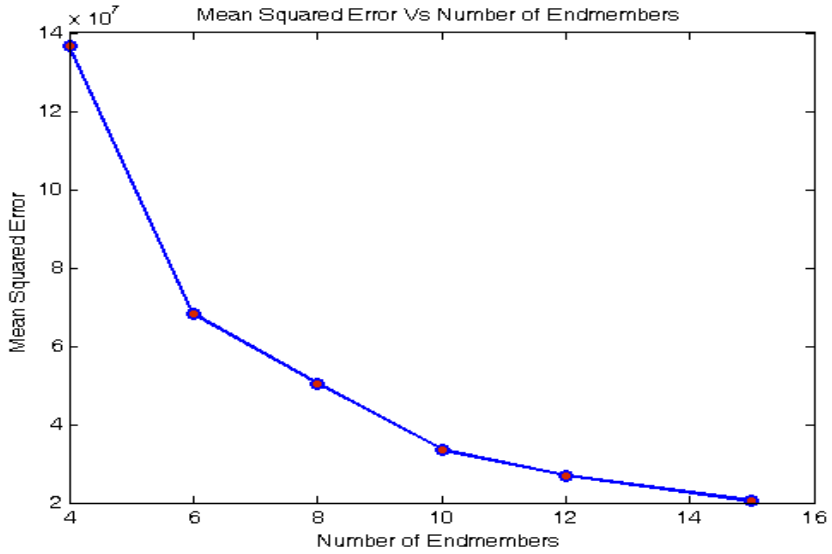


Figure 6–28: A.P. Hill mean squared error for different number of endmembers

Table 6–7: $\mathbf{R}^2$ and RMS results for different number of endmembers(A. P. Hill Image)

Num. of Endmembers	$\mathbf{R}^2$ Average	$\mathbf{R}^2$ Stdev	RMS Average	RMS Stdev
4	0.9902	0.0610	8.0925	5.7095
6	0.9956	0.0274	6.2359	3.1639
8	.9968	0.0208	5.2704	2.8892
10	0.9971	0.0251	4.4434	2.0721
12	0.9974	.0249	4.0173	1.7745
15	0.9978	0.0224	3.553	1.5364

average is slowly change. There is a tradeoff between the number of endmembers and the RMS residual average. The smaller RMS residual average, which corresponds to larger number of endmembers is the better. We need to evaluate the computational cost due to the increase of the number of endmembers.

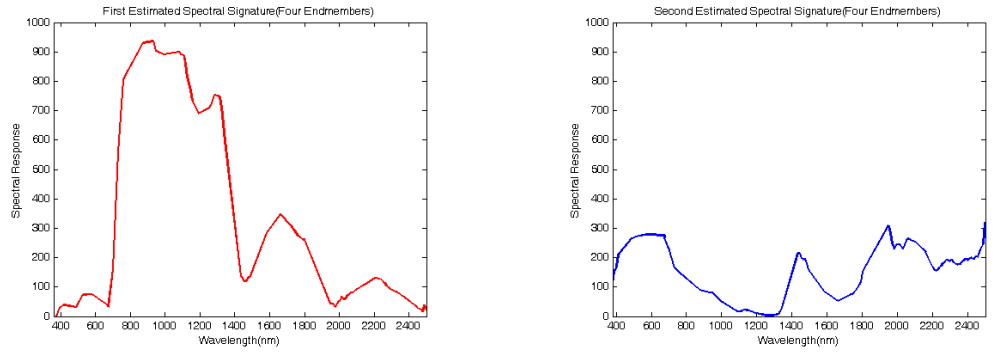
Figure 6–29 and 6–30 show the estimated four endmembers and their corresponding abundance maps, respectively. The first abundance map Figure 6–30(a) indicates that the high concentration of the first endmember shown in Figure 6–29(a) is in the area classified as green ag field#1 in the classified segment shown in Figure 6–27.

The second endmember abundance map is shown in Figure 6–30(b). The high concentration of the second endmember with spectral response shown in Figure 6–29(b) is corresponding to the gravel in the classified segment.

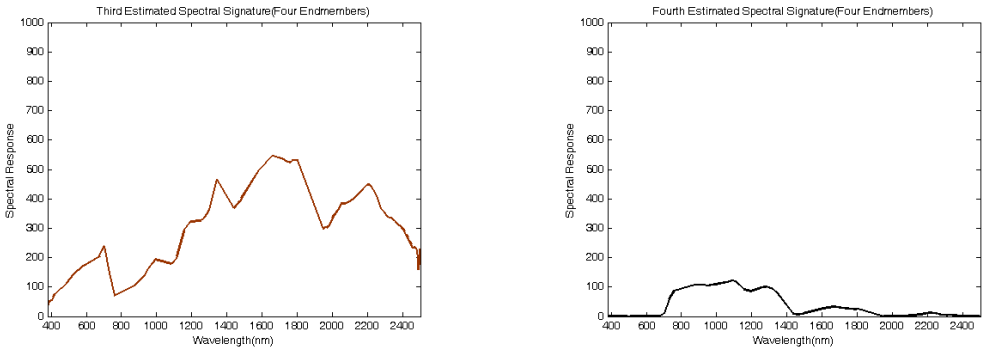
The third endmember abundance map is shown in Figure 6–30(c). The high concentration of the second endmember with spectral response shown in Figure 6–29(c) is corresponding to the soil #5 in the classified segment.

The fourth endmember abundance map is shown in Figure 6–30(d). The high concentration of the second endmember with spectral response shown in Figure 6–29(d) is corresponding to the shaded vegetation in the classified segment.

We present results for different number of endmembers to study the effect of having less or more endmembers on the results. To evaluate the quality of the estimated endmembers and how good are they to model the measured image, we present the RMS residual map for each case. Figure 6–31, 6–34, and 6–37 show the RMS residual image for four, twelve, fifteen endmembers, respectively. For the estimated endmembers to fit well the measured image, the RMS residual image



(a) (b)



(c) (d)

Figure 6–29: A. P. Hill estimated four spectral signatures

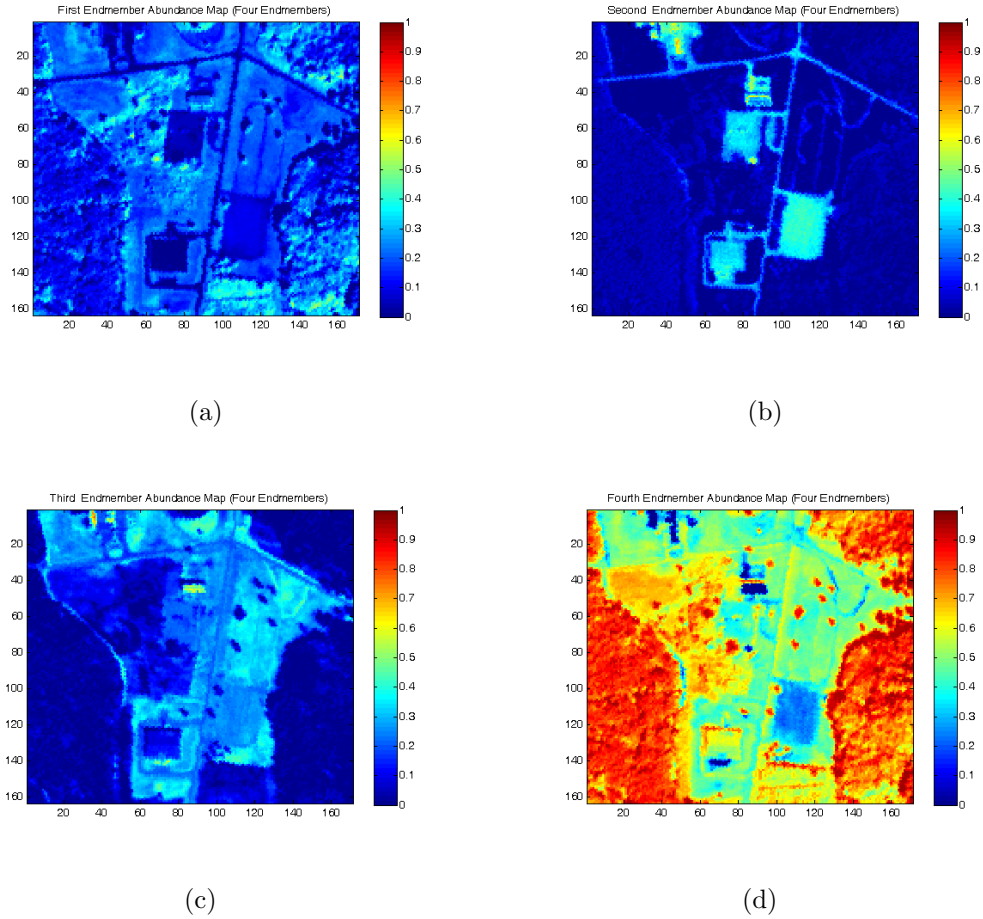


Figure 6–30: A. P. Hill abundance maps for four endmembers

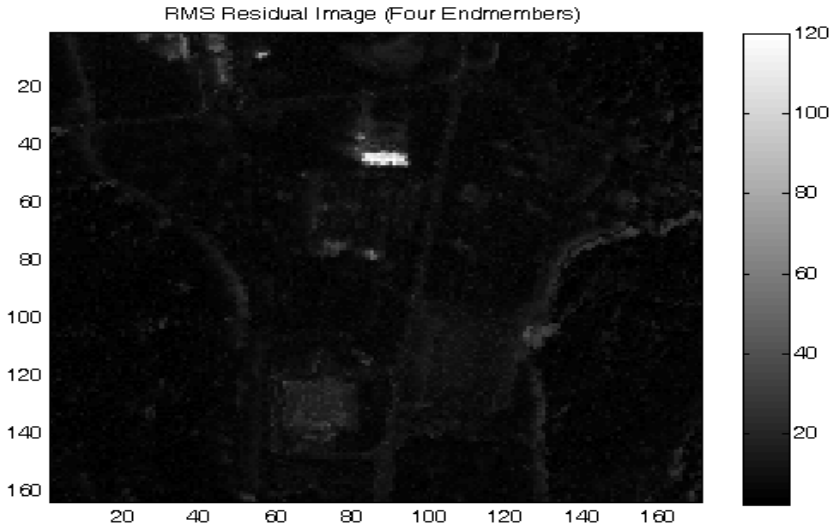
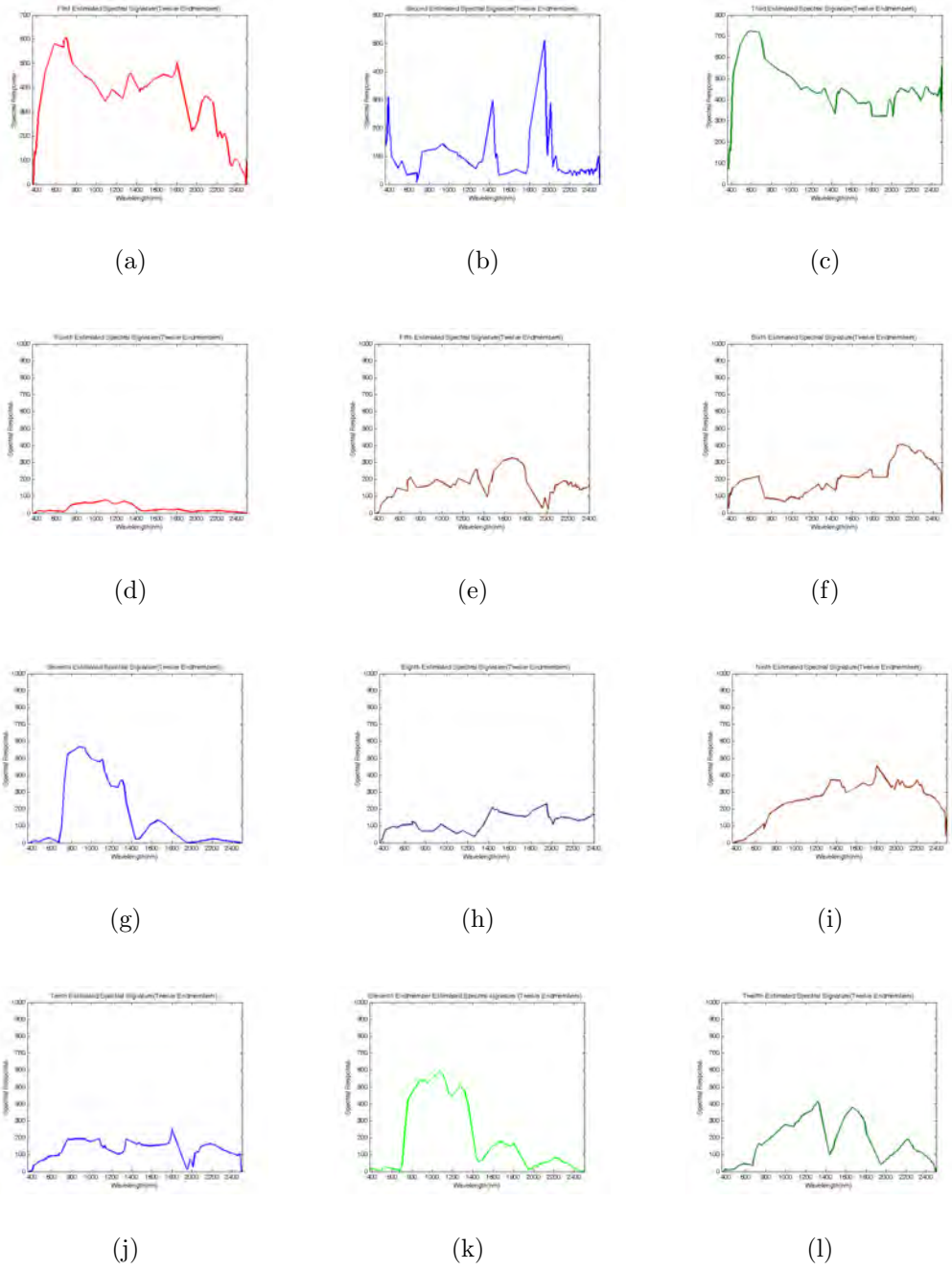


Figure 6–31: Fort A.P. Hill four endmembers RMS residual map.



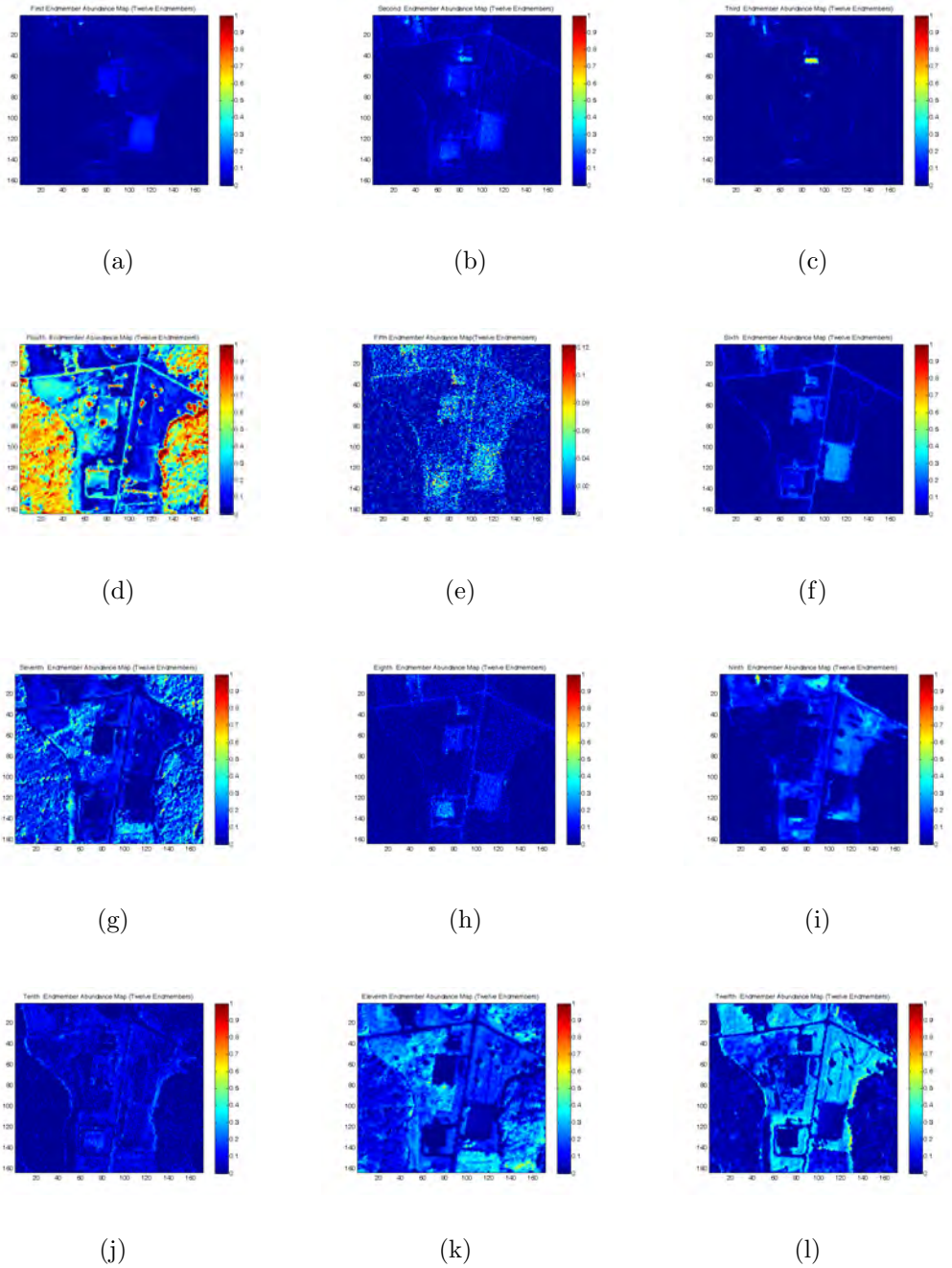
(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

Figure 6–32: A. P. Hill estimated twelve spectral signatures



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

Figure 6–33: A. P. Hill abundance maps for twelve endmembers

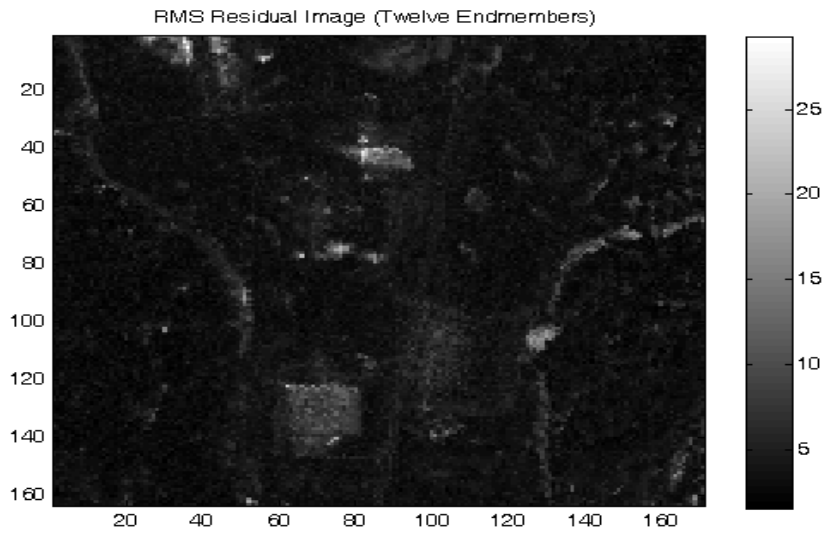


Figure 6–34: Fort A.P. Hill twelve endmembers RMS residual map.

should not show any spatial structure. From the RMS residual images for the three number of endmembers cases, it can be seen that in the three cases, we still have spatial structure, but the RMS value in the four endmembers case is higher. We expect twelve endmembers to be good choice for the number of endmembers.

6.5 Biochemical Data

To show the potential of the developed algorithms in meeting the CenSSIS vision of diverse problems similar solutions, we study the application of the unsupervised unmixing algorithms to microscopy data. We conducted two experiments. The first experiment is to correct for dispersive component estimated using the unmixing developed algorithms. The other experiment is to use unsupervised unmixing developed algorithms to find the cancerous areas in the tissue image. We present and discuss results from both experiments. We also compare these results to those obtained using other algorithms.

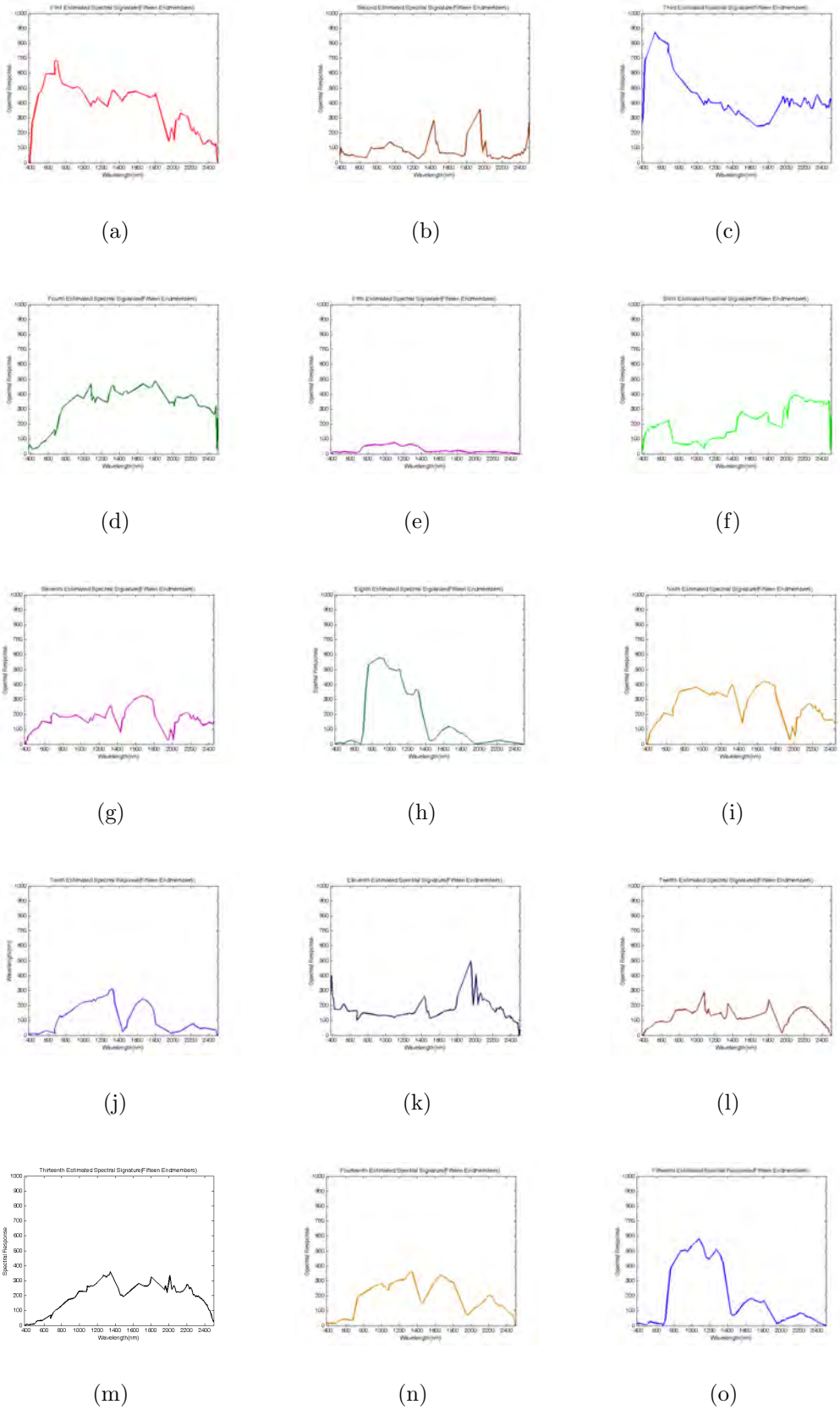
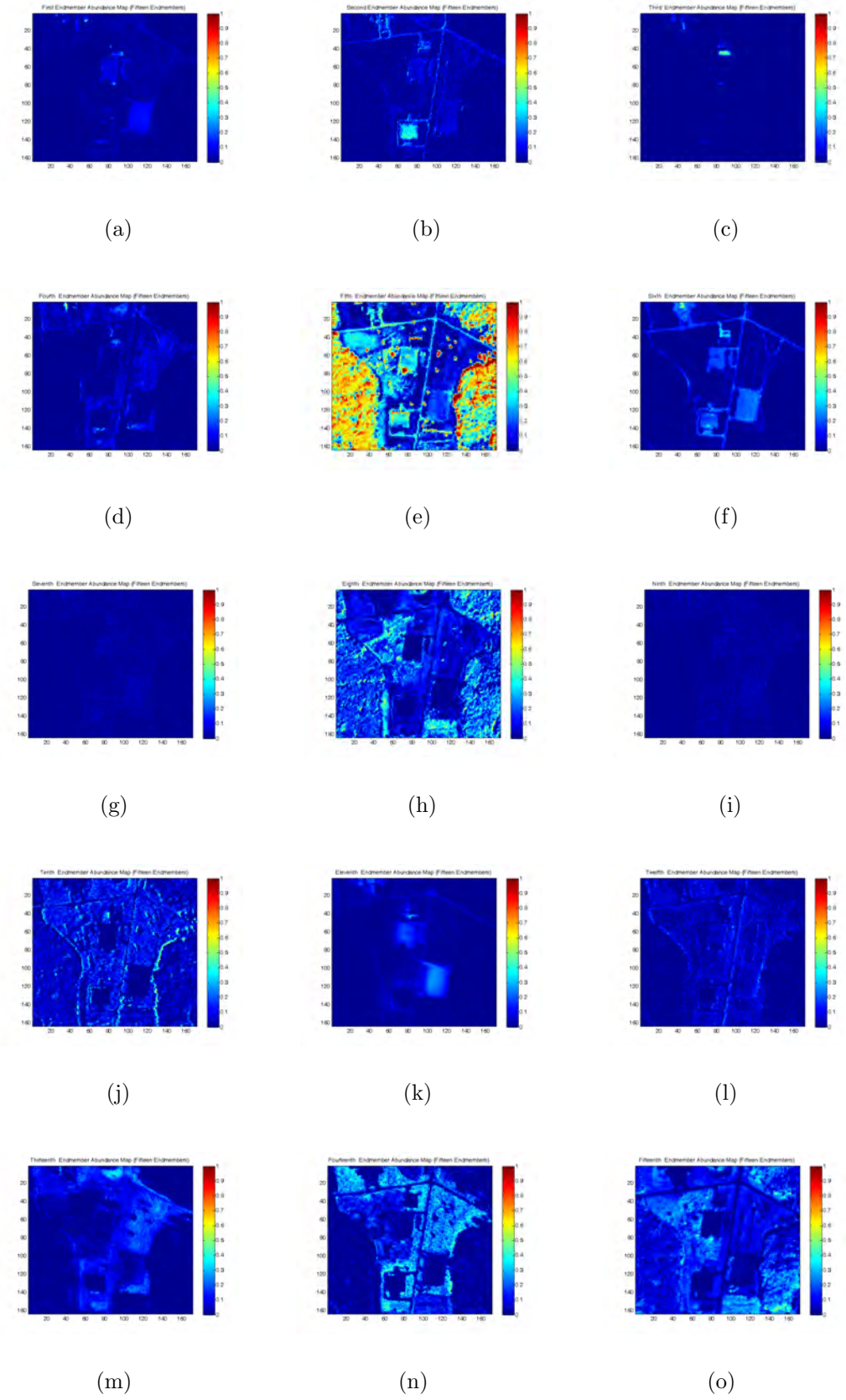


Figure 6–35: A. P. Hill estimated fifteen spectral signatures



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

(m) (n) (o)

Figure 6–36: A. P. Hill abundance maps for fifteen endmembers

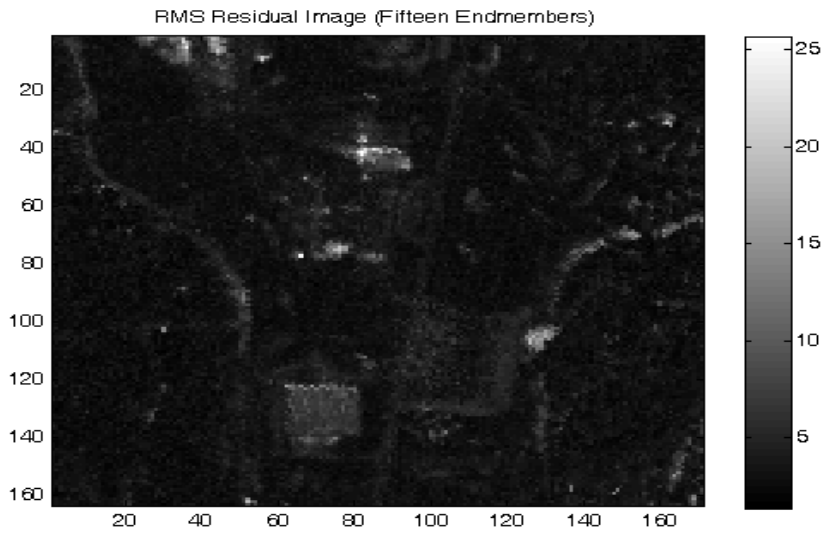


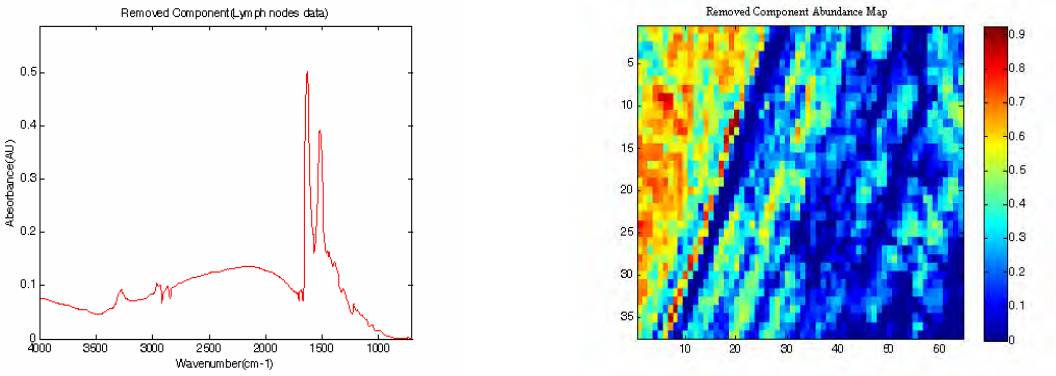
Figure 6–37: Fort A.P. Hill fifteen endmembers RMS residual map.

6.5.1 Dispersion Artifact Correction

Dispersion artifact occurs in infrared microspectral data collected in transflection(reflection/absorption) mode [54, 56]. The artifact occurs along the edges of tissue samples when the tissue does not adhere well to substrate [54, 56]. The presence of the artifact causes an unusual ratio of the amide I/II bands (Peaks ratio), a significant shift (to lower wavenumbers) of the bands. The peak shifts and amide ratio will have a significant influence on the statistical analysis of IR spectra. In many cases, these effects totally dominate the statistical analysis, for example, Hierarchical Cluster Analysis (HCA) as the magnitude of the changes caused by the artifact are more significant than the subtle spectral changes that we are trying to find Cancerous tissue.

The first experiment was conducted using IR lymph nodes. The dispersion artifact component was located in the original image and the response was saved. We use the two-stage approach to process the lymph nodes image. We compare the resulting components with the saved response from the original lymph nodes

image. We selected the closest component to the saved one in the spectral angle sense (Figure 6–38(a)). From the abundance map of the selected component((Figure 6–38(b)), we know the concentration of that component in each pixel of the original image. We subtracted weight of the selected component from the whole pixels of the original image. We reconstructed the lymph nodes with the artifact component removed. Finally, we compared the response of the shown pixel before and after dispersion artifact component removal.

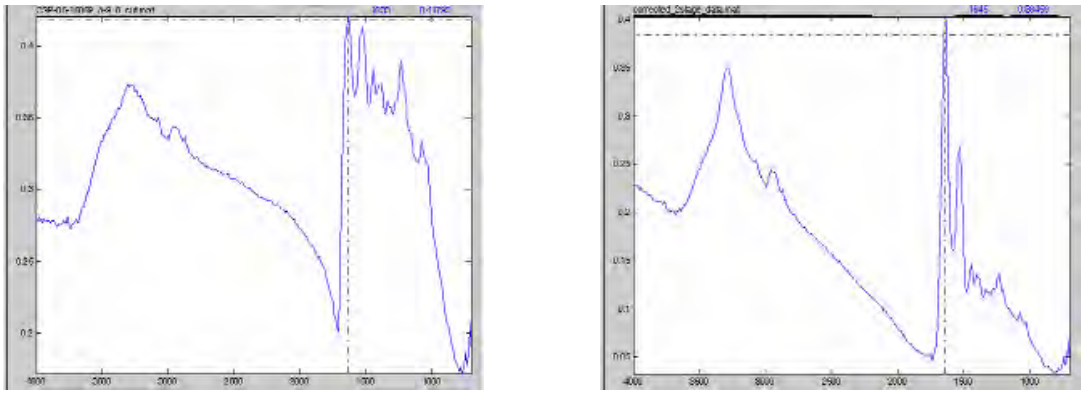


(a) (b)

Figure 6–38: Removed component (a)response (b)abundance map

Figure 6–39(a) shows a pixel response from the IR image before dispersion artifact correction. Figure 6–39(b) shows the same pixel after the dispersion artifact correction. We can see in Figure 6–39 the effect of the dispersion artifact on the response of the pixels. The dispersion artifact cause the ratio between the main two peaks is very small and the location of the peaks is shifted to lower wavenumbers.

Figure 6–40 shows a screen shot taken from the CYTOSPEC software (see [55] for more information about the software). The CYTOSPEC is a toolbox designed to process the Raman and the IR hyperspectral data. The screen shots show the original IR image (top right), the clusters means spectra(left), and the Hierarchical



(a) (b)

Figure 6–39: Pixel response (a)before dispersion correction (b)after dispersion correction

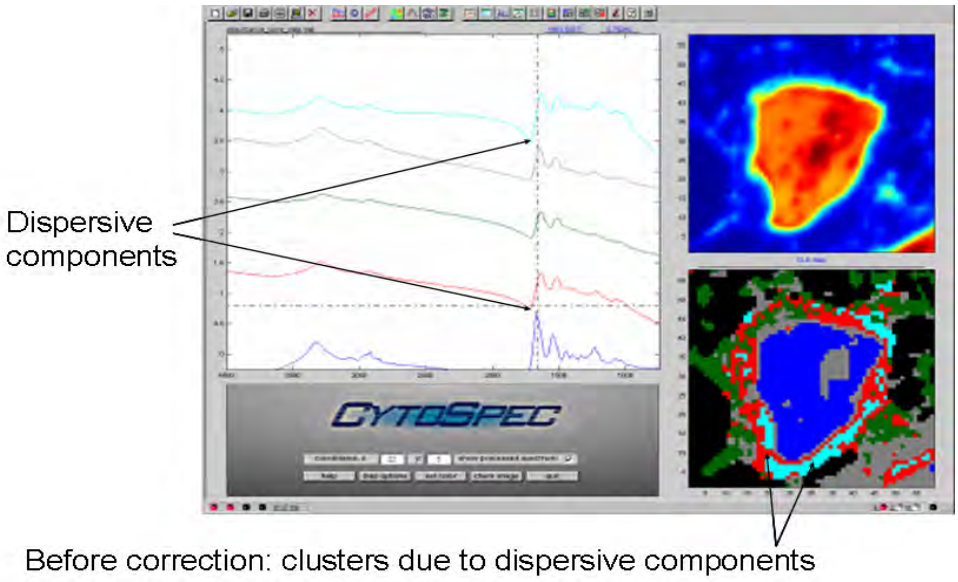


Figure 6–40: Dispersion artifact in IR data.

Cluster Analysis (HCA) clusters map (bottom right). The Figure also shows the location of the dispersive component in the clusters means spectra and the in the clusters map.

Figure 6–41 shows a screen shot obtained after correction of the dispersive component for the same image as in Figure 6–40.

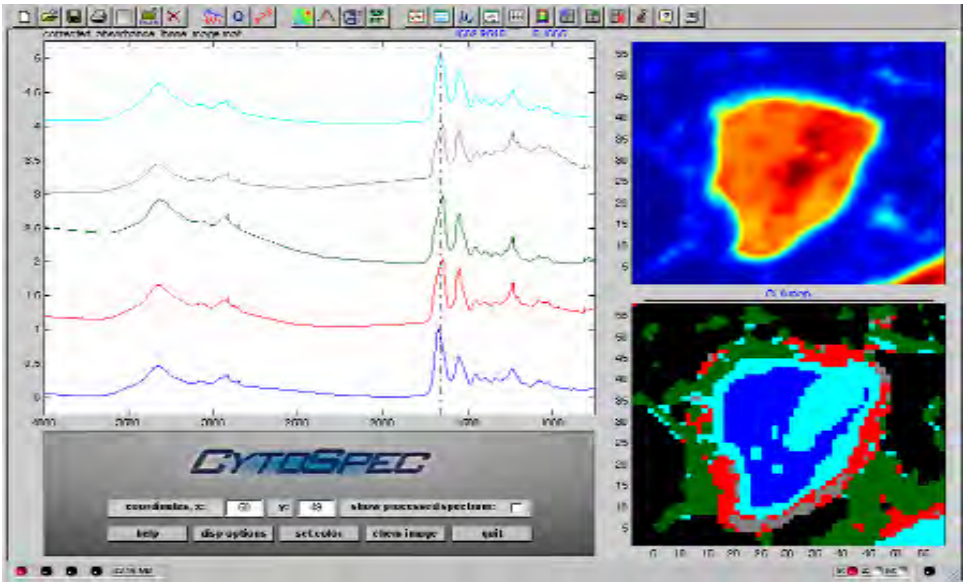


Figure 6–41: Dispersion artifact correction in IR data.

6.5.2 Bladder Tissue Results

In this section, we use a bladder tissue to check for cancerous areas. The bladder tissue sample is shown in Figure 6–42. The Hierarchical Cluster Analysis (HCA) is used to process this image in [56]. Using the HCA, the aim is to find the clusters and their centroids. HCA is time consuming; the times needed to cluster the Bladder tissue is more than 18 hours. We guessed that using unsupervised unmixing we can estimated the endmembers and employ them the same way as the centroids. We tried to minimize the processing time, and explore more potential of our developed unsupervised unmixing algorithms in similar data sets.

Figure 6–44(a) shows the corresponding abundance map of the component shown in Figure 6–43. The marked regions show the high concentration of the component. Figure 6–44(b) shows the cluster map obtained using the hierarchical clustering analysis (HCA). Using the HCA is computationally expensive, it took about 18 hours to get the clustering map. It can be seen from Figure 6–44 that

Figure 6–42: Bladder tissue.

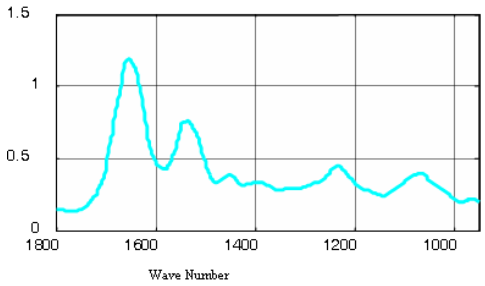
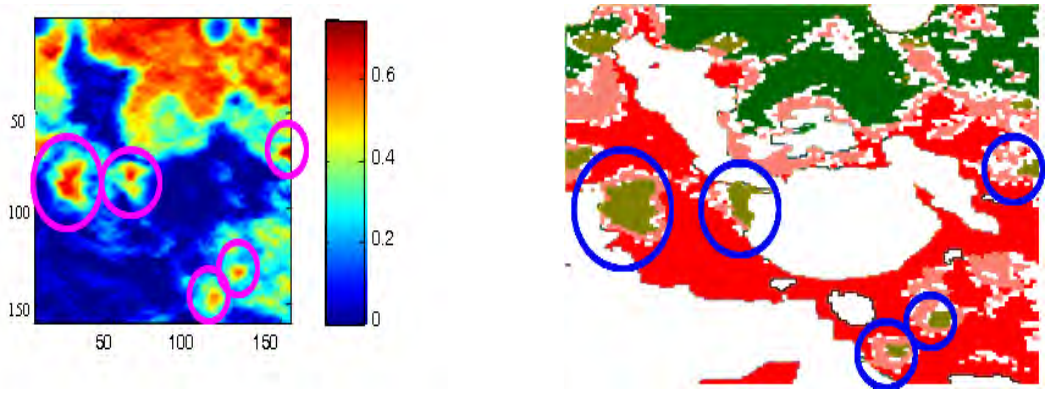


Figure 6–43: Bladder tissue response.

the marked regions obtained from the unmixing algorithm agree to those marked regions obtained using HCA.



(a) (b)

Figure 6–44: Component (a)abundance map (b)HCA clusters map

The unmixing algorithm execution time is 30 minutes where the HCA approach execution time is around 18 hours. This is an advantage of the unmixing algorithms over the HCA.

6.6 Summary

In this Chapter, we present results obtained from processing real data. We use Enrique Reef HYPERION image, Cuprite AVIRIS image, Fort A. P. Hill image, and biochemical data. For all data set we conducted experiments to estimate the spectral response of endmembers, and the corresponding abundance fractions. We use the model fitness metrics to evaluate the obtained results. In the Enrique Reef case, we compare the estimated spectral signatures to PPI and N-FINDR algorithms. We also compare the abundance maps to the testing IKONOS reference image. For the Cuprite image, in addition to evaluating the estimated spectral signatures using the quality of estimation metrics mentioned in Chapter 4, we look at the ground truth of the Cuprite image. The Fort A. P. Hill image is used. We compare the abundance maps to the thematic map of the classified segment. We assume four endmembers in the image. The R-squared plot and the RMS residual map show that the estimated endmembers do not fit well the measured image.

In the biochemical data case, we present two different application of the developed unmixing algorithms. One application is the dispersion artifact correction, while the other is recovering regions which has the potential to be cancerous. In both applications, the developed algorithms perform well and show their potential for such applications.

CHAPTER 7

Ethical Issues

As defined in Meriam-Webster, Ethics is a set of moral principles or values. Sometimes you will hear someone refer to "computing ethics", which is a whole range of ethical questions surrounding computer science and the use of computers. As the field of this research, the focus of this chapter will be on showing the main ethical issues which are related to this type of research. During this research, we touched important borders where ethical issues are strongly required and clearly appear. As a Ph.D. research and when a literature review is done, we need to look at others work in the area of interest. Researcher should be honest enough to avoid repetition of work and especially in evaluating others work. Without being aware of this point one of very important ethical rules could be broken and then the research will be wrongly structured. Other important issue in this research is the honesty in the research results. Accuracy in results analysis is important and should be carried over by a big responsibility and honesty, if not as time passes inaccurate research environment will be growing and this will be reflected negatively on the quality of research. Giving credit to other people for using part of their work is important issue. You as a researcher can use the available information in your research but you can not be in any way the owner of that work.

It is important to add that we should ensure a proper and achievable goals and objectives for this research. As final issue, I would like to discuss in this part of the research is that we are required to make this research beneficial and available for other researchers. Knowledge should be reachable to all scientists in need for this segment of the research environment.

CHAPTER 8

Conclusions and Future Work

8.1 Dissertation Summary

In this section, we summarize the major results from the dissertation. In Chapter 1, we presented the hyperspectral imagery and describe their useful characteristics such as high spectral and spatial resolution. We presented the problem spectral umixing using the linear mixing model. The objectives and the main contribution of this research are described.

Chapter 2 presents background about hyperspectral imaging. The hyperspectral imaging concept, sensors types, and image formation are introduced. The potential of the hyperspectral data in different applications and the work done to develop capable algorithms to process this huge amount of data is introduced. The mixing problem using linear and nonlinear mixing models is presented. Previous work in unmixing is reviewed. Different methods to estimate the linear dimensionality of hyperspectral images were presented. The standard two-stage unmixing procedures was discussed in detail. Standard unmixing is a two-stage approach where the endmembers are determined in the first stage and abundances are estimated in a second stage. The techniques described in the literature for each stage of the standard unmixing are described.

Chapter 3 presents the positive matrix factorization. The geometrical interpretation of the positive matrix factorization and its relation to polyhedral cones is described. We present different approaches used to solve the general positive matrix factorization problem and discuss how to use PMF to solve the unsupervised unmixing problem. Since our approach is based on the constrained PMF, we present the formulation of the unmixing problem as a constrained PMF. We present optimality conditions that any solution to the constrained PMF problem should satisfy. We present the use of Lagrange multiplier to derive the optimality conditions of the problem.

Chapter 4 presents some approaches to compute the constrained PMF. We present a Gauss-Seidel or two-stage approach. We show how the two-stage approach solves the optimization problem given in Equation 2.3 by alternating between two iterative optimization steps until convergence is achieved. In the first stage, we fix the endmember matrix and optimize with respect to the abundance fractions matrix. In the second stage, we fix the abundance fraction matrix and optimize with respect to the endmembers matrix. This problem decomposition was selected since each stage is a constrained linear least squares problem. We also present a penalty approach to compute the constrained PMF. This algorithm focuses on incorporating the constraints into the general PMF problem to get what is called augmented form of the PMF problem. Changing the variable of the augmented form transforms the problem again into the general form. We use the multiplicative rules to update the parameters and in each step we check the satisfaction of the constraints.

Since the developed algorithms are iterative, we study initialization schemes and discuss their impact on the convergence and quality of results.

We discussed the convergence and the stopping criteria of the algorithms. We introduced the solution smoothing factor as a stopping criteria. We presented analysis of estimation quality. We focused on the quality of estimated endmembers and the model fitness. For the estimated endmembers quality, we used the percentage error and the spectral angle. This is possible in the simulated data case. We also presented metrics to measure the fitness of the estimated endmembers to the measured data. We used $\mathbf{R}^2$ value, and root-mean-square residual.

Chapter 5 presents results obtained using simulated data. An experiment to evaluate the initialization schemes and their impact on computing the constrained PMF algorithms was conducted. For this task, we used five endmembers selected from the spectral library of the cuprite image and generated 1000 pixels using the selected endmembers and randomly generated abundances fractions. The generated abundance fractions satisfied the positivity and the sum-to-one constraints. We used the random initialization, random Acol (also called random average) initialization, random C (also called longest norm average) initialization, and the singular value decomposition subset selection initialization. Since in the first three initialization schemes we have randomization effect, the experiment was repeated four times. The average values of the four trials were used as the final results. We estimated the endmembers and compared the estimates with the true spectra. We used the metrics mentioned in Chapter 4, to measure the endmembers estimation quality. We studied the effect of number of spectral bands, number of endmembers, and the noise on the results, we used the RMS residual metric to compare between the results obtained for different number of spectral bands.

To evaluate the developed algorithms, we conducted two additional simulation experiments. One experiment was conducted using a simulated data arranged in a cube with the corners as the endmembers. The other experiment was using the

1000 pixels generated using the five endmembers selected from the Cuprite image spectral library. In this experiment we look at the performance of the developed algorithms to estimated known endmembers. We also used the metrics mentioned in Chapter 4, to look at the quality of estimation and the model fitness using the mentioned metrics.

In Chapter 6, we use real data to evaluate the performance of the algorithms. We use Enrique Reef Hyperion image, Cuprite AVIRIS image, Fort A.P. Hill image, and Biochemical IR images. Herein, we use the R-squared value plot, and the RMS residual plot to look at the quality of results. We present results from the biochemical data to show the potential of the unmixing algorithms in different applications.

8.2 Conclusions

In this research, we developed an unsupervised algorithm to solve the *unmixing problem* in hyperspectral imagery using the constrained positive matrix factorization (cPMF). This algorithm consists of three different steps: *determination of number of endmembers*, *initialization*, and *computing the constrained positive matrix factorization*. The schematic diagram of the algorithm is shown in Figure 8–1. we showed that cPMF could be used to solve the unsupervised *unmixing problem* by extracting simultaneously, the endmembers, and their abundances, by solving the optimization Problem (2.3). We used the two-stage Gauss-Seidel iterative approach to compute the constrained positive matrix factorization.

Our developed algorithm has different advantages over other algorithms as N-FINDR, VCA, and PPI. For example, N-FINDR, VCA, and PPI are geometric approaches, and assume that endmembers are pure pixel from the image, while our algorithm does not have this assumption. This assumption is not always true since pure pixels are rarely found in real hyperspectral images. Our algorithm estimate the

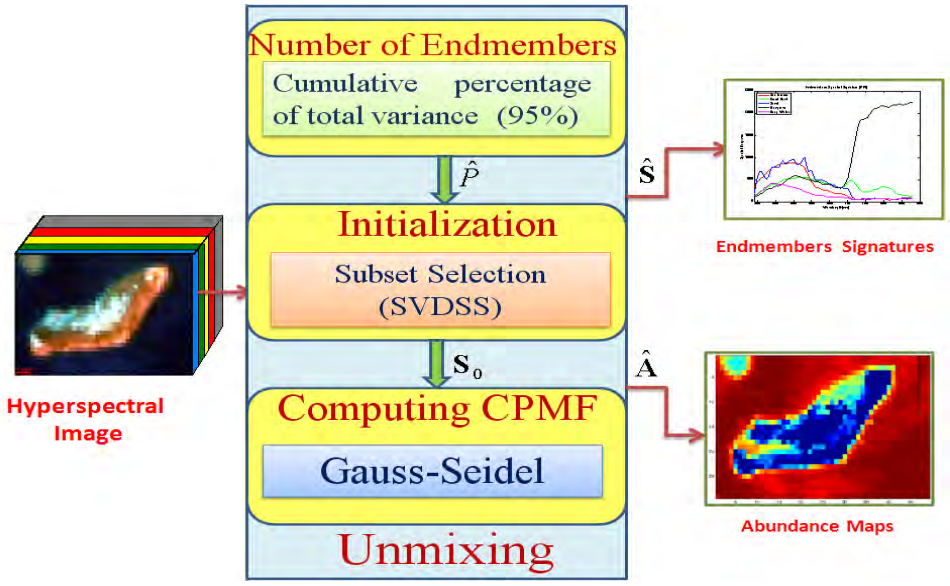


Figure 8–1: Unsupervised unmixing algorithm

endmembers and their abundance simultaneously by solving an optimization problem, while N-FINDR, VCA, and PPI extract the endmembers and their abundances in separate stages, hence being more suitable for automated image analysis.

We evaluated the developed algorithms using simulated data to be able to compare the results to a known reference. Since we use simulated data, we used two metrics to evaluate the performance of the developed algorithms. We used the spectral angle between the estimated and the true spectral signatures and the percentage error of the estimated spectral signature to evaluate the goodness of the estimated endmembers. Both metrics showed complete agreement between the estimated and the true spectral signatures. comparison between initialization schemes are also shown. Results showed that the SVDSS initialization scheme introduced in this work is the best among the used initialization schemes.

In Chapter 6, we presented results using real data. The real data used to test the potential of the developed algorithms to estimate the endmembers from a real hyperspectral image. We used the Enrique Reef HYPERION image to test the

developed algorithms. We choose this image because the obtained results could be compared to the image of the same region acquired using the IKONOS sensor. We use the $\mathbf{R}^2$ value plot and the RMS residual plot to evaluate the obtained results. The obtained results are acceptable with respect to the metrics.

We used the Fort A. P. Hill AVIRIS data. We conducted different experiments using different numbers of endmembers. Our impression was that when we have large number of endmembers, some estimated component are noise. We used twelve endmembers as the best number of endmembers based on the linear dimension estimate. To evaluate the effect of number of endmembers, we compared results for four, twelve, and fifteen endmembers. We compared the obtained results with a thematic map of the same image. The thematic map was obtained using the spectral library developed for the Fort A. P. Hill image.

We also applied our algorithms to hyperspectral IR microscopy imagery. Results show that unsupervised unmixing can be used to retrieve and correct for instrument artifacts and in cancer detection. The first application was to estimate and remove a dispersive artifact component from the IR image. We showed the effects that removing the artifact had in improving the clustering results. We also tried to estimate the component with high concentration in regions of the Bladder tissue image that has to potential to be cancerous. Our results agreed to those results obtained using the HCA. Our algorithms had an advantage over the HCA in that the processing time is 30 minutes where the processing time for the HCA is more than 18 hours. These results motivate the use of the developed algorithm in different biomedical applications.

In summary, the developed algorithms solve the unsupervised unmixing problem well and have the potential to process different type of data. This clearly demonstrates CenSSIS vision of diverse problems similar solutions.

8.3 Future Work

Future work is possible in different directions. In the developed algorithms, some issues are not solved yet. Reliable determination of the number of endmembers is still a big unsolved issue. Although linear dimensionality was used here as a first approach, this approach could not estimated correctly the number of endmembers in the cases shown in Figure 2–11. We still think that positive rank is the appropriate rank to look at but there are no methodologies to estimate it.

Because of the MATLAB memory management problem, when the number of pixels and the number of bands are large, the step of computing the cPMF of our developed algorithm face an out of memory problem. We need to address how to remove these limitations. This issue can be addressed using the different memory management approaches. ENVI/IDL has better memory management capabilities and should be considered for future implementations.

Also, Because of the bilinear structure that the unmixing problem has, it is possible to develop parallel algorithms instead of the developed sequential algorithms. The parallel algorithms could solve the image size limitation problem. In [68], an algorithm based on PMF and alternating quadratic programming (AQP)was presented. The unmixing problem was written in subproblems which motivates the parallelization of the unmixing problem. In [67], a framework for parallel processing of hyperspectral imagery is presented. This framework could be followed to achieve the parallelization of our developed algorithms. Parallelizing our algorithm is a complex process. Although the abundance estimation problem can be solved for each

pixel independently the endmember determination requires to look at the full image hence, complicating problem decomposition. We expect the parallel implementation to have direct effect on the results, and the convergence speed, so we need to partition the image in such a way that does not affect the results. I suggest to explore the partition along the spectral dimension. that way we make sure that each process is working on the same spatial area. We need to study the effect of noisy bands on the results using this partition along the spectral dimension.

We presented some results using the biochemical data. The results show the potential of the developed algorithms for different applications. One direction of the future work could be to study the capability of the developed unmixing algorithms to process biochemical data and extract information from this type of data.

In this research, we assumed no variability on the endmembers, this assumption is not always valid. We suggest to investigate how to further refine existing linear mixing models used to model spectral mixing in a measured spectral signature to account for variability of the endmember and how this modeling assumption affects the solution of the unmixing problem using supervised and unsupervised approaches.

REFERENCES

[1] N. Keshava and J.F. Mustrad. "A Survey of Spectral Unmixing." In *Lincoln Laboratory Journal*, Vol. 14, No.1, pp. 55-78, 2003.

[2] Chang C. I.,Ren H. et. al. "Subpixel Target Size Estimation for Remotely Sensed Imagery." In *Proceedings of SPIE: Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery IX*, Vol. 5093, April 2003.

[3] C. L. Lawson and R. J. Hanson. *Solving Least Squares Problems*. Prentice-Hall. 1974.

[4] M. Velez-Reyes, S. Rosario, A. Puetz and R. B. Lockwood. "Iterative Algorithms for Unmixing of Hyperspectral Imagery." In *Proceedings of SPIE: Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery IX*, Vol. 5093, April 2003.

[5] M. E. Winter. "Fast Autonomous Spectral Endmember Determination in Hyperspectral Data." In *Proceedings of the Thirteenth International Conference on Applied Geologic Remote Sensing*, Vol. II, pp. 337-344, Canada 1999.

[6] J. Nascimento and J. Dias. "Vertex component analysis: a fast algorithm to extract endmembers spectra from hyperspectral data." In *Proceedings of the First IbPRIA*, vol. 2652, pp. 626-635, Spain, 2003.

[7] N. Keshava and J. F. Mustrad."Spectral unmixing." In *IEEE Signal Processing Magazine*, pp. 44-57, 2002.

[8] L. Parra, C. Spence, P. Sajda, A. Ziehe, and K. R. Mller. "Unmixing Hyperspectral Data." In *Advances in Neural Information Processing*, Vol. 12 (Proc. NIPS*99), pp. 942- 948, MIT Press, 2000.

[9] D. D. Lee and H. S. Seung. "Learning the Parts of Objects by Non-negative Matrix Factorization." In *Nature* Vol. 401, pp. 788-791, 1999.

[10] D. D. Lee and H. Sebastian Seung. "Algorithms for Non-negative Matrix Factorization." In *Advances in Neural Information Processing*, pp. 556-562, 2001.

[11] D. D. Lee and H. S.Seung. "Unsupervised Learning by Convex and Conic Coding." In *Proceedings of the Conference on Neural Information Processing Systems*, Vol. 9, pp. 515-521, 1997.

[12] D. Donoho and V. Stodden. "When does Non-Negative Matrix Factorization give a correct Decomposition into Parts?." In *NIPS Workshop on ICA: Sparse Representations in Signal Processing*, Whistler, Canada,2003.

[13] P. Sajda, S. Du and L. Parra. "Nonnegative Matrix Factorization for Rapid Recovery of Constituent Spectra in Magnetic Resonance Chemical Shift Imaging of the Brain." In *IEEE Transactions on Medical Imaging*, Vol. 23, No. 12, December 2004.

[14] Y. H. Hu, H. B. Lee and F. L. Scarpace. "Optimal Linear Spectral Unmixing." In *IEEE Transactions on Geoscience, and Remote Sensing*, Vol. 37, No. 12, January 1999.

[15] M. Velez-Reyes and S. Rosario. "Solving Abundance Estimation in Hyperspectral Unmixing as a Least Distance Problem." In *Proceedings of the IEEE International Geosciences and Remote Sensing Symposium*, Alaska, 2004.

[16] M. Chu, F. Diele, R. Plemmons, and S. Ragni. "Optimality,Computation, and Interpretations of Nonnegative Matrix Factorizations." In *SIAM Journal on Matrix Analysis*, October 2004.

[17] K. Kontantinides and K. Yao. "Statistical Analysis of effective Singular Values in Matrix Rank Determination." In *IEEE Transaction on Acoustics, Speech, and Signal Processing*. Vol. 36, No. 5, May 1988.

[18] G. Solyar, C. Chang, and A. Plaza. "Endmember generation by projection pursuit." In *Proceedings of SPIE: Algorithms and Technology for Multispectral, Hyperspectral, and Ultraspectral Imagery XI*, vol. 5806 , April 2005.

[19] J. W. Boardman, F. A. Kruse, and R. O. Green. "Mapping Target Signatures via Partial Unmixing of AVIRIS Data." In *Proceedings of JPL Airborne Earth Science Workshop, Pasadena*, CA, 1995. http://aviris.jpl.nasa.gov.

[20] A. Plaza and C. Chang. "An Improved N-FINDR Algorithm in Implementation. In *Proceedings of SPIE: Algorithms and Technology for Multispectral, Hyperspectral, and Ultraspectral Imagery XI*, Vol. 5806 , April 2005.

[21] *Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)* http://aviris.jpl.nasa.gov.

[22] P. Lindstrom and P. A. Wedin. "Gauss-Newton Based Algorithms for onstrained Nonlinear Least Squares Problems." http://citeseer.ist.psu.edu/237590.html.

[23] J. M. van den Hof and J. H. Van Schuppen. " Positive Matrix Factorization via Extremal Polyhedral Cones". In *Linear Algebra and its Application*, Vol. 293, pp. 171-186, 1999.

[24] R. Basedow, P. Silverglate, et. al. "The HYDICE Instrument Design." *In Proceedings of the International Symposium on Spectral Sensing Research*, Vol. 1, pp. 430-445, 1992.

[25] NASA. "Hyperion Satellite Sensor." http://edc.usgs.gov/products/satellite/eol1.html.

[26] ENVI, "the Environment for Visualizing Images User Guide." 3.4 ed. Research Systems Inc., 2000. http://www.rsinc.com/envi/index.cfm.

[27] M. E. Winter. "Comparison of Approaches for Determining End-Members in Hyperspectral Data." In *Proceedings of the Aerospace Conference*, Vol. 3 , pp. 305-313, 2000.

[28] M.D.Craig. "Minimum Volume Transforms for Remotely Sensed Data." In *IEEE Transactions on Geoscience and Remote Sensing*, Vol.32 , pp. 542-552, 1994.

[29] J. M. van den Hof and J. H. van Schuppen. "Positive Matrix Factorization via Extermal Polyhedral Cones." *Linear Algebra and its Applications*, Vol.293, pp. 171-186, 1999.

[30] P. Paatero and U. Tapper. "Positive Matrix Factorization: A Non-negative Factor Model with Optimal Utilization of Error Estimates of data values." *Environmetrics*, Vol. 5, pp. 111-126, 1994.

[31] A. N. Langville, M. W. Berry, M. Browne, V. P. Pauca, and R. J. Plemmons, "A Survey of Algorithms and Applications for the Nonnegative Matrix Factorization." In *Computational Statistics and Data Analysis*. Elsevier. Submitted Jan. 2006.

[32] A. N. Langville, M. W. Berry, M. Browne, V. P. Pauca, and R. J. Plemmons. "Algorithms and Applications for the approximate Nonnegative Matrix Factorization." In *Computational Statistics and Data Analysis*. Elsevier. 2007

[33] P. Hoyer. "Non-negative Matrix Factorization with Sparseness Constraints." In *Journal of Machine Learning Research*, Vol.5, pp. 1457-1469, 2004.

[34] F. Shahnaz, M. Berry, P. Pauca, and R. Plemmons. "Document Clustering using Nonnegative Matrix Factorization." In *Information Processing and Managenent*, Vol. 42, No. 2, pp. 373-386, 2006.

[35] P. Pauca, J. Piper, and R. Plommens. "Nonnegative Matrix Factorization for Spectral Data Analysis." to appear In *Linear Algebra and its Application*, 2005.

[36] S. Marinetti, L. Finesso, E. Marsilio. " Matrix Factorization Methods: Application to Thermal NDT/E." In *Proceedings of the Fifth Workshop, Advances in Signal Processing for Non Destructive Evaluation Materials*, Canada, 2006.

[37] J. A. O'Sullivan. "Properties of the Information Value Decomposition." In *Proceedings of IEEE International Symposioums on Information Theory*, page 491, Italy, 2000.

[38] E. K. P. Chong, and S. H. Zak. *An Introduction to Optimization*. John Wiley and Sons, 2001.

[39] A. N. Langville, C. D. Meyer, R. Albright, J. Cox, and D. Duling. "Initializations for the Nonnegative Matrix Factorization." In *Proceedings of the Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, August 23-26, 2006.

[40] S. Wild. *Seeding non-negative Matrix Factorizations with Spherical K-means Clustering*. Masters thesis, University of Colorado, 2003.

[41] S. Wild, J. Curry, and A. Dougherty. "Motivating Non-negative Matrix Factorizations." In *Eighth SIAM Conference on Applied Linear Algebra*, Philadelphia, 2003.

[42] G. H. Golub and C. F. Van Loan. *Matrix Computations*. 3rd ed. , Johns Jopkins University Press, MD, 1996.

[43] M. Velez-Reyes, L. O. Jimenez, et. al. "Comparison of Matrix Factorization Algorithmes for Band Selection in Hyperspectral Imagery." In *Proceedings of SPIE: Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery VI*, Vol. 4049, April 2000.

[44] S. Wild, J. Curry, and A. Dougherty. "Improving Non-negative Matrix Factorizations Through Structured Initialization." In *Journal of Pattern Recognition*, Vol. 37, No. 11, pp. 2217-2232, 2004.

[45] A. Hyvarinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley and Sons, 2001.

[46] L. O. Jimenez, and D. A. Landgrebe. "Hyperspectral Data Analysis and Supervised Feature Reduction via Projection Pursuit." In *IEEE Trans on Geoscience*

and Remote Sensing, vol. 37. , no. 6, pp. 2653-2667, Nov. 1999.

[47] A. Ifarraguerri and C. Chang. "Hyperspectral image segmentation with convex cones" In *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no 2, pp. 756-770, March 1999.

[48] H. C. Schau. "Acceleration of Tomographic Hyperspectral Restoration Algorithms." In *Proceedings of SPIE: Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIII*, Vol. 6233, May 2006.

[49] J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley and Sons. 1999.

[50] J. B. Adams and A. R. Gillespie. *Remote Sensing of Landscapes with Spectral Images : A physical Modeling Approach*. Cambridge University Press. 2006.

[51] S. M. Jong and F. D. van de Meer. *Remote Sensing Image Analysis Including the Spatial Domain*. Springer . 2006.

[52] K. B. Petersen and M. S. Pedersen. The Matrix Cookbook. http://www.iro.umontreal.ca/~kegl/ift3390/2006_1/Lectures/l01_MatrixCookbook.pdf

[53] J. J. Cipar, R. Lockwood, T. Cooley and P. Grigsby. "Background Spectral library for Fort A.P. Hill, Virginia." In *Proceedings of SPIE: Remote Sensing and Modeling of Ecosystems for Sustainability*. Vol.5544, No.1, pp. 35-46, 2004.

[54] M. Romeo, and M. Diem. "Correction of Dispersive Line Shape Artifact Observed in Diffuse Reflection Infrared Spectroscopy and Absorption/Reflection(Transflection) Infrared Micro-Spectroscopy." In *Vibrational Spectroscopy*,vol. 38, pp. 129-132, 2005.

[55] P. Lasch. *A Matlab Based Application for Infrared Imaging*. http://www.cytospec.com

[56] M. Romeo, and M. Diem. "Infrared Spectral Imaging of Lymph Nodes: Strategies for Analysis and Artifact Reduction." In *Vibrational Spectroscopy*, vol. 38,

pp. 115-119, 2005.

[57] J. S. Tyo. " Image Acquisition, Storage, and Interpretation." http://www.eece.unm.edu/faculty/tyo/hsi_gtri/aqcuisition_gtri.

[58] D. G. Luenberger. "Linear and Nonlinear Programming." Addison-Wesley Publishing CO. Second Edition. 1984.

[59] C.-J. Lin. "On the Convergence of Multiplicative Update Algorithms for Non-negative Matrix Factorization." To appear in *IEEE Transactions on Neural Networks*, 2007.

[60] C.-J. Lin. "Projected gradient methods for non-negative matrix factorization." To appear in *Neural Computation*, 2007.

[61] A. Sen, and M. Srivastava. *Regression Analysis: Theory, Methods, and Applications*. Springer-Verlag. 1990.

[62] U.S. Geological Survey (USGS). http://speclab.cr.usgs.gov/PAPERS/tetracorder/.

[63] J. Nascimento. *Unsupervised Hyperspectral Unmixing*. PhD Dissertation, *Universidade Technica de Lisboa*. Spain. 2006.

[64] C. Kwan, B. Ayhan, G. Chen, J. Wang, B. Ji, and C. Chang. " A Novel Approach for Spectral Unmixing, Classification, and Concentration Estimation of Chemical and Biological Agents." In *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 44, No. 2, February 2006.

[65] A. Umaña-Díaz. *Determining the Dimensionality of Hyperspectral Imagery.* Masters thesis, University of Puerto Rico, 2003.

[66] K. Thome, P. Slater, S. Biggar, B. Markham, Barker. "Radiometric Calibration of Landsat." In *Photogrammetric Engineering Remote Sensing*. Vol. 63, no. 7, pp. 853-858. July 1997.

[67] S. A. Robila and L. G. Maciak .“A Parallel Unmixing Algorithm for Hyperspectral Images.” In *Proceedings of SPIE: Intelligent Robots and Computer Vision XXIV: Algorithms, Techniques, and Active Vision*, Vol. 6384, October 2006.

[68] M. Parente, A. Zymnis, J. Skaf, and J. Bishop.“ASpectral unmixing with nonnegative matrix factorization.” In *Proceedings of SPIE: Remote Sensing for Environmental Monitoring, GIS Applications, and Geology VI*, Vol. 6366, 2006.

[69] National Geospatial Intelligence Agency (NGA). “Hyperspectral Imagery (HSI) Dimension Reduction.” http://www.ipam.ucla.edu/publications/gss2005/gss2005_5579.ppt.

[70] M. Velez-Reyes. *Decomposition Algorithms for Parameter estimation.* Doctoral Dissertation, Massachusets Institute of Technology, 1992.

APPENDIX A

Matrix Operators and Differentiation

A.1 Vector and Matrix Differentiation

Let f be a scalar valued vector function and let $\beta \in \Re^n$,

$$f : \Re^n \to \Re$$

The Jacobian of f is given by

$$\frac{\partial f(\beta)}{\partial \beta} = (\frac{\partial f(\beta)}{\partial \beta_1}, ..., \frac{\partial f(\beta)}{\partial \beta_n}) \tag{A.1}$$

The gradient of f is given by

$$\nabla f(\beta) = \left(\frac{\partial f(\beta)}{\partial \beta}\right)^T \tag{A.2}$$

Similarly, if

$$f : \Re^n \to \Re$$

$$[\frac{\partial f(\mathbf{A})}{\partial \mathbf{A}}]_{ij} = \frac{\partial f(\mathbf{A})}{\partial a_{ij}} \tag{A.3}$$

We write $\Re^{m\times n}$ for the space of all real-valued matrices. Then

- For $\mathbf{A} \in \Re^{m\times n}$ and $\beta \in \Re^{\mathrm{n}}$

$$\frac{\partial(\mathbf{A}\beta)}{\partial \beta} = \mathbf{A} \tag{A.4}$$

- For $\mathbf{A} \in \Re^{m\times n}$ and $\beta \in \Re^{\mathrm{n}}$

$$\frac{\partial \beta^T \mathbf{A}\beta}{\partial \beta} = \beta^T(\mathbf{A}^T + \mathbf{A}) \tag{A.5}$$

- For $\mathbf{A} \in \Re^{m\times n}$, $\beta \in \Re^{\mathrm{n}}$ and $\alpha \in \Re^{\mathrm{m}}$

$$\frac{\partial \alpha^T \mathbf{A}^T \beta}{\partial \mathbf{A}} = \beta\alpha^T \tag{A.6}$$

$$\frac{\partial \alpha^T \mathbf{A}\beta}{\partial \mathbf{A}} = \alpha\beta^T \tag{A.7}$$

$$\frac{\partial \alpha^T \mathbf{A}\alpha}{\partial \mathbf{A}} = \frac{\partial \alpha^T \mathbf{A}^T \alpha}{\partial \mathbf{A}} = \alpha\alpha^T \tag{A.8}$$

$$\frac{\partial \mathbf{A}}{\partial a_{ij}} = \mathbf{J}^{ij} \tag{A.9}$$

where

$$\mathbf{J}^{ij}$$

is the single nonzero-entry matrix, with 1 at the entry (i,j) and zero elsewhere.

$$\frac{\partial \alpha^T \mathbf{A}^T \mathbf{A}\beta}{\partial \mathbf{A}} = \mathbf{A}(\alpha\beta^T + \beta\alpha^T) \tag{A.10}$$

$$\frac{\partial \mathbf{A}^T\mathbf{B}\mathbf{A}}{\partial a_{ij}} = \mathbf{A}^T\mathbf{B}\mathbf{J}^{ij} + \mathbf{J}^{ij}\mathbf{B}\mathbf{A}$$
$$(\mathbf{J}^{\mathrm{ij}})_{kl} = \delta_{ik}\delta_{jl}, \delta_{\mathrm{ik}} = \begin{cases} 1 & \text{if} \quad \mathrm{i} = \mathrm{k} \\ 0 & \text{otherwise} \end{cases} \tag{A.11}$$

For a full list of matrix identities and facts see [52].

A.2 The Kronecker Product

The Kronecker product is a binary matrix operator that maps two arbitrarily dimension matrices into a large matrix with special block structure.

Definition A.2.1. *(Kronecker Product): Given the matrix* $\mathbf{A}^{m\times n}$ *and the matrix* $\mathbf{B}^{p\times q}$

$$\mathbf{A} = \begin{bmatrix} a_{11} & . & . & . & a_{1m} \\ . & . & & & . \\ . & & . & & . \\ . & & & . & . \\ a_{n1} & . & . & . & a_{nm} \end{bmatrix} \mathbf{B} = \begin{bmatrix} b_{11} & . & . & . & b_{1q} \\ . & . & & & . \\ . & & . & & . \\ . & & & . & . \\ b_{p1} & . & . & . & b_{pq} \end{bmatrix} \tag{A.12}$$

their Kronecker product, denoted $\mathbf{A}\otimes\mathbf{B}$*, is the* $np\times mq$ *matrix with the block structure*

$$\mathbf{A}\otimes\mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & . & . & . & a_{1m}\mathbf{B} \\ . & . & & & . \\ . & & . & & . \\ . & & & . & . \\ a_{n1}\mathbf{B} & . & . & . & a_{nm}\mathbf{B} \end{bmatrix} \tag{A.13}$$

A.2.2 Kronecker Product Operator Properties

Assume that **A**, **B**, **C**, and **D** are real valued matrices. The following identities hold for appropriately dimensioned matrices

1. The Kronecker product is a bi-linear operator. Given $\alpha \in R$

$$\begin{aligned} \mathbf{A} \otimes (\alpha \mathbf{B}) &= \alpha(\mathbf{A} \otimes \mathbf{B}) \\ (\alpha \mathbf{A}) \otimes \mathbf{B} &= \alpha(\mathbf{A} \otimes \mathbf{B}) \end{aligned} \tag{A.14}$$

2. Kronecker product distributes over addition

$$\begin{aligned} (\mathbf{A} + \mathbf{B}) \otimes \mathbf{C} &= (\mathbf{A} \otimes \mathbf{C}) + (\mathbf{B} \otimes \mathbf{C}) \\ \mathbf{A} \otimes (\mathbf{B} + \mathbf{C}) &= (\mathbf{A} \otimes \mathbf{B}) + (\mathbf{A} \otimes \mathbf{C}) \end{aligned} \tag{A.15}$$

3. The Kronecker product is associative

$$(\mathbf{A}) \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C}) \tag{A.16}$$

4. The Kronecker product is not in general commutative, i.e. usually

$$\mathbf{A} \otimes \mathbf{B} \neq \mathbf{B} \otimes \mathbf{A} \tag{A.17}$$

5. Transpose distributes over the Kronecker (does not invert order)

$$(\mathbf{A} \otimes \mathbf{B})^T = \mathbf{A}^T \otimes \mathbf{B}^T \tag{A.18}$$

6. Matrix multiplication, when dimension are appropriate,

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC} \otimes \mathbf{BD}) \tag{A.19}$$

7. When **A** and **B** are square and full rank

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = (\mathbf{A}^{-1} \otimes \mathbf{B}^{-1}) \tag{A.20}$$

8. When $\mathbf{A}^{n\times n}$ and $\mathbf{B}^{m\times n}$, the determinant of a Kronecker product is given by

$$\det(\mathbf{A} \otimes \mathbf{B}) = (\det(\mathbf{A}))^m.(\det(\mathbf{B}))^n \tag{A.21}$$

9. The trace of a Kronecker product is

$$trace(\mathbf{A} \otimes \mathbf{B}) = trace(\mathbf{A}).trace(\mathbf{B}) \tag{A.22}$$

10. The vec of a matrix multiplication, when dimension are appropriate for the product **ABC** to be well defined, is given by

$$vec(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})vec(\mathbf{B}) \tag{A.23}$$

A.3 The Vec Operator

There are situations in which it is useful to transform a matrix to a vector that has as its elements the elements of the matrix. One such situation is when we have a complicated matrix system of equations and a solution is needed. Usually, it is easier to work with vector because most of algorithms are designed for vectors. Least squares problem is well known example; all the algorithms are based on vector form equations. Consequently, when we have a matrix form least squares problem, it is usually more convenient mathematically to transform our problem into vector form problem. In Chapter 4, we use this technique to simplify our problem and to utilize developed solution to similar vector form problems.

The operator that transforms a matrix to a vector is known as the vec operator. If the $m \times n$ matrix $\mathbf{A}$ has $\mathbf{a}_i$ as its ith column, then vec($\mathbf{A}$) is the $mn \times 1$ vector given by

$$vec(\mathbf{A}) = \begin{bmatrix} \mathbf{a}_1 \\ . \\ . \\ . \\ \mathbf{a}_n \end{bmatrix} \tag{A.24}$$

Example. *If* $\mathbf{A}$ *is the* 3×3 *given by*

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

then vec($\mathbf{A}$*) is the* 9×1 *vector given by*

$$vec(\mathbf{A}) = \begin{bmatrix} 1 \\ 4 \\ 7 \\ 2 \\ 5 \\ 8 \\ 3 \\ 6 \\ 9 \end{bmatrix}$$

The vec operator has the following properties [49]:

1. *If* $\mathbf{a} \in R^{n\times 1}$*, is a vector , then* $vec(\mathbf{a}) = \mathbf{a}$
2. *If* $\mathbf{A}^{m\times n}$*, and* $\mathbf{v}^{n\times 1}$ *then*

$$vec(\mathbf{Av}) = \mathbf{Av}$$

3. *The vec operator is linear.*

$$vec(\mathbf{A} + \mathbf{B}) = vec(\mathbf{A}) + vec(\mathbf{B}) \tag{A.25}$$

4. *Multiplication of three matrices*

$$vec(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})vec(\mathbf{B}) \tag{A.26}$$

5. *Multiplication of two matrices*

$$vec(\mathbf{AB}) = (\mathbf{I}_n \otimes \mathbf{A})vec(\mathbf{B}) = (\mathbf{B}^T \otimes \mathbf{I}_m)vec(\mathbf{A}) \tag{A.27}$$

6. *The trace of two matrices product*

$$(\mathbf{vec}(\mathbf{B}^T)^T\mathbf{vec}(\mathbf{A}) = \mathbf{tr}(\mathbf{AB}) = \mathbf{tr}(\mathbf{BA}) = (\mathbf{vec}(\mathbf{A}^T))^T\mathbf{vec}(\mathbf{B}) \tag{A.28}$$

7. *The trace of three matrices product*

$$tr(\mathbf{ABC}) = (vec(\mathbf{A}^T))^T(\mathbf{C}^T \otimes \mathbf{I})vec(\mathbf{B}) \tag{A.29}$$