

REPORT DOCUMENTATION PAGE			Form Approved OMB NO. 0704-0188		
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 17-06-2015		2. REPORT TYPE Ph.D. Dissertation		3. DATES COVERED (From - To) -	
4. TITLE AND SUBTITLE Context Dependent Spectral Unmixing				5a. CONTRACT NUMBER W911NF-13-1-0066	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHORS Hamdi Jenzri (PhD student), Hichem Frigui (Advisor)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAMES AND ADDRESSES University of Louisville 2301 S. Third Street Jouett Hall Louisville, KY 40208 -1838				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211				10. SPONSOR/MONITOR'S ACRONYM(S) ARO	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S) 63184-CS.9	
12. DISTRIBUTION AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.					
14. ABSTRACT A hyperspectral unmixing algorithm that finds multiple sets of endmembers is proposed. The algorithm, called Context Dependent Spectral Unmixing (CDSU), is a local approach that adapts the unmixing to different regions of the spectral space. It is based on a novel function that combines context identification and unmixing. This joint objective function models contexts as compact clusters and uses the linear mixing model as the basis for unmixing. ~~Several variations of the CDSU, that provide additional desirable features, are also proposed.~~					
15. SUBJECT TERMS Context-Dependent Clustering; Subpixel target detection; hyperspectral imaging					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT UU	b. ABSTRACT UU	c. THIS PAGE UU			Hichem Frigui
					19b. TELEPHONE NUMBER 502-852-2009

Report Title

Context Dependent Spectral Unmixing

ABSTRACT

A hyperspectral unmixing algorithm that finds multiple sets of endmembers is proposed.

The algorithm, called Context Dependent Spectral Unmixing (CDSU), is a local approach that adapts the unmixing to different regions of the spectral space. It is based on a novel function that combines context identification and unmixing. This joint objective function models contexts as compact clusters and uses the linear mixing model as the basis for unmixing.

Several variations of the CDSU, that provide additional desirable features, are also proposed.

First, the Context Dependent Spectral unmixing using the Mahalanobis Distance (CDSUM) offers the advantage of identifying non-spherical clusters in the high dimensional spectral space. Second, the Cluster and Proportion Constrained Multi-Model Unmixing (CC-MMU and PC-MMU) algorithms use partial supervision information, in the form of cluster or proportion constraints, to guide the search process and narrow the space of possible solutions. The supervision information could be provided by an expert, generated by analyzing the consensus of multiple unmixing algorithms, or extracted from co-located data from a different sensor. Third, the Robust Context Dependent Spectral Unmixing (RCDSU) introduces possibilistic memberships into the objective function to reduce the effect of noise and outliers in the data. Finally, the Unsupervised Robust Context Dependent Spectral Unmixing (U-RCDSU) algorithm learns the optimal number of contexts in an unsupervised way. The performance of each algorithm is evaluated using synthetic and real data. We show that the proposed methods can identify meaningful and coherent contexts, and appropriate endmembers within each context.

The second main contribution of this thesis is consensus unmixing. This approach exploits the diversity and similarity of the large number of existing unmixing algorithms to identify an accurate and consistent set of endmembers in the data. We run multiple unmixing algorithms using different parameters, and combine the resulting unmixing ensemble using consensus analysis. The extracted endmembers will be the ones that have a consensus among the multiple runs.

The third main contribution consists of developing subpixel target detectors that rely on the proposed CDSU algorithms to adapt target detection algorithms to different contexts. A local detection statistic is computed for each context and then all scores are combined to yield a final detection score. The context dependent unmixing provides a better background description and limits target leakage, which are two essential properties for target detection algorithms.

CONTEXT DEPENDENT SPECTRAL UNMIXING

By

Hamdi Jenzri

B.E., Telecommunications, Higher School of Communications of Tunis, 2009

A Dissertation

Submitted to the Faculty of the

J.B. Speed School of Engineering of the University of Louisville

in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

Department of Computer Engineering and Computer Science

University of Louisville

Louisville, Kentucky

August 2014

Copyright 2014 by Hamdi Jenzri

All rights reserved

CONTEXT DEPENDENT SPECTRAL UNMIXING

By

Hamdi Jenzri

B.E., Telecommunications, Higher School of Communications of Tunis, 2009

A Dissertation Approved On

August 11, 2014

by the following Dissertation Committee:

Hichem Frigui, Ph.D., Dissertation Director

Paul Gader, Ph.D.

Olfa Nasraoui, Ph.D.

Roman V. Yampolskiy, Ph.D.

Amir Amini, Ph.D.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Dr. Hichem Frigui, for his continuous guidance and support over the course of this work. Dr. Frigui has been an inspiring professor, mentor, and advisor. He helped build up my confidence and shape my career as a researcher. It has been a great learning opportunity as well as a pleasure working with him. I sincerely hope that we will remain both collaborators and friends for many years to come.

I would like to thank Dr. Paul Gader, Dr. Olfa Nasraoui, Dr. Roman V. Yampolskiy, and Dr. Amir Amini for agreeing to serve on my dissertation committee and being a part of this special milestone.

I would like to thank my colleagues in the Multimedia Research Laboratory and the Computer Engineering and Computer Science Department for their support and friendship.

I thank my friends in Louisville for the wonderful time we spent together. They helped me make this city home away from home.

Last but not least, I would like to thank my wife and my extended family for the love, patience, and support they always provide in all my endeavors.

ABSTRACT

CONTEXT DEPENDENT SPECTRAL UNMIXING

Hamdi Jenzri

August 11, 2014

A hyperspectral unmixing algorithm that finds multiple sets of endmembers is proposed. The algorithm, called Context Dependent Spectral Unmixing (CDSU), is a local approach that adapts the unmixing to different regions of the spectral space. It is based on a novel function that combines context identification and unmixing. This joint objective function models contexts as compact clusters and uses the linear mixing model as the basis for unmixing.

Several variations of the CDSU, that provide additional desirable features, are also proposed. First, the Context Dependent Spectral unmixing using the Mahalanobis Distance (CDSU_M) offers the advantage of identifying non-spherical clusters in the high dimensional spectral space. Second, the Cluster and Proportion Constrained Multi-Model Unmixing (CC-MMU and PC-MMU) algorithms use partial supervision information, in the form of cluster or proportion constraints, to guide the search process and narrow the space of possible solutions. The supervision information could be provided by an expert, generated by analyzing the consensus of multiple unmixing algorithms, or extracted from co-located data from a different sensor. Third, the Robust Context Dependent Spectral Unmixing (RCDSU) introduces possibilistic memberships into the objective function to reduce the effect of noise and outliers in the data. Finally, the Unsupervised Robust Context Dependent Spectral Unmixing (U-RCDSU) algorithm learns the optimal number of contexts in an unsupervised way. The performance of each algorithm is evaluated using synthetic and real data. We show that the proposed methods can identify meaningful and coherent contexts, and appropriate endmembers within each context.

The second main contribution of this thesis is consensus unmixing. This approach exploits the diversity and similarity of the large number of existing unmixing algorithms to identify an

accurate and consistent set of endmembers in the data. We run multiple unmixing algorithms using different parameters, and combine the resulting unmixing ensemble using consensus analysis. The extracted endmembers will be the ones that have a consensus among the multiple runs.

The third main contribution consists of developing subpixel target detectors that rely on the proposed CDSU algorithms to adapt target detection algorithms to different contexts. A local detection statistic is computed for each context and then all scores are combined to yield a final detection score. The context dependent unmixing provides a better background description and limits target leakage, which are two essential properties for target detection algorithms.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iv
ABSTRACT	v
LIST OF TABLES	x
LIST OF FIGURES	xii
LIST OF ALGORITHMS	xvi
CHAPTER	
1 INTRODUCTION	1
1.1 Hyperspectral image data and analysis	1
1.2 Linear Mixture Model	4
1.3 Motivations and overview of the proposed research	5
1.3.1 Motivations	5
1.3.2 Contributions	6
2 LITERATURE REVIEW	8
2.1 Linear spectral unmixing	8
2.1.1 Pure pixel based unmixing algorithms	8
2.1.2 Minimum volume based unmixing algorithms	10
2.2 Target detection using hyperspectral imaging	17
2.2.1 Definitions	18
2.2.2 Target detection using unstructured background models	20
2.2.3 Target detection using structured background models	21
3 CONTEXT DEPENDENT SPECTRAL UNMIXING	23
3.1 Motivations	23
3.1.1 Example 1	23
3.1.2 Example 2	24
3.1.3 Example 3	25
3.2 Context Dependent Spectral Unmixing	27

3.3	Context Dependent Spectral Unmixing Using the Mahalanobis Distance	31
3.4	Semi-supervised Context Dependent Spectral Unmixing	32
3.4.1	Cluster Constrained Multi-Model Unmixing	33
3.4.2	Proportion Constrained Multi-Model Unmixing	35
3.5	Robust Context Dependent Spectral Unmixing	36
3.6	Unsupervised Robust Context Dependent Spectral Unmixing	39
4	ROBUST UNMIXING USING CONSENSUS ANALYSIS	42
4.1	Motivations	42
4.2	Consensus Unmixing	43
5	CONTEXT DEPENDENT HYPERSPECTRAL SUBPIXEL TARGET	
	DETECTION	46
5.1	Motivations	46
5.2	Context Dependent Target Detectors	48
5.2.1	Context Dependent AMSD	48
5.2.2	Context Dependent OSP Detector	48
5.2.3	Context Dependent HSD	49
6	EXPERIMENTAL RESULTS	50
6.1	Data sets	50
6.1.1	Simulated data	50
6.1.2	Real data	52
6.2	Context Dependent Spectral Unmixing	54
6.2.1	Evaluation using simulated data	55
6.2.2	Evaluation using real data	63
6.3	Context Dependent Spectral Unmixing Using the Mahalanobis Distance	72
6.4	Cluster Constrained Multi-Model Unmixing	73
6.5	Proportion Constrained Multi-Model Unmixing	77
6.6	Robust Context Dependent Spectral Unmixing	77
6.7	Unsupervised Robust Context Dependent Spectral Unmixing	80
6.7.1	Evaluation using simulated data	80
6.7.2	Evaluation using real data	81

6.8	Robust Unmixing Using Consensus Analysis	85
6.8.1	Evaluation using simulated data	86
6.8.2	Evaluation using real data	86
6.9	Context Dependent Hyperspectral Subpixel Target Detection	89
6.9.1	Evaluation using implanted targets	89
6.9.2	Evaluation using actual targets	93
7	CONCLUSIONS AND POTENTIAL FUTURE WORK	99
7.1	Conclusions	99
7.2	Potential Future Work	102
7.2.1	Large scale evaluation	102
7.2.2	Target detection	103
7.2.3	Multi-sensor fusion	103
7.2.4	Non-linear unmixing	103
	REFERENCES	104
A	Proof of Theorem 3.2.1	110
B	Proof of Theorem 3.3.1	114
C	Proof of Theorem 3.4.1	117
D	Proof of Theorem 3.4.2	120
E	Proof of Theorem 3.5.1	122
	CURRICULUM VITAE	124

LIST OF TABLES

TABLE		Page
6.1	Parameters used for the CDSU and P-COMMEND algorithms on the USGS simulated hyperspectral data	58
6.2	Composition of the clusters generated by CDSU and P-COMMEND on the Usgs1C2M3 data with SNR = 20 dB	59
6.3	Parameters used for the CDSU and P-COMMEND algorithms on the Pavia University data	66
6.4	Composition of the clusters generated by P-COMMEND and CDSU, in percentage, for the Pavia University data, taking into account the labeled points only	67
6.5	Parameters used for the CC-MMU and CDSU algorithms on the University of Southern Mississippi data	74
6.6	Number of satisfied constraints using CC-MMU and CDSU with Euclidean and Mahalanobis distances on the University of Southern Mississippi data	74
6.7	Parameters used for P-COMMEND, CDSU _M and RCDSU on the Usgs2C2M3 data	78
6.8	Average ($\pm$ standard deviation) of the SMAE over 25 runs for P-COMMEND, CDSU _M and RCDSU on the Usgs2C2M3 data	78
6.9	Comparison of high proportion pixels (in %) among ground truth classes using consensus unmixing, N-FINDR, VCA, ICE and PPI on the Pavia University data	89
6.10	Average ($\pm$ standard deviation) of the AUC over 25 runs for the CD, EigVect, MVSA, NFINDR and PPI methods (Indian Pines data)	93
6.11	Targets in the University of Southern Mississippi data	94
6.12	Average ($\pm$ standard deviation) of the AUC over 25 runs for the CD, EigVect, MVSA, NFINDR and PPI methods (University of Southern Mississippi data, brown target)	95
6.13	Average ($\pm$ standard deviation) of the AUC over 25 runs for the CD, EigVect, MVSA, NFINDR and PPI methods (University of Southern Mississippi data, dark green target)	96

6.14	Average ($\pm$ standard deviation) of the AUC over 25 runs for the CD, EigVect, MVSA, NFINDR and PPI methods (University of Southern Mississippi data, faux vineyard green target)	97
6.15	Average ($\pm$ standard deviation) of the AUC over 25 runs for the CD, EigVect, MVSA, NFINDR and PPI methods (University of Southern Mississippi data, pea green target)	98

LIST OF FIGURES

FIGURE	Page
1.1 Illustration of the hyperspectral imaging concept.	2
1.2 Linear mixing from a checkerboard mixture of materials with a single reflection. . . .	3
1.3 Nonlinear mixing from an intimate mixture of materials.	4
1.4 Illustration of the 2-simplex. Green points represent spectral vectors. Red points represent vertices of the simplex and correspond to the endmembers.	5
2.1 Target detector structure.	19
2.2 Threshold selection trade-offs.	19
3.1 Example 1: (a) Synthetic two-dimensional data with one convex hull. (b) Endmem- bers detected by the ICE algorithm. (c) Endmembers detected by the P-COMMEND algorithm.	24
3.2 Example 2: (a) Synthetic two-dimensional data with two convex hulls. (b) Endmem- bers detected by the ICE algorithm with $M = 3$. (c) Endmembers detected by the ICE algorithm with $M = 6$. (d) Endmembers detected by the P-COMMEND algorithm. 25	25
3.3 Example 3: (a) Synthetic two-dimensional data with three convex hulls (colors rep- resent true cluster assignments). (b) Result of the P-COMMEND algorithm.	26
4.1 First 3 PC of the data (blue), endmembers from all unmixing runs (red), and true endmembers (green).	43
5.1 Target (red cross) leakage with non robust unmixing (red dots). No target leakage with robust unmixing (green dots).	47
5.2 Multiple endmember sets (green dots) versus a one global set (ref dots) for target (red cross) detection.	48
6.1 The D2EC2M3 synthetic data with two elongated convex hulls.	51
6.2 USGS spectra used to generate the synthetic hyperspectral data	51
6.3 Pavia University data	53
6.4 RGB image of the University of Southern Mississippi data	54
6.5 RGB image of the Indian Pines data	55

6.6	Result of the CDSU algorithm on the D2C1M3 data.	56
6.7	Result of the CDSU algorithm on the D2C2M3 data.	56
6.8	Result of the CDSU algorithm on the D2C3M3 data.	57
6.9	Error metrics for CDSU and P-COMMEND on the Usgs1C2M3 data across the 25 runs and at all noise levels: (a) <i>SME</i> , (b) <i>SMAE</i> , (c) <i>AME</i>	59
6.10	True (solid lines) and estimated (dashed lines) endmembers for the Usgs1C2M3 data with SNR = 20 dB using CDSU and P-COMMEND.	60
6.11	Membership values in cluster 1 for the Usgs1C2M3 data with SNR = 20 dB. Two principal components of the data are scatter plotted.	61
6.12	Membership values in cluster 2 for the Usgs1C2M3 data with SNR = 20 dB. Two principal components of the data are scatter plotted.	61
6.13	Running time (in seconds) of CDSU and P-COMMEND on the USGS simulated data across the 25 runs and at all noise levels.	62
6.14	Spectral signatures of the clusters generated by the FCM algorithm on the Usgs1C2M3 data with SNR = 20 dB.	64
6.15	Spectral signatures of the clusters generated by the CDSU algorithm on the Usgs1C2M3 data with SNR = 20 dB.	65
6.16	Evolution of the FCM objective function on the Usgs1C2M3 data (SNR = 20 dB).	66
6.17	Evolution of the objective function of CDSU on the Usgs1C2M3 data (SNR = 20 dB).	66
6.18	Cluster assignments of CDSU and P-COMMEND on the Pavia University data	67
6.19	Composition, in percentage, of the clusters generated by CDSU and P-COMMEND on the Pavia University data, taking into account the labeled points only.	68
6.20	Proportion maps estimated by the P-COMMEND algorithm for the Pavia University data. Each row of 3 proportion maps represents one context, and each column represents an endmember in that context.	69
6.21	Proportion maps estimated by the CDSU algorithm for the Pavia University data. Each row of 3 proportion maps represents one context, and each column represents an endmember in that context.	70
6.22	Estimated endmembers for the Pavia University data using CDSU and P-COMMEND.	71
6.23	Results of CDSU and CDSU _M on the D2EC2M3 data.	72
6.24	The University of Southern Mississippi data.	73

6.25	Proportion maps from CC-MMU on the University of Southern Mississippi data. Rows correspond to clusters and columns correspond to endmembers.	75
6.26	Proportion maps from CDSU on the University of Southern Mississippi data. Rows correspond to clusters and columns correspond to endmembers.	76
6.27	True (solid line) and estimated (dashed line) endmembers of CDSU _M and RCDSU for the Usgs2C2M3 data with 5% noise points.	79
6.28	Initial cluster assignments on the Usgs2C2M3 data with 5% noise points using RCDSU with $C_{max} = 5$	80
6.29	Cluster merging on the Usgs2C2M3 data with 5% noise points using U-RCDSU. .	81
6.30	Final cluster assignments on the Usgs2C2M3 data with 5% noise points using U- RCDSU.	81
6.31	True (solid line) and estimated (dashed line) endmembers for the Usgs2C2M3 data with 5% noise points using U-RCDSU.	82
6.32	Evolution of the cluster assignments using U-RCDSU with $C_{max} = 10$ on the Pavia University data.	83
6.33	Proportion maps from U-RCDSU on the Pavia University data. Rows correspond to clusters and columns correspond to endmembers.	84
6.34	Proportion maps from U-RCDSU on the University of Southern Mississippi data. Rows correspond to clusters and columns correspond to endmembers.	85
6.35	iVAT image of $\mathcal{H}_a$ of the Usgs3C1M4 data.	86
6.36	The PC of the Usgs3C1M4 data, the voters, the extracted endmembers and the true endmembers.	87
6.37	iVAT image of $\mathcal{H}_a$ of the Pavia University data.	87
6.38	RGB and proportion maps of the Pavia University data using consensus analysis.	88
6.39	Band 115 of the Indian Pines data with 100 implanted targets.	90
6.40	Scatter plot of the possibilistic memberships of the targets as a function of their proportions in the pixels (Indian Pines data).	91
6.41	Scatter plot of the CD-OSP statistic as a function of the target proportion in the pixel (Indian Pines data).	91
6.42	Receiver operating characteristic (ROC) curves (Indian Pines data).	92
6.43	Subimage of the University of Southern Mississippi data with target types, lo- cations and sizes	94

6.44 Principal components of the University of Southern Mississippi data (blue dots)	
with targets (red dots).	97

LIST OF ALGORITHMS

ALGORITHM	Page
2.1 Iterated Constrained Endmembers	12
2.2 Piece-wise Convex Multiple Model Endmember Detection	13
3.1 Context Dependent Spectral Unmixing (CDSU)	30
3.2 Context Dependent Spectral Unmixing using the Mahalanobis Distance (CDSU _M)	33
3.3 Cluster Constrained Multi-Model Unmixing (CC-MMU)	35
3.4 Proportion Constrained Multi-Model Unmixing (PC-MMU)	37
3.5 Robust Context Dependent Spectral Unmixing (RCDSU)	39
3.6 Unsupervised Robust Context Dependent Spectral Unmixing (U-RCDSU)	41
4.1 Robust Unmixing Using Consensus Analysis	45

CHAPTER 1

INTRODUCTION

1.1 Hyperspectral image data and analysis

Hyperspectral imaging contributes significantly to earth observation and remote sensing [1–13]. It is also used in food safety [14–17], pharmaceutical process monitoring and quality control [18–22], as well as in biomedical [23, 24], industrial [25], biometric [26] and forensic applications [27].

Hyperspectral sensors capture both the spatial and spectral information of a scene. They collect radiance data in hundreds of contiguous wavelengths. The focus is mainly on the visible, near-infrared and shortwave infrared spectral bands (in the range between $0.4\mu m$ and $2.5\mu m$ [7]). As the sensor collects data over a region, a data cube is generated. Figure 1.1 illustrates this concept. The data cube can be interpreted as a stack of two-dimensional images captured over a range of wavelengths. Each element of the data cube corresponds to the radiance measured at a particular wavelength at one ground location [1, 28, 29].

Spectral and spatial resolutions are two important characteristics of a hyperspectral sensor. The spectral resolution of a sensor corresponds to the range of wavelengths over which radiance values are measured and combined to become a single band in a hyperspectral image. The spatial resolution corresponds to the size of the physical area on the ground from which radiance measurements are taken for a single image pixel. As the area corresponding to one pixel increases, the spatial resolution of the image decreases [1, 28].

The main appeal for hyperspectral imaging is that different materials reflect and emit varying amounts of radiance across the electromagnetic spectrum. In other words, different materials generally have unique spectral signatures. Thus, hyperspectral sensors can be used to identify and distinguish between different materials in a scene [28].

In a hyperspectral image, a pixel can be *pure* or *mixed*. A *pure* pixel corresponds to a single material's radiance values. On the other hand, a *mixed* pixel is a pixel which combines the radiance values of multiple materials. Typically, most pixels are mixed as a result of the low spatial resolution [1, 28]. This is because as the pixel's corresponding area on the ground increases,

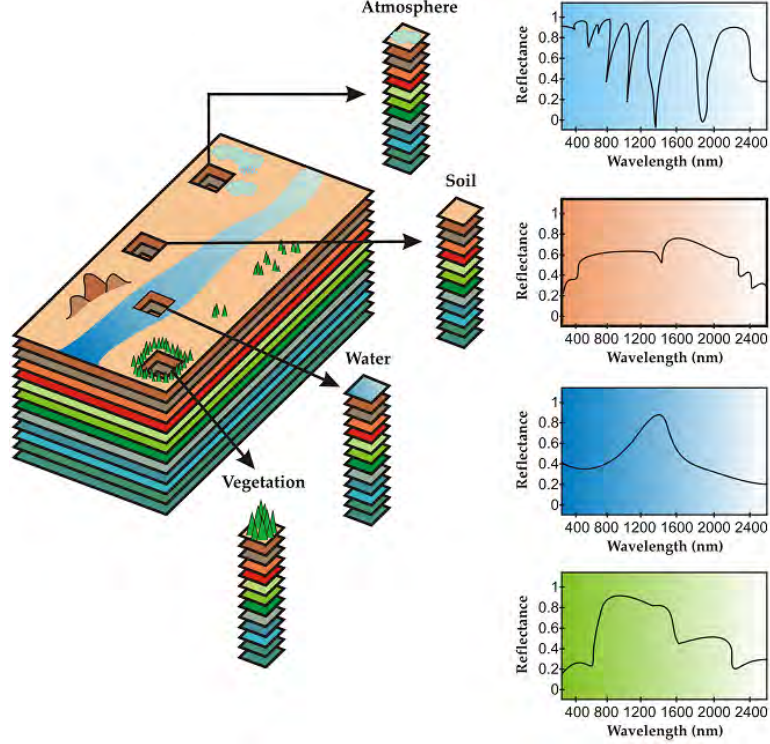


Figure 1.1: Illustration of the hyperspectral imaging concept.¹

neighboring materials are likely to be combined into the image pixel. Mixed pixels also occur when the different materials are mixed on the ground. Beach sand is a common example for this type of mixed pixel since grains of different materials are intermingled [1].

Pure spectral signatures, or the *constituent spectra*, in an imaged scene are referred to as *endmembers* [1]. In the purest sense, endmembers can represent unique elements, e.g., calcium, iron, and copper. However, in the practical sense of hyperspectral imaging, the endmembers more likely represent disparate macroscopic entities, e.g., desert, forest, metal, and salt water [30]. Due to the presence of mixed pixels in a hyperspectral image, *spectral unmixing* is often performed to decompose mixed pixels into their respective endmembers and *abundances*. Abundances are the proportions of the endmembers in each pixel. In estimating the endmembers and abundances in pixel spectra, unmixing algorithms incorporate philosophical assumptions regarding the physical mechanisms and mathematical structure by which the reflectance properties from disparate substances combine to yield the mixed pixel spectra. In other words, spectral unmixing relies on the definition of a mixing model.

Mixing models can be characterized as either *linear* or *nonlinear* [1,31]. The *linear mixing*

¹This image was taken from [29].

model (also known as the *convex geometry model*) holds when the mixing scale is macroscopic [32] and the incident light interacts with just one material, as is the case in checkerboard type scenes [33,34]. In this case, the mixing occurs within the sensor itself. It is due to the fact that the spatial resolution of the sensor is not fine enough. The light from the materials, although almost completely separated, is mixed within the measuring sensor. Figure 1.2 depicts linear mixing where the reflecting surface is portrayed as a checkerboard mixture, and the incident radiation bounces only once on its surface. If the total surface area is conceived to be divided proportionally according to the fractional abundances of the constituent substances, then the reflected radiation will convey, with the same proportions, the characteristics of the associated media.

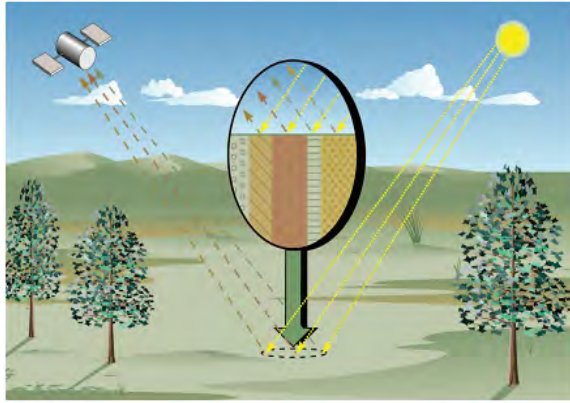


Figure 1.2: Linear mixing from a checkerboard mixture of materials with a single reflection.¹

The nonlinear mixing is usually due to the light interaction with multiple materials in the scene. These interactions can be at a classical or multilayered level, or at a microscopic or intimate level. Mixing at the classical level occurs when light scattered from one or more objects, is reflected off additional objects, and eventually measured by the hyperspectral sensor. Microscopic mixing occurs when two materials are homogeneously mixed [35]. In this case, the interactions consist of photons emitted from the molecules of one material that are absorbed by molecules of another material, which may in turn emit more photons. Figure 1.3 illustrates nonlinear mixing from an intimate mixture.

Despite its simplicity, the linear mixing model has proved to be an acceptable approximation of the light scattering mechanisms in many real scenarios. Furthermore, in contrast to nonlinear mixing, the linear mixing model is the basis of a plethora of unmixing algorithms [1,30,36–48].

¹This image was taken from [1].

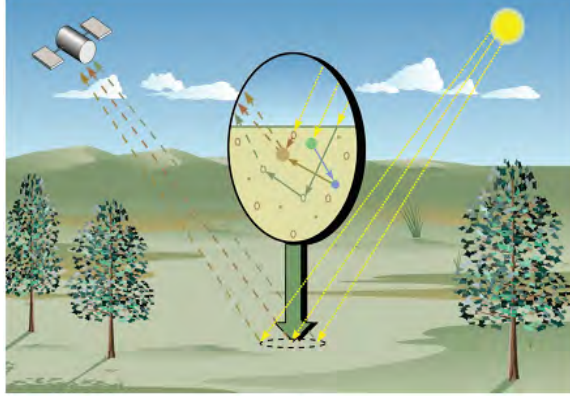


Figure 1.3: Nonlinear mixing from an intimate mixture of materials.¹

1.2 Linear Mixture Model

The standard model used to perform spectral unmixing is the *convex geometry model* (also known as the *linear mixing model*). It states that every pixel's spectral signature is a convex combination of endmembers in the scene. This has been shown to hold in cases where the endmembers are mixed by the spatial resolution of the imaging sensor [1, 36]. If the convex geometry holds, the endmembers are the spectra found at the corners of a convex region enclosing all the spectra in the hyperspectral scene. This model is defined as [1]:

$$\mathbf{x}_j = \sum_{k=1}^M p_{jk} \mathbf{e}_k + \epsilon_j, \quad j = 1, \dots, N \quad (1.1)$$

where $\mathbf{x}_j$ ($1 \times d$) is the spectral signature of pixel j , d is the number of spectral bands, N is the number of pixels in the image, M is the number of endmembers, ϵ_j ($1 \times d$) is an error term, p_{jk} is the proportion of endmember k in pixel j , and $\mathbf{e}_k$ ($1 \times d$) is the k^{th} endmember. The proportions satisfy the following constraints:

$$p_{jk} \geq 0, \quad \forall k = 1, \dots, M; \quad \text{and} \quad \sum_{k=1}^M p_{jk} = 1, \quad \forall j = 1, \dots, N. \quad (1.2)$$

In the hyperspectral unmixing literature, the constraints in (1.2) are referred to as *abundance non-negativity constraint* (ANC) and *abundance sum constraint* (ASC), respectively. Equation (1.1) can be rewritten in a matrix format as:

$$\mathbf{x}_j = \mathbf{p}_j \mathbf{E} + \epsilon_j, \quad j = 1, \dots, N. \quad (1.3)$$

where $\mathbf{E}$ ($M \times d$) is the mixing matrix containing the endmembers, $\mathbf{p}_j$ ($1 \times M$) are the proportions of pixel j in the M endmembers, and ϵ_j ($1 \times d$) is an error term. The constraints in (1.2) can also

¹This image was taken from [1].

be rewritten in a matrix format as:

$$\mathbf{p}_j \geq \mathbf{0}, \text{ and } \mathbf{1}_{1 \times M} \mathbf{p}_j^T = 1, \forall j = 1, \dots, N. \quad (1.4)$$

where $\mathbf{1}_{1 \times M}$ is a $1 \times M$ vector of ones.

Assuming that the rows of $\mathbf{E}$ are affinely independent, i.e., $\mathbf{e}_2 - \mathbf{e}_1, \mathbf{e}_3 - \mathbf{e}_1, \dots, \mathbf{e}_M - \mathbf{e}_1$ are linearly independent, then the set

$$\left\{ \mathbf{x} = \mathbf{pE}, \text{ such that } \sum_{k=1}^M p_k = 1, p_k \geq 0, k = 1, \dots, M \right\} \quad (1.5)$$

i.e., the convex hull of the rows of $\mathbf{E}$ ($\text{conv}\{\mathbf{E}\}$), is a $(M-1)$ -simplex in $\mathbb{R}^d$. Figure 1.4 illustrates a 2-simplex for a hypothetical mixing matrix $\mathbf{E}$ containing three endmembers. In this figure, the green points denote spectral vectors, and the red points are the vertices of the simplex and correspond to the endmembers. Note that the inference of the mixing matrix $\mathbf{E}$ is equivalent to identifying the vertices of the simplex. This is referred to as geometrical-based unmixing.

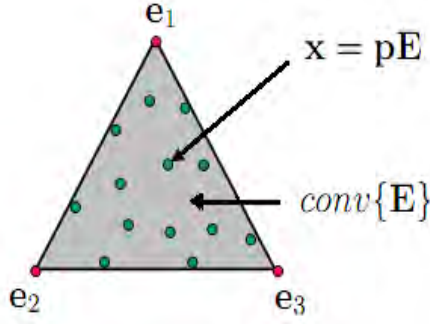


Figure 1.4: Illustration of the 2-simplex. Green points represent spectral vectors. Red points represent vertices of the simplex and correspond to the endmembers.

Given a data set $\mathbf{X}$ ($N \times d$) containing N d -dimensional spectral vectors, the linear hyperspectral unmixing problem, with reference to the linear model (1.3), consists of estimating the mixing matrix $\mathbf{E}$ and the fractional abundance vectors $\mathbf{p}_j$ for each pixel $j = 1, \dots, N$.

1.3 Motivations and overview of the proposed research

1.3.1 Motivations

Most of the existing linear spectral unmixing algorithms assume that the hyperspectral data points lie in a single convex region with one set of endmembers. However, it may be the case that multiple sets of endmembers, defining several overlapping convex regions, can better describe the hyperspectral image. This issue has been addressed in [49–53], where the linear mixing model has

been extended to multiple sets of endmembers. Each endmember set is found using the convex geometry model resulting in a piece-wise convex representation of the hyperspectral data. Another limitation of existing spectral unmixing algorithms is that they do not take into account the distribution of the data in the spectral space while unmixing. This is the case even for the piece-wise convex representation in [49].

To address the above limitations of linear spectral unmixing, we propose a local hyperspectral unmixing algorithm, called *Context Dependent Spectral Unmixing* (CDSU). CDSU takes into account the distribution of the data in the spectral space while identifying multiple sets of endmembers. In other words, the unmixing process is adapted to different regions of the spectral space.

Another challenge with most unmixing algorithms is that they require the knowledge of the number of endmembers to be extracted before hand. Moreover, different algorithms have different assumptions and modes of operation, usually yielding different results. Even the same algorithm may not result in the same endmembers when run multiple times. This is mainly due to the non-deterministic behavior of the algorithm. To address this limitation, we investigate using multiple algorithms with different parameters to identify an accurate and consistent set of endmembers using consensus analysis.

Spectral unmixing is a goal in itself, where one is interested in identifying the materials present in the scene. Unmixing is also an initial step to other hyperspectral imaging applications, such as target detection. In fact, spectral unmixing is used to describe the background with a set of endmembers, based on which a detection statistic is computed for every pixel. Background description is of paramount importance in target detection. A better description allows for a better detection. Hence, we propose using the context dependent unmixing framework to design a new class of detectors called Context Dependent Target Detectors.

1.3.2 Contributions

Our main contributions can be summarized as follows:

- We propose a Context Dependent Spectral Unmixing (CDSU) [54] algorithm. CDSU is based on optimizing an objective function that combines *context identification* and *spectral unmixing*. The context or region identification component strives to partition the input spectral space into different clusters (called *contexts*). The spectral unmixing component strives to learn optimal endmembers and abundances within each cluster.

- We propose an extension to CDSU using the Mahalanobis distance (CDSU_M) [55]. CDSU_M supports non-spherical cluster shapes.
- We propose two semi-supervised versions of CDSU that use partial supervision information to constrain the problem, guide the optimization and narrow the space of possible solutions. The Cluster Constrained Multi-Model Unmixing (CC-MMU) [55] algorithm uses cluster assignment constraints on the pixels, while the Proportion Constrained Multi-Model Unmixing (PC-MMU) algorithm uses constraints on the proportions of the pixels.
- We propose a Robust Context Dependent Spectral Unmixing (RCDSU) [56] algorithm. RCDSU handles noise and outliers in the data, and finds the optimal number of contexts in an unsupervised way (U-RCDSU) [56].
- We propose a robust unmixing approach based on consensus analysis [57]. We run multiple unmixing algorithms using different parameters, and the goal is to find a consensus unmixing by combining the unmixing ensemble resulting from those algorithms.
- We propose a new class of target detection algorithms, called Context Dependent Target Detectors [58], that takes advantage of the context dependent unmixing framework. The detection is performed locally within the extracted contexts, and a global detection statistic is computed as a weighted sum of the local scores.

The remainder of this dissertation is organized as follows. Chapter 2 provides a review of some linear spectral unmixing and target detection algorithms. Chapter 3 introduces the proposed context dependent spectral unmixing algorithm and its variations. Chapter 4 introduces the proposed robust unmixing using consensus analysis. Chapter 5 introduces the proposed context dependent target detection algorithms. Chapter 6 provides experimental results and analyses of the proposed methods. Finally, chapter 7 provides conclusions and potential future work.

CHAPTER 2

LITERATURE REVIEW

This chapter provides a review of some existing linear hyperspectral unmixing and target detection algorithms. Methods that are relevant to our work are described in details.

2.1 Linear spectral unmixing

Spectral unmixing algorithms can be categorized into two main categories: pure pixel based and minimum volume based approaches.

2.1.1 Pure pixel based unmixing algorithms

The pure pixel based algorithms belong to the minimum volume based approaches. They have the additional assumption that the data have at least one pure pixel per endmember. In other words, they assume that there is at least one spectral vector on each vertex of the data simplex. This assumption is a strong requisite that may not hold in many datasets. These algorithms find the set of purest pixels in the data. They have been often used in linear hyperspectral unmixing applications, mainly because of their light computational burden and clear conceptual meaning [29]. Algorithms relying on the pixel purity assumption include the Pixel Purity Index (PPI) algorithm [59] and the N-FINDR algorithm [60], both of which are described in the next subsections.

2.1.1.1 The Pixel Purity Index

The Pixel Purity Index (PPI) [59] is a commonly used algorithm for determining the purest pixels in a given hyperspectral image. It ranks image pixels based on their purity indices. Then, the M pixels with the highest purity values are returned as potential endmembers. The number of endmembers, M , is assumed to be known. PPI is often used for generating candidate endmembers which are then used as inputs to other endmember extraction algorithms [61] or loaded into a visualization tool for users to manually select endmembers from the list of potential candidates [62]. The PPI algorithm assigns each pixel a purity value by repeatedly projecting all of the pixels onto

skewers, defined as a large set of random vectors. The algorithm is initialized by setting the purity of each pixel to zero. The pixel purity values are updated following each random projection by adding one to the values of the pixels that fall near either end of every projection. Since PPI values are generated using random vectors, the results are dependent on the number of random projections and the threshold for determining if a pixel's projection is considered near an end-point [59].

2.1.1.2 N-FINDR

N-FINDR [60] is based on the fact that the volume defined by a simplex formed by the purest pixels is larger than any other volume defined by any other combination of pixels. This algorithm finds the set of pixels defining the largest volume by inflating a simplex inside the data. The algorithm begins by randomly selecting a set of M pixels from the image to be the initial endmember set $\mathbf{E}$. Then, each endmember is replaced, in succession, by all other pixels in the image. After each replacement, the volume of the space defined by the current set of potential endmembers is computed. When a replacement increases the volume, the replacement is maintained. The algorithm cycles through all image pixels and endmembers until no further replacements are made.

The volume enclosed by each set of potential endmembers is computed using:

$$V(\mathbf{E}^*) = \frac{1}{(M-1)!} \text{abs}(|\mathbf{E}^*|), \quad (2.1)$$

where $\text{abs}(\cdot)$ refers to the absolute value, $|\cdot|$ refers to the determinant, and

$$\mathbf{E}^* = [\mathbf{1}_{M \times 1}, \mathbf{E}]. \quad (2.2)$$

If the dimensionality of the data is larger than $(M-1)$, then a dimensionality reduction method, such as Principal Components Analysis or Maximum Noise Fraction, must be employed [63,64]. The data dimensionality must be one less than the desired number of endmembers since the determinant of a non-square matrix is not defined [60]. In addition to assuming that pure pixels can be found in the image, N-FINDR requires the knowledge of the number of endmembers in advance.

Other pure pixel based algorithms include the Iterative Error Analysis (IEA) algorithm [65], the Vertex Component Analysis (VCA) algorithm [66], the Simplex Growing Algorithm (SGA) [67], and the Sequential Maximum Angle Convex Cone (SMACC) algorithm [68]. The IEA implements a series of linear constrained unmixings, each time choosing as endmembers those pixels which minimize the remaining error in the unmixed image. The VCA algorithm iteratively projects data onto

a direction orthogonal to the subspace spanned by the detected endmembers. The new endmember signature corresponds to the maximum of the projection. The algorithm iterates until all endmembers are exhausted. The SGA iteratively grows a simplex by finding the vertices, one at a time, corresponding to the maximum volume. The SMACC algorithm represents the endmembers using a convex cone. It starts with a single endmember and every iteration the data vector, making the maximum angle with the existing cone, is chosen as the next endmember. The algorithm terminates when all of the data vectors are within the convex cone, to some tolerance. A quantitative and comparative analysis of these methods is given in [39].

2.1.2 Minimum volume based unmixing algorithms

The minimum volume approaches seek a mixing matrix $\mathbf{E}$ that minimizes the volume of the simplex defined by its rows (endmembers), subject to the constraint that it contains the observed spectral vectors. The pure pixel constraint is no longer enforced.

Examples of minimum volume based algorithms include the Nonnegative Matrix Factorization Minimum Volume Transform (NMF-MVT) algorithm [69], the Minimum Volume Simplex Analysis (MVSA) algorithm [70], and the Simplex Identification via Split Augmented Lagrangian (SISAL) algorithm [71].

In the following, we focus on two minimum volume based algorithms that are closely related to our proposed approach. The first one is the Iterated Constrained Endmembers (ICE) algorithm [61] which fits a simplex to the data while penalizing its volume. The second one is the Piecewise Convex Multiple Model Endmember Detection (P-COMMEND) algorithm [49] which models a hyperspectral image using a piece-wise convex representation.

2.1.2.1 ICE: Iterated Constrained Endmembers

The ICE algorithm [61] is based on the joint optimization of two terms. The first term is the residual sum of squares (RSS) based on the convex geometry model in equation (1.1). This term is defined as

$$RSS = \sum_{j=1}^N \left(\mathbf{x}_j - \sum_{k=1}^M p_{jk} \mathbf{e}_k \right) \left(\mathbf{x}_j - \sum_{k=1}^M p_{jk} \mathbf{e}_k \right)^T, \quad (2.3)$$

where $\mathbf{x}_j$ ($1 \times d$) is the spectral signature of pixel j , d being the number of spectral bands, N is the number of pixels in the image, M is the number of endmembers, p_{jk} is the proportion of endmember k for pixel j , and $\mathbf{e}_k$ ($1 \times d$) is the k^{th} endmember. The proportions satisfy the constraints in equation (1.2).

The second term of the ICE algorithm is the sum of squared distances (SSD) between all the simplex vertices. This term represents an approximation of the volume of the simplex. Therefore, by adding this term, ICE finds endmembers that provide a tight fit around the data. The SSD term is defined as

$$SSD = \sum_{k=1}^{M-1} \sum_{l=k+1}^M (\mathbf{e}_k - \mathbf{e}_l)(\mathbf{e}_k - \mathbf{e}_l)^T. \quad (2.4)$$

The objective function of the ICE algorithm is a weighted sum of both RSS and SSD terms [62]. It is defined as

$$J = (1 - \mu) \frac{RSS}{N} + \mu \frac{SSD}{M(M-1)}, \quad (2.5)$$

where μ is a tradeoff or a regularization parameter in $(0, 1)$ used to balance the RSS and SSD terms. In [61], the authors recommend using one value of μ for all datasets. For this to be possible, the objective function should be approximately independent of the sample size N and the number of endmembers M . This is the reason for normalizing RSS by N and SSD by $M(M-1)$ in (2.5).

It is instructive to consider the limiting behavior of μ near its extreme values, 0 and 1. As μ tends to 0, the limiting solution is an $(M-1)$ -simplex which totally encloses the data points while ignoring the tightness of fit. On the other hand, as μ tends to 1, the algorithm results in a trivial solution where all the endmembers converge to one point, the mean of the data.

The ICE algorithm minimizes the objective function in (2.5) iteratively. First, given endmember estimates, the proportions for each pixel are estimated. Initially, endmembers may be set to randomly chosen pixels from the image. Estimating the proportions requires a least squares minimization of each term in equation (2.3). Since each of these terms is quadratic and subjected to the linear constraints in equation (1.2), the minimization can be achieved using quadratic programming. After solving for the proportions, the endmembers are updated using the current proportion estimates [61].

Like most iterative solutions to nonlinear continuous parameter optimization problems, the ICE algorithm will asymptotically approach a local minimum of the objective function [61]. Therefore, the iterative procedure is stopped when the estimated parameters do not change significantly between successive iterations. Algorithm 2.1 illustrates the steps of the ICE algorithm.

Although ICE is an effective algorithm for finding endmembers, it can provide only a single set of endmembers for the entire input data set. This may not provide a compact description of the hyperspectral scene. An attempt to alleviate this shortcoming was proposed in [49], where the linear mixture model has been extended to multiple sets of endmembers. It is called the Piece-wise Convex Multiple Model Endmember Detection (P-COMMEND) algorithm.

Algorithm 2.1 Iterated Constrained Endmembers

Inputs: $\mathbf{X}$: the data points ($N \times d$).
 M : the number of endmembers.
 μ : the regularization parameter $\mu \in (0, 1)$.
Outputs: $\mathbf{E}$: the estimated endmembers.
 $\mathbf{P}$: the estimated proportions.

Initialize $\mathbf{E}$
repeat
 Update $\mathbf{P}$.
 Update $\mathbf{E}$.
until parameters do not change significantly
return $\mathbf{E}, \mathbf{P}$

2.1.2.2 P-COMMEND: Piece-wise Convex Multiple Model Endmember Detection

P-COMMEND [49] is a hyperspectral unmixing algorithm that finds multiple sets of endmembers. It models a hyperspectral image using a piece-wise convex representation to characterize non-convex hyperspectral data. It assumes that a hyperspectral scene contains multiple distinct regions that do not share common materials. Each region is defined by a simplex with a set of endmembers.

P-COMMEND estimates endmember sets and proportion values by minimizing

$$J = \sum_{i=1}^C \left(\sum_{j=1}^N u_{ij}^m (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i) (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T + \alpha \sum_{k=1}^{M-1} \sum_{l=k+1}^M (\mathbf{e}_{ik} - \mathbf{e}_{il})(\mathbf{e}_{ik} - \mathbf{e}_{il})^T \right), \quad (2.6)$$

subject to

$$u_{ij} \in [0, 1], \forall i, j, \sum_{i=1}^C u_{ij} = 1, \forall j \quad (2.7)$$

and

$$\mathbf{p}_{ij} \geq 0, \text{ and } \mathbf{1}_{1 \times M} \mathbf{p}_{ij}^T = 1, \forall i, j. \quad (2.8)$$

where $\mathbf{x}_j$ ($1 \times d$) is the spectral signature of pixel j , d is the number of spectral bands, N is the number of pixels in the image, M is the number of endmembers, and C is the number of models, convex regions, or sets of endmembers. In (2.6), $\mathbf{p}_{ij}$ ($1 \times M$) is the vector of proportions associated with pixel j with respect to model i , $\mathbf{E}_i$ ($M \times d$) is the mixing matrix corresponding to model i , and $\mathbf{e}_{ik}$ ($1 \times d$) is the k^{th} row of $\mathbf{E}_i$ representing the k^{th} endmember in the i^{th} endmember set. In (2.6), u_{ij} represents the membership of pixel j in model i , indicating the degree to which pixel j contributes to the endmembers of convex set i . Finally, $m \in (1, +\infty)$ is a fuzzifier controlling the degree of sharing between the models, and α is a fixed parameter used to balance the two terms of the objective function.

The first term in (2.6) computes the residual sum of squares (RSS) between each input hyperspectral data point and its estimate using the estimated endmembers and proportion values. The second term is used to constrain the size of each simplex by minimizing the sum of squared distances (SSD) between each pair of endmembers within each set. Note that, when $C = 1$, the P-COMMEND objective function reduces to the ICE objective function in equation (2.5) (to a normalization factor).

In [49], the authors showed that, using Lagrange multipliers optimization along with the Karush-Kuhn-Tucker (KKT) conditions, the objective function in (2.6) can be minimized by updating the endmembers, the proportions and the fuzzy memberships using

$$\mathbf{E}_i = \left[\alpha(M\mathbf{I}_{M \times M} - \mathbf{1}_{M \times M}) + \sum_{j=1}^N u_{ij}^m \mathbf{p}_{ij}^T \mathbf{p}_{ij} \right]^{-1} \left[\sum_{j=1}^N u_{ij}^m \mathbf{p}_{ij}^T \mathbf{x}_j \right], \quad (2.9)$$

$$\mathbf{p}_{ij}^T = \max \left(\left[\mathbf{E}_i \mathbf{E}_i^T \right]^{-1} \left[\mathbf{E}_i \mathbf{x}_j^T + \frac{1 - \mathbf{1}_{1 \times M} (\mathbf{E}_i \mathbf{E}_i^T)^{-1} \mathbf{E}_i \mathbf{x}_j^T}{\mathbf{1}_{1 \times M} (\mathbf{E}_i \mathbf{E}_i^T)^{-1} \mathbf{1}_{M \times 1}} \mathbf{1}_{M \times 1} \right], 0 \right), \quad (2.10)$$

and

$$u_{ij} = \frac{\left[(\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i) (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T \right]^{\frac{1}{1-m}}}{\sum_{q=1}^C \left[(\mathbf{x}_j - \mathbf{p}_{qj} \mathbf{E}_q) (\mathbf{x}_j - \mathbf{p}_{qj} \mathbf{E}_q)^T \right]^{\frac{1}{1-m}}}. \quad (2.11)$$

In (2.9) and (2.10), the notation $[\mathbf{A}]^{-1}$ refers to the inverse of matrix $\mathbf{A}$.

The P-COMMEND algorithm is outlined in Algorithm 2.2 [49].

Algorithm 2.2 Piece-wise Convex Multiple Model Endmember Detection

Inputs: $\mathbf{X}$: the data points ($N \times d$).

C : the number of models.

M : the number of endmembers for each model.

m : the fuzzifier, $m \in (1, +\infty)$.

α : the weight of the second term in the objective function.

Outputs: $\mathbf{U}$: the fuzzy membership matrix of the data samples.

$\mathbf{E}_i$: the sets of endmembers in all models.

$\mathbf{P}_i$: the sets of proportions in all models.

Initialize $\mathbf{U}$ and $\mathbf{E}_i$.

repeat

 Update $\mathbf{P}_i$ using (2.10).

 Update $\mathbf{E}_i$ using (2.9).

 Update $\mathbf{U}$ using (2.11).

until parameters do not change significantly

return $\mathbf{U}$, $\mathbf{E}_i$, $\mathbf{P}_i$

In the implementation of the P-COMMEND algorithm [49], the membership values are initialized using the Fuzzy C-Means algorithm [72] (which is, in turn, randomly initialized), and the

endmember sets $\mathbf{E}_i$ are initialized using the Minimum Volume Simplex Analysis (MVSA) algorithm [70]. The algorithm is stopped whenever the estimated parameters do not change significantly between successive iterations.

Other variations of the P-COMMEND algorithm have been proposed. These include a multiple model endmember detection algorithm based on spectral and spatial information [50], a spatially-smooth piece-wise convex endmember detection algorithm [51], a competitive agglomeration piece-wise convex multiple model endmember detection algorithm [52], and a piece-wise convex spatial-spectral unmixing algorithm using possibilistic and fuzzy clustering [53].

The multiple model endmember detection algorithm based on spectral and spatial information [50] adds a spatial information term to the P-COMMEND objective function. The idea is to fit the hyperspectral data using a convex geometry model locally. The assignment of a point to a model is done according to its spectral and spatial information. Using the same notation as for the P-COMMEND objective function above, the objective function is defined as:

$$J = \sum_{i=1}^C \left[\sum_{j=1}^N u_{ij}^m (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i) (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T + \rho \sum_{j=1}^N u_{ij}^m (\mathbf{y}_j - \mathbf{c}_i) (\mathbf{y}_j - \mathbf{c}_i)^T + \alpha \sum_{k=1}^{M-1} \sum_{l=k+1}^M (\mathbf{e}_{ik} - \mathbf{e}_{il}) (\mathbf{e}_{ik} - \mathbf{e}_{il})^T \right], \quad (2.12)$$

subject to the constraints in (2.7) and (2.8).

In (2.12), $\mathbf{y}_j$ is the 2-dimensional spatial coordinate vector of pixel j , $\mathbf{c}_i$ is the spatial center of the points assigned to the i^{th} model, and ρ is a scaling parameter. Using Lagrange multipliers optimization along with the Karush-Kuhn-Tucker (KKT) conditions, it was shown [50] that the objective function in (2.12) can be minimized by updating the endmembers and the proportions using (2.9), (2.10), and the fuzzy memberships and centers using

$$u_{ij} = \frac{[(\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)(\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T + \rho(\mathbf{y}_j - \mathbf{c}_i)(\mathbf{y}_j - \mathbf{c}_i)^T]^{\frac{1}{1-m}}}{\sum_{q=1}^C [(\mathbf{x}_j - \mathbf{p}_{qj} \mathbf{E}_q)(\mathbf{x}_j - \mathbf{p}_{qj} \mathbf{E}_q)^T + \rho(\mathbf{y}_j - \mathbf{c}_q)(\mathbf{y}_j - \mathbf{c}_q)^T]^{\frac{1}{1-m}}}, \quad (2.13)$$

and

$$\mathbf{c}_i = \frac{\sum_{j=1}^N u_{ij}^m \mathbf{y}_j}{\sum_{j=1}^N u_{ij}^m}. \quad (2.14)$$

The spatially-smooth piece-wise convex endmember detection algorithm (Spatial P-COMMEND) [51] extends P-COMMEND by incorporating spatial information to aid in estimating endmembers and abundance values. Spatial information is incorporated by encouraging neighboring pixels in the image to have similar membership values to the different convex regions. This is accomplished by

adopting the spatially-smooth Fuzzy Local Information C-Means method, FLICM, developed in [73]. The FLICM algorithm adds a *fuzzy factor* term, G , to the objective function that influences the updates of the membership values by incorporating spatial information. The G term, adapted for use in the Spatial P-COMMEND algorithm, is defined as:

$$G_{ij} = \sum_{\substack{k \in N_j \\ k \neq j}} \frac{1}{d_{jk} + 1} (1 - u_{ik})^m \|\mathbf{x}_k - \mathbf{p}_{ik} \mathbf{E}_i\|_2^2, \quad (2.15)$$

where j is the center pixel in the local window under consideration, N_j is the neighborhood around the center pixel (such as a 3×3 window), and d_{jk} is the Euclidean distance in pixel space of the image indices between pixels j and k . Therefore, the G term scales the influence of neighboring pixels based on their distances to the center pixel in the index space. When a neighboring pixel has high membership in a convex region, the center pixel under consideration is encouraged to also have a high membership in that region. The fuzzy factor term is updated every iteration and treated as a constant during the updates of the endmembers, abundances and membership values. Adding the term in (2.15) and using the same notation as for the P-COMMEND objective function, the objective function of the Spatial P-COMMEND is defined as:

$$J = \sum_{i=1}^C \left(\sum_{j=1}^N u_{ij}^m \left[(\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)(\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T + G_{ij} \right] + \alpha \sum_{k=1}^{M-1} \sum_{l=k+1}^M (\mathbf{e}_{ik} - \mathbf{e}_{il})(\mathbf{e}_{ik} - \mathbf{e}_{il})^T \right), \quad (2.16)$$

subject to the constraints in (2.7) and (2.8).

In Spatial P-COMMEND, the update equations for the endmembers and abundances remain the same as in equations (2.9) and (2.10) respectively. The membership update equation becomes

$$u_{ij} = \frac{[(\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)(\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T + G_{ij}]^{\frac{1}{1-m}}}{\sum_{q=1}^C [(\mathbf{x}_j - \mathbf{p}_{qj} \mathbf{E}_q)(\mathbf{x}_j - \mathbf{p}_{qj} \mathbf{E}_q)^T + G_{qj}]^{\frac{1}{1-m}}}. \quad (2.17)$$

The Spatial P-COMMEND algorithm performs alternating optimization on the endmembers, abundances and memberships until some stopping criterion is reached such as convergence or a maximum number of iterations.

The competitive agglomeration piece-wise convex multiple model endmember detection algorithm (CAP) [52] integrates the competitive agglomeration algorithm [74] into P-COMMEND in order to estimate the number of convex regions needed for a given data set. The competitive agglomeration algorithm uses a regularization term with the sum of squares of the cardinalities (sum of the memberships) of the clusters. Using the same notation as for the P-COMMEND objective

function, the objective function of CAP is defined as:

$$J = \sum_{i=1}^C \left[\sum_{j=1}^N u_{ij}^2 (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i) (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T + \alpha \sum_{k=1}^{M-1} \sum_{l=k+1}^M (\mathbf{e}_{ik} - \mathbf{e}_{il})(\mathbf{e}_{ik} - \mathbf{e}_{il})^T - \beta \left(\sum_{j=1}^N u_{ij} \right)^2 \right], \quad (2.18)$$

subject to the constraints in (2.7) and (2.8). In (2.18), β is a scaling parameter.

The update equations for the endmembers and proportion values remain the same as in P-COMMEND (equations (2.9) and (2.10) respectively) with a fuzzifier value of 2. The update equation for the membership values becomes

$$u_{ij} = \frac{\beta N_i + \lambda_j}{(\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)(\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T}, \quad (2.19)$$

where

$$\lambda_j = \frac{1 - \beta \sum_{k=1}^C \frac{N_k}{\|\mathbf{x}_j - \mathbf{p}_{kj} \mathbf{E}_k\|_2^2}}{\sum_{k=1}^C \frac{1}{\|\mathbf{x}_j - \mathbf{p}_{kj} \mathbf{E}_k\|_2^2}}, \quad (2.20)$$

and

$$N_k = \sum_{j=1}^N u_{kj}. \quad (2.21)$$

Competitive agglomeration encourages sparsity in the membership values. When the membership values associated with a single convex region drop below a prescribed threshold, the convex region can be removed. Following the discussion in [74], the parameter for the competitive agglomeration term is adjusted every iteration using the following annealing schedule:

$$\beta(t) = \beta_0 e^{\frac{-t}{\tau}} \frac{\sum_{i=1}^C \sum_{j=1}^N u_{ij}^2 \|\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i\|_2^2}{\sum_{i=1}^C \left(\sum_{j=1}^N u_{ij} \right)^2}, \quad (2.22)$$

where t is the iteration number, and τ and β_0 are constants. This schedule for the β parameter tries to balance the residual error term and the competitive agglomeration term while giving a larger weight to the error term as the number of iterations increases.

The piece-wise convex spatial-spectral unmixing algorithm using possibilistic and fuzzy clustering [53] associates both fuzzy and typicality membership values with each pixel to learn sets of endmembers (i.e. convex regions). This algorithm integrates concepts from the Fuzzy Local Information C-Means (FLICM) method [73], and the Possibilistic Fuzzy C-Means (PFCM) method [75] into the P-COMMEND objective function. Using the same notation as for P-COMMEND, the

objective function is defined as:

$$J = \sum_{i=1}^C \left[\sum_{j=1}^N \left(au_{ij}^m [(\mathbf{x}_j - \mathbf{p}_{ij}\mathbf{E}_i)(\mathbf{x}_j - \mathbf{p}_{ij}\mathbf{E}_i)^T + G_{ij}] + bt_{ij}^n (\mathbf{x}_j - \mathbf{p}_{ij}\mathbf{E}_i)(\mathbf{x}_j - \mathbf{p}_{ij}\mathbf{E}_i)^T \right) + \alpha \sum_{k=1}^{M-1} \sum_{l=k+1}^M (\mathbf{e}_{ik} - \mathbf{e}_{il})(\mathbf{e}_{ik} - \mathbf{e}_{il})^T + \gamma_i \sum_{j=1}^N (1 - t_{ij})^n \right], \quad (2.23)$$

subject to the constraints in (2.7), (2.8), and

$$t_{ij} \in [0, 1], \forall i, j, \text{ and } \sum_{i=1}^C t_{ij} \leq 1, \forall j. \quad (2.24)$$

In (2.23), $a, b, n > 1$ and γ_i are positive constants.

In [53], it was shown that the optimization of (2.23) yields an update equation for the proportions that is the same as the P-COMMEND algorithm (equation (2.10)). Similarly, the membership update equation is as in (2.17). The endmembers need to be updated using

$$\mathbf{E}_i = \left[\alpha(M\mathbf{I}_{M \times M} - \mathbf{1}_{M \times M}) + \sum_{j=1}^N (au_{ij}^m + bt_{ij}^n) \mathbf{p}_{ij}^T \mathbf{p}_{ij} \right]^{-1} \left[\sum_{j=1}^N (au_{ij}^m + bt_{ij}^n) \mathbf{p}_{ij}^T \mathbf{x}_j \right], \quad (2.25)$$

and the update equation for the typicality is

$$t_{ij} = \frac{1}{1 + \left(\frac{b}{\gamma_i} \|\mathbf{x}_j - \mathbf{p}_{ij}\mathbf{E}_i\|_2^2 \right)^{\left(\frac{1}{n-1} \right)}}. \quad (2.26)$$

In the above equation, the γ_i values are set to be the mean of the $\|\mathbf{x}_j - \mathbf{p}_{ij}\mathbf{E}_i\|_2^2$ values for all the pixels in the associated convex region. Therefore, γ_i are updated in every iteration.

Although the P-COMMEND algorithm and its variations allow the detection of multiple sets of endmembers for a hyperspectral image, they do not take into account the distribution of the data in the spectral space. In fact, the estimated endmember sets are the result of the piece-wise convex and spatial representation of the data. Thus, they do not have a spectral meaning.

In the next chapter, we introduce our approach for hyperspectral unmixing which alleviates this shortcoming by taking into account the distribution of the data in the spectral space while unmixing.

2.2 Target detection using hyperspectral imaging

Besides material identification through spectral unmixing, remote sensing using hyperspectral images is used for a variety of other civilian and military applications. These can be categorized into 4 main tasks [76]:

- *Target detection*: searching the pixels of a hyperspectral image for “rare” pixels with known spectral signatures,
- *Anomaly detection*: searching the pixels of a hyperspectral image for pixels whose spectra significantly differ from the local background,
- *Change detection*: finding the “significant” spectral changes over time between two hyperspectral images of the same geographic region,
- *Classification*: assigning a label to each pixel of a hyperspectral image.

In this work, we are interested in target detection. Investigating the other applications is a potential future work. In the following, we review the use of hyperspectral imaging for target detection.

2.2.1 Definitions

A target is defined as any object or material being sought in a hyperspectral image. Targets that occupy multiple pixels are referred to as *multipixel* or *resolved* targets. The detection of resolved targets can exploit spatial and spectral properties of the image. *Subpixel* targets, on the other hand, occupy only a part of the pixel. The remaining part is filled with one or more materials, which are collectively referred to as *background* [28]. The detection of subpixel targets exploits spectral properties only. Hyperspectral imaging provides an invaluable tool for subpixel target detection, as it can identify and distinguish between different materials having different spectral signatures.

Typically, the number of targets in a scene is too small to support the estimation of the statistical properties of the target class, and pattern classification algorithms that require such information are hence not applicable.

Instead, the design and evaluation of detection algorithms can be achieved using an area of statistics known as binary hypothesis testing. In particular, the likelihood ratio (LR) test is used for many target detectors. It minimizes the risk associated with incorrect decisions, and leads to detectors that are optimum for a wide range of performance criteria, including the maximization of separation between target and background spectra [28].

Given an observed spectrum $\mathbf{x}$, we want to choose between two competing hypotheses:

$$\begin{aligned} \mathbf{H}_0: & \text{Target absent} \\ \mathbf{H}_1: & \text{Target present} \end{aligned} \tag{2.27}$$

The likelihood ratio is the ratio of the conditional probability density functions under the two hypotheses:

$$\Lambda(\mathbf{x}) = \frac{p(\mathbf{x}|\mathbf{H}_1)}{p(\mathbf{x}|\mathbf{H}_0)}. \quad (2.28)$$

If $\Lambda(\mathbf{x})$ exceeds a certain threshold η , then $\mathbf{H}_1$ is selected as true. This suggests the detector structure shown in figure 2.1. The test pixel $\mathbf{x}$ gets mapped into a scalar $y = \Lambda(\mathbf{x})$, referred to as *detection*

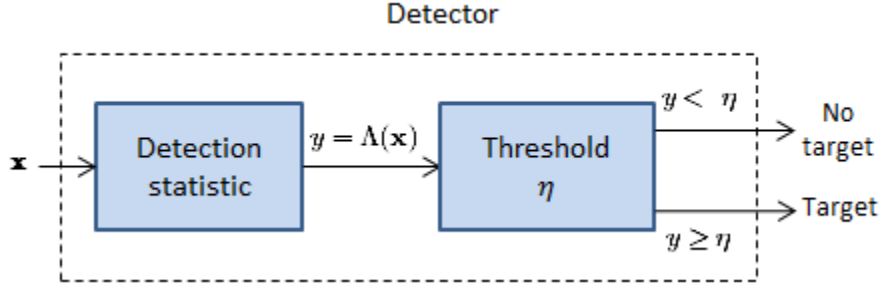


Figure 2.1: Target detector structure.

statistic, which is compared to η to decide whether a target is present.

The choice of η controls the number of correct detections and the number of detection errors (target misses and false alarms). This is illustrated in figure 2.2.

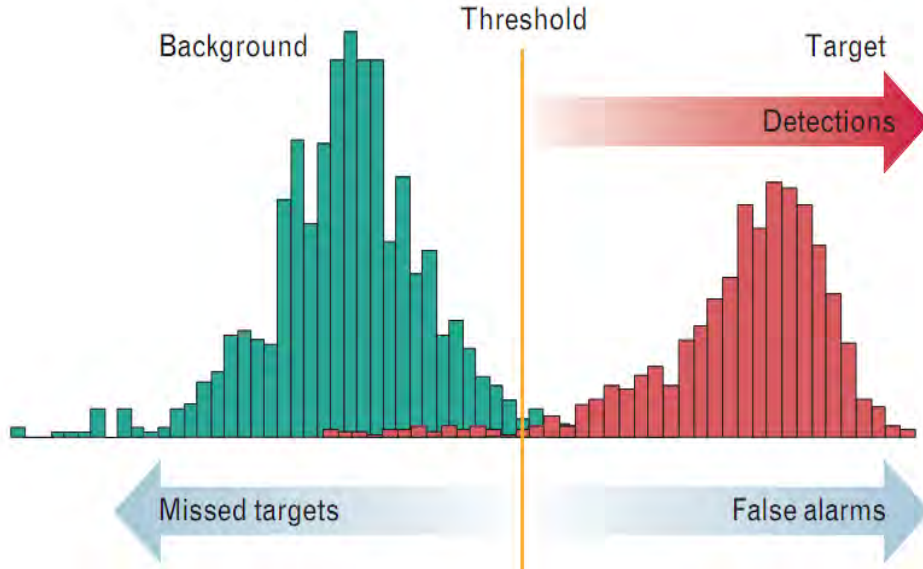


Figure 2.2: Threshold selection trade-offs.¹

There is a compromise between choosing a low threshold to increase the probability of detection P_D , and a high threshold to keep the probability of false alarms P_{FA} low. The trade-off between P_D

¹This image was taken from [28].

and P_{FA} is described by the receiver operating characteristic (ROC) curve, which plots P_D versus P_{FA} as a function of all possible values of the threshold η . Therefore, ROC curves provide a means to evaluate the detector performance or compare detectors independently of threshold selection.

In subpixel target detection, the spectrum of the target is mixed with the spectra of the background. Finding a good model for the background is key for detection [28]. Different models lead to different detectors. If a statistical distribution is used to model the background, the model is said to be *unstructured* [28]. In contrast, when a subspace is used to model the background, the model is said to be *structured* [28]. In the following, we describe few subpixel target detectors based on each kind of the background models, with more emphasis on the structured background model.

2.2.2 Target detection using unstructured background models

In this case, the background is modeled using a multivariate normal distribution with zero mean and covariance matrix $\mathbf{\Gamma}$. This assumes that the background is homogeneous. Due to the zero mean background assumption, the sample mean of the image should be removed from all image pixels and target spectra beforehand.

A well known unstructured background target detector is the adaptive coherence/cosine estimator (ACE) detector [77]:

$$T_{ACE}(\mathbf{x}) = \frac{(\mathbf{s}_t \hat{\mathbf{\Gamma}}^{-1} \mathbf{x}^T)^2}{(\mathbf{s}_t \hat{\mathbf{\Gamma}}^{-1} \mathbf{s}_t^T)(\mathbf{x} \hat{\mathbf{\Gamma}}^{-1} \mathbf{x}^T)}, \quad (2.29)$$

where $\mathbf{x}$ is the spectral signature of the test pixel, $\mathbf{s}_t$ is the spectrum of the target, and $\hat{\mathbf{\Gamma}}$ is the maximum likelihood estimate of the covariance matrix $\mathbf{\Gamma}$.

If we consider the square-root decomposition of $\hat{\mathbf{\Gamma}} = \hat{\mathbf{\Gamma}}^{\frac{1}{2}} \hat{\mathbf{\Gamma}}^{\frac{1}{2}}$, the transformation $\mathbf{z}^T = \hat{\mathbf{\Gamma}}^{-\frac{1}{2}} \mathbf{x}^T$ is called adaptive whitening. The ACE detector can be expressed as:

$$T_{ACE}(\mathbf{x}) = \frac{(\tilde{\mathbf{s}}_t \mathbf{z}^T)^2}{(\tilde{\mathbf{s}}_t \tilde{\mathbf{s}}_t^T)(\mathbf{z} \mathbf{z}^T)} = \frac{(\tilde{\mathbf{s}}_t \mathbf{z}^T)^2}{\|\tilde{\mathbf{s}}_t\|^2 \|\mathbf{z}\|^2} = \cos^2 \theta \quad (2.30)$$

where $\tilde{\mathbf{s}}_t^T = \hat{\mathbf{\Gamma}}^{-\frac{1}{2}} \mathbf{s}_t^T$ is the whitened target signature. This shows that, in the whitened coordinate space, $T_{ACE}(\mathbf{x})$ is equal to the cosine square of the angle θ between the test pixel and the target.

A similar unstructured target detector was developed by E. J. Kelly [78]:

$$T_{Kelly}(\mathbf{x}) = \frac{(\mathbf{s}_t \hat{\mathbf{\Gamma}}^{-1} \mathbf{x}^T)^2}{(\mathbf{s}_t \hat{\mathbf{\Gamma}}^{-1} \mathbf{s}_t^T)(N + \mathbf{x} \hat{\mathbf{\Gamma}}^{-1} \mathbf{x}^T)}, \quad (2.31)$$

where N denotes the number of pixels in the image.

2.2.3 Target detection using structured background models

If a subspace is used to model the background, the detection problem involves choosing between the following competing hypotheses:

$$\begin{aligned} \mathbf{H}_0: \quad \mathbf{x} &= \mathbf{p}_b \mathbf{E} + \mathbf{w} && (\text{Target absent}) \\ \mathbf{H}_1: \quad \mathbf{x} &= p_t \mathbf{s}_t + \mathbf{p}_b \mathbf{E} + \mathbf{w} && (\text{Target present}) \end{aligned} \quad (2.32)$$

where $\mathbf{s}_t$ denotes the spectrum of the target (specified by the user), the matrix $\mathbf{E}$ denotes the background subspace (estimated from the data), the vector $\mathbf{p}_b$ and the scalar p_t represent the proportions of the background and the target in the test pixel $\mathbf{x}$, and $\mathbf{w}$ is an error term assumed to be normally distributed with zero mean and covariance $\mathbf{\Gamma}_w = \sigma_w^2 \mathbf{I}$.

2.2.3.1 Adaptive Matched Subspace Detector (AMSD)

In practice, the parameters $\mathbf{p}_b$, p_t and σ_w^2 are unknown. Using their maximum likelihood estimates (MLE) results in the generalized likelihood ratio (GLR) [76]:

$$\text{GLR}(\mathbf{x}) \triangleq \left[\frac{\mathbf{x} \mathbf{P}_b^\perp \mathbf{x}^T}{\mathbf{x} \mathbf{P}_S^\perp \mathbf{x}^T} \right]^{d/2}, \quad (2.33)$$

where d is the dimension of $\mathbf{x}$,

$$\mathbf{P}_b^\perp = \mathbf{I} - \mathbf{E}^T (\mathbf{E} \mathbf{E}^T)^{-1} \mathbf{E}, \quad (2.34)$$

is the background orthogonal projection error matrix, and $\mathbf{P}_S^\perp$ is the error matrix of the orthogonal projection on the composite target and background space $\mathbf{S} = [\mathbf{s}_t; \mathbf{E}]$, computed by replacing $\mathbf{E}$ with $\mathbf{S}$ in (2.34). Notice that σ_w^2 is not required for the computation of GLR because it cancels out.

The GLR in equation (2.33) is used to design a detection statistic called Adaptive Matched Subspace Detector (AMSD) [76]:

$$T_{\text{AMSD}}(\mathbf{x}) = \frac{\mathbf{x} \mathbf{P}_b^\perp \mathbf{x}^T}{\mathbf{x} \mathbf{P}_S^\perp \mathbf{x}^T}. \quad (2.35)$$

In (2.35), the numerator is proportional to the residual of the projection of $\mathbf{x}$ on the background space, while the denominator is proportional to the residual of the projection of $\mathbf{x}$ on the combined target and background space. Thus, the ratio becomes large when the test pixel representation in the combined target and background space $\mathbf{S}$ is better than its representation in the background space $\mathbf{E}$. In other words, the larger the ratio in (2.35) the more likely that pixel $\mathbf{x}$ contains the target.

2.2.3.2 Orthogonal Subspace Projection (OSP) Detector

The proportion p_t of the target $\mathbf{s}_t$ in pixel $\mathbf{x}$ can be estimated as [76]:

$$p_t = \frac{\mathbf{s}_t \mathbf{P}_{\mathbf{b}}^\perp \mathbf{x}^T}{\mathbf{s}_t \mathbf{P}_{\mathbf{b}}^\perp \mathbf{s}_t^T}, \quad (2.36)$$

where $\mathbf{P}_{\mathbf{b}}^\perp$ is the same as in equation (2.34). The numerator in (2.36) has been proposed as a detection statistic under the Orthogonal Subspace Projection (OSP) name [76]:

$$T_{\text{OSP}}(\mathbf{x}) = \mathbf{s}_t \mathbf{P}_{\mathbf{b}}^\perp \mathbf{x}^T. \quad (2.37)$$

The operation $\mathbf{P}_{\mathbf{b}}^\perp \mathbf{x}^T$ removes from $\mathbf{x}^T$ the part which belongs to the background $\mathbf{E}$. The residual is then projected on the target $\mathbf{s}_t$. The larger this quantity is, the more likely the test pixel $\mathbf{x}$ contains the target.

2.2.3.3 Hybrid Subspace Detector (HSD)

Unlike OSP and AMSD, which perform an unconstrained least squares estimate of the abundances using orthogonal projection [76], the Hybrid Subspace Detector (HSD) estimates these abundances using the fully constrained least squares algorithm [79]. This assures the satisfaction of the non-negativity and sum-to-one constraints and provides a meaning to the abundances [79]. HSD uses the following detection statistic:

$$T_{\text{HSD}}(\mathbf{x}) = \frac{(\mathbf{x} - \mathbf{p}_{\mathbf{b}} \mathbf{E}) \mathbf{\Gamma}^{-1} (\mathbf{x} - \mathbf{p}_{\mathbf{b}} \mathbf{E})^T}{(\mathbf{x} - \mathbf{p}_{\mathbf{s}} \mathbf{S}) \mathbf{\Gamma}^{-1} (\mathbf{x} - \mathbf{p}_{\mathbf{s}} \mathbf{S})^T}, \quad (2.38)$$

where $\mathbf{E}$ is the background subspace matrix, $\mathbf{p}_{\mathbf{b}}$ is the estimate of the abundances of the test pixel in $\mathbf{E}$, $\mathbf{S} = [\mathbf{s}_t; \mathbf{E}]$ is the composite target and background space, $\mathbf{p}_{\mathbf{s}}$ is the estimate of the abundances of the test pixel in $\mathbf{S}$, and $\mathbf{\Gamma}^{-1}$ denotes the inverse of the covariance matrix of the background.

The OSP, AMSD, HSD, and other structured target detectors use global methods to model the background subspace. These methods include eigenvector decomposition of the data correlation matrix [76], and global unmixing algorithms such as ICE [61], MVSA [70], and NFINDR [60]. These global methods may not provide a good description of the hyperspectral data when the scene contains multiple distinct regions that do not share common materials. In this case, unmixing methods that can learn multiple background models, that correspond to regions with different characteristics, are more appropriate. Moreover, most global unmixing methods face the challenge of target leakage into the background, which is the contribution of the target pixels to the background model, due to their presence in the scene.

CHAPTER 3

CONTEXT DEPENDENT SPECTRAL UNMIXING

In this chapter, we introduce our approach, called Context Dependent Spectral Unmixing (CDSU), to hyperspectral unmixing. It finds multiple sets of endmembers and takes into account the distribution of the data in the spectral space while unmixing. First, we present the motivations behind this approach. Then, we propose a novel objective function and derive the necessary conditions to optimize it.

3.1 Motivations

To motivate our proposed approach, we present three examples of data sets and analyze the performance of the previously presented ICE and P-COMMEND algorithms on them. The data sets are chosen to be 2-dimensional for easier visualization.

3.1.1 Example 1

In this example, we consider a 2-dimensional data set of 250 points generated using the convex geometry model in (1.1). A zero error term is considered. One set of three endmembers is used:

$$\mathbf{E}_1 = \begin{pmatrix} 1 & 3 \\ 2 & 3 \\ 1.5 & 5 \end{pmatrix}. \quad (3.1)$$

The proportions are chosen randomly from a standard uniform distribution and are normalized to sum to one. The data is illustrated in figure 3.1(a). Blue points denote the data vectors and green points denote the true endmembers used to generate the data set.

The result of the ICE algorithm on this data set, using $M = 3$ and $\mu = 0.0009$, is shown in figure 3.1(b). The detected endmembers are shown in red. The result of the P-COMMEND algorithm, using $C = 1$, $M = 3$, $m = 2$, and $\alpha = 0.01$, is shown in figure 3.1(c). The detected endmembers are shown in red.

As it can be seen, for this simple example, both ICE and P-COMMEND succeeded in identifying

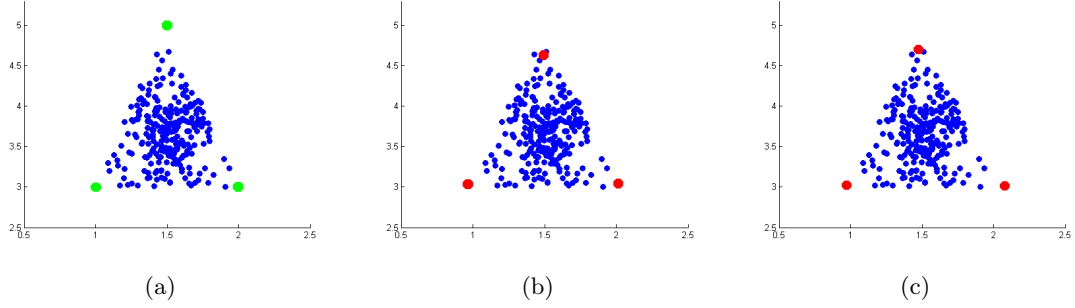


Figure 3.1: Example 1: (a) Synthetic two-dimensional data with one convex hull. (b) Endmembers detected by the ICE algorithm. (c) Endmembers detected by the P-COMMEND algorithm.

endmembers at the vertices of the convex hull enclosing the data points providing a tight fit around them.

3.1.2 Example 2

In this example, we consider a 2-dimensional data set of 500 points generated using the convex geometry model in (1.1) with a zero error term. For this example, we use two sets of endmembers. Each set has three endmembers and was used to generate 250 points. The first set is the same as in (3.1), and the second one is:

$$\mathbf{E}_2 = \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 1.5 & 2.5 \end{pmatrix}. \quad (3.2)$$

Similar to example 1, the proportions are chosen randomly from a standard uniform distribution and are normalized to sum to one. This data is plotted in figure 3.2(a), where blue points denote the data vectors and green points denote the true endmembers used to generate the data.

The results of the ICE algorithm on this data set, using $M = 3$ and $M = 6$, are shown in figures 3.2(b) and (c) respectively. A value of $\mu = 0.0009$ was used for both cases. The detected endmembers are shown in red. As it can be seen, the ICE algorithm failed to identify correct endmembers for this data set. This is due to the fact that ICE assumes a single convex region for the entire data, and this assumption is not valid for this data set. Consequently, the detected erroneous endmembers and the associated abundances will have a negative impact on any further processing or analysis based on them.

The result of the P-COMMEND algorithm, using $C = 2$, $M = 3$, $m = 2$, and $\alpha = 0.01$, is shown in figure 3.2(d). The detected endmembers are shown in red for one cluster and in green for the other

one. As it can be seen, P-COMMEND succeeded in identifying endmember sets at the vertices of

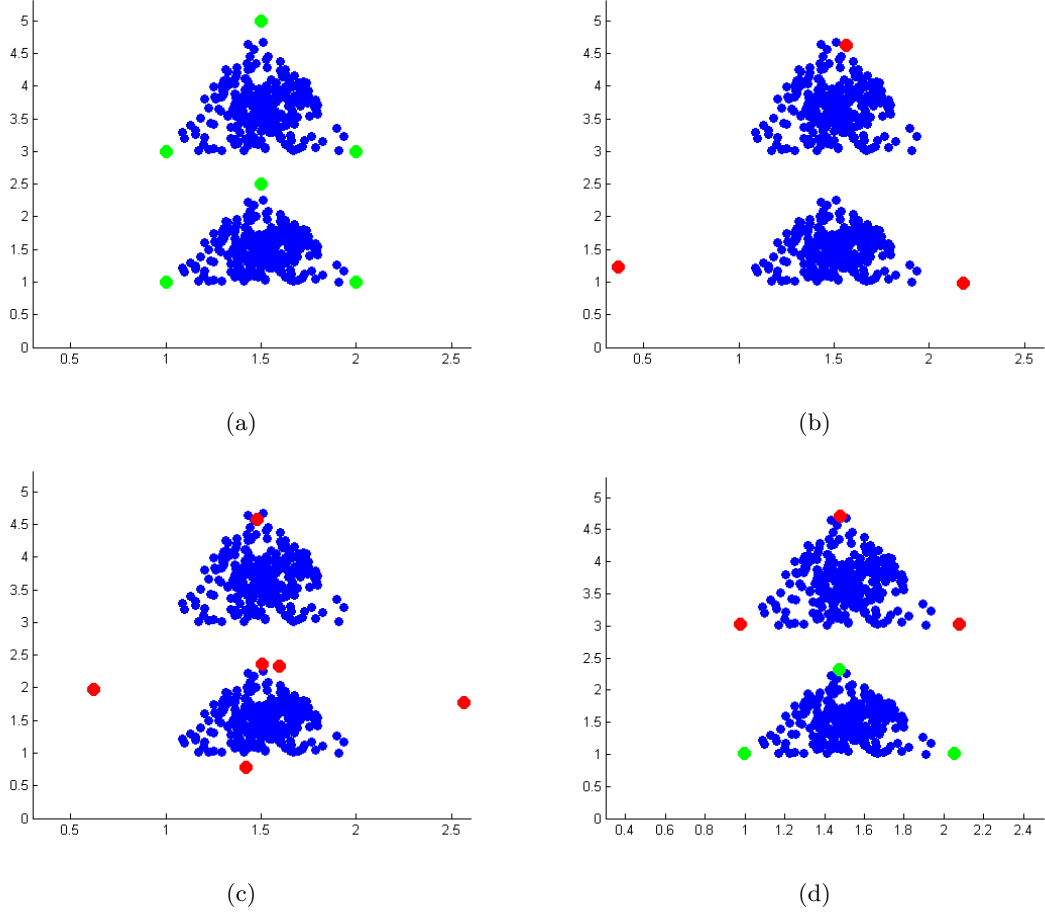


Figure 3.2: Example 2: (a) Synthetic two-dimensional data with two convex hulls. (b) Endmembers detected by the ICE algorithm with $M = 3$. (c) Endmembers detected by the ICE algorithm with $M = 6$. (d) Endmembers detected by the P-COMMEND algorithm.

the convex hulls enclosing the data points and thus, providing a tight fit around them.

3.1.3 Example 3

In this example, we consider a 2-dimensional data set of 1000 points, arranged so that they form three clusters. These are illustrated in figure 3.3(a), where clusters are represented using different colors.

We run the P-COMMEND algorithm on this data set, using $C = 3$, $M = 3$, $m = 2$, $\alpha = 0.1$ ¹ and using the true cluster assignment in figure 3.3(a) as the initial membership values $\mathbf{U}$. We use this initialization to illustrate the fact that the results of P-COMMEND are not due to convergence to local minima. As for the initialization of the endmember sets $\mathbf{E}_i$, we pick three random points

¹Other values of α can lead to better results

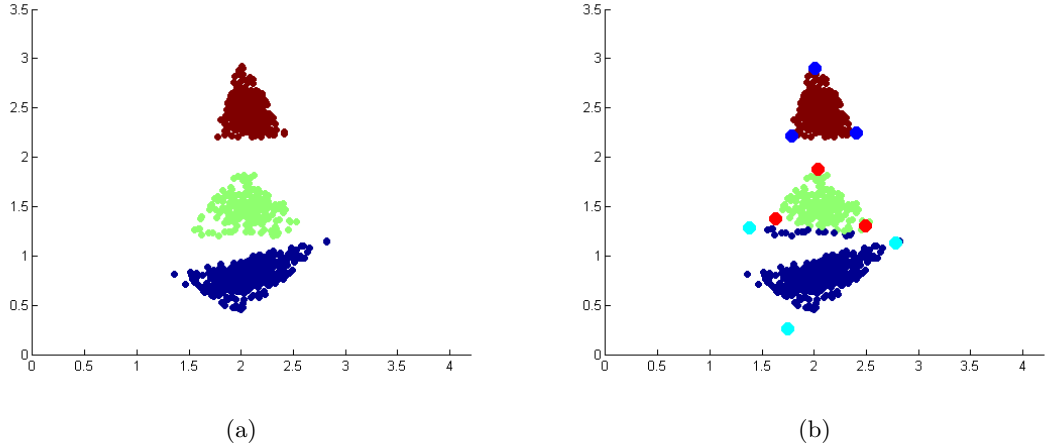


Figure 3.3: Example 3: (a) Synthetic two-dimensional data with three convex hulls (colors represent true cluster assignments). (b) Result of the P-COMMEND algorithm.

from each cluster. The result after convergence is illustrated in figure 3.3(b), where clusters are represented using different colors, and the endmembers for each cluster are represented using bolder points. The assignment of a point to a cluster is based on its highest membership value. As it can be seen, P-COMMEND diverged from the initial cluster assignments and led to clusters that do not match the distribution of the data. These clusters resulted from the convex geometry of the estimated endmembers. Consequently, the detected erroneous endmembers and the associated abundances will have a negative impact on any further processing or analysis based on them.

Based on the above three examples, the motivation behind our approach is two-fold. First, there is a need for an unmixing algorithm that is able to find multiple sets of endmembers for an input hyperspectral data and not only a unique set. In general, multiple sets would provide a better description of the hyperspectral scene. This is because hyperspectral images tend to be large and may contain multiple distinct regions that do not share common materials, and hence need multiple sets of endmembers to fully describe it. Second, there is a need to find sets of endmembers that represent semantically meaningful regions of the hyperspectral image. In other words, we need to take into account the distribution of the data in the spectral space while unmixing.

The idea of the proposed approach is based on context-based processing which has been shown to be important in many signal processing applications [80–82]. In [80, 81], the authors presented a fusion method, called Context-Dependent Multi-Sensor Fusion (CDMSF), which fuses the results of multiple landmine detection algorithms that use different types of features, different classification methods, and different sensors. The approach was motivated by the fact that the relative performance of different detectors can vary significantly depending on the sensor, mine

type, geographical site, soil and weather conditions, and burial depth.

In [82], the authors point out that CDMSF [80,81] treats the partitioning of the feature space and the selection of local expert classifiers as two independent processes performed sequentially. They claim that these two tasks are not independent, and their optimization should be combined. To alleviate this limitation, they proposed a generic framework for context-dependent fusion, called Context Extraction for Local Fusion (CELF) [82], that jointly optimizes the partitioning of the feature space and the fusion of the classifiers. They defined a novel objective function that, when optimized, produces contexts via unsupervised clustering while simultaneously providing optimal fusion parameters for each context. The authors have shown that CELF can identify meaningful and coherent clusters where different expert algorithms can be identified. Their experiments have also indicated that their proposed fusion approach outperforms all individual detectors.

Motivated by the above context dependent processing, our proposed unmixing approach adapts the unmixing to different regions of the spectral space.

3.2 Context Dependent Spectral Unmixing

The Context Dependent Spectral Unmixing (CDSU) algorithm is a local approach that adapts the unmixing to different regions of the spectral space. It is based on a novel objective function that combines context identification and spectral unmixing into a joint function. This objective function models contexts as compact clusters and uses the linear mixing model as the basis for unmixing. The unmixing provides optimal endmembers and abundances for each context.

In the following, we assume that we have N spectral signatures, $\mathbf{X} = \{\mathbf{x}_j \in \mathbb{R}^d, j = 1, \dots, N\}$, obtained from a hyperspectral scene, each having d spectral bands. CDSU combines context identification and spectral unmixing into a joint objective function. It takes into account the distribution of the data in the spectral space when finding the regions or contexts, and not only the convex geometry of the spectral unmixing as in [49]. CDSU achieves these two tasks by minimizing the following objective function:

$$J = \sum_{i=1}^C \sum_{j=1}^N u_{ij}^m (\mathbf{x}_j - \mathbf{c}_i)(\mathbf{x}_j - \mathbf{c}_i)^T + \alpha \sum_{i=1}^C \left[\sum_{j=1}^N u_{ij}^m (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)(\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T + \beta_i \sum_{k=1}^{M-1} \sum_{l=k+1}^M (\mathbf{e}_{ik} - \mathbf{e}_{il})(\mathbf{e}_{ik} - \mathbf{e}_{il})^T \right] \quad (3.3)$$

subject to

$$u_{ij} \in [0, 1], \forall i, j, \sum_{i=1}^C u_{ij} = 1, \forall j, \quad (3.4)$$

and

$$\mathbf{p}_{ij} \geq 0, \text{ and } \mathbf{1}_{1 \times M} \mathbf{p}_{ij}^T = 1, \forall i, j. \quad (3.5)$$

In the above, $\mathbf{x}_j$ is a $1 \times d$ vector representing the j^{th} pixel spectrum, C is a user-specified constant that represents the number of contexts to be extracted, and M is the number of endmembers for each context. In (3.3), $\mathbf{p}_{ij}$ is a $1 \times M$ vector representing the proportion values for pixel j with respect to the i^{th} context, and $\mathbf{E}_i$ is a $M \times d$ matrix such that each row of $\mathbf{E}_i$, $\mathbf{e}_{ik}$, is the $1 \times d$ vector representing the k^{th} endmember in the i^{th} context. The membership values, u_{ij} , indicate the fuzzy degree to which the j^{th} sample belongs to the i^{th} context. Finally, α and $\boldsymbol{\beta} = [\beta_1, \dots, \beta_C]$ are positive constants used to balance the three terms of the objective function.

The first term in (3.3) is an unsupervised learning component. It is the sum of intra-cluster distances and is the objective function used in the Fuzzy C-Means (FCM) algorithm [72]. It seeks to partition the N samples into C clusters, and represent each cluster by a center $\mathbf{c}_i$. Each sample $\mathbf{x}_j$ will be assigned to each cluster i with a membership degree u_{ij} . In this term, $m \in (1, +\infty)$ is used to control the degree of fuzziness [72]. The second and third terms in (3.3) represent the spectral unmixing component. This component attempts to learn cluster-dependent endmembers and abundances. The second term is the residual sum of squares (RSS) between actual pixels and their estimates from the endmembers and abundances. The third term is the sum of squared distances (SSD) between each pair of endmembers in an endmember set, representing an approximation of the volume enclosed by these endmembers. Note that these last two terms represent the objective function of the P-COMMEND algorithm [49] with the exception that we use C balancing parameters β_i ($i = 1, \dots, C$), one for each context, unlike P-COMMEND which uses a unique balancing parameter α . This allows for more flexibility since the simplexes in all contexts may not be of equal volumes.

When all terms are combined and the parameters α and $\boldsymbol{\beta}$ are chosen properly, the algorithm seeks to partition the data into compact clusters while learning the endmembers and abundances for each cluster.

Using the matrix notation

$$\sum_{k=1}^{M-1} \sum_{l=k+1}^M (\mathbf{e}_{ik} - \mathbf{e}_{il})(\mathbf{e}_{ik} - \mathbf{e}_{il})^T = M \text{trace}(\mathbf{E}_i \mathbf{E}_i^T) - \mathbf{1}_{1 \times M} \mathbf{E}_i \mathbf{E}_i^T \mathbf{1}_{M \times 1}, \quad (3.6)$$

the objective function in (3.3) can be rewritten as

$$J(\mathbf{E}_i, \mathbf{P}_i, \mathbf{U}, \mathbf{c}_i) = \sum_{i=1}^C \sum_{j=1}^N u_{ij}^m (\mathbf{x}_j - \mathbf{c}_i)(\mathbf{x}_j - \mathbf{c}_i)^T + \alpha \sum_{i=1}^C \left[\sum_{j=1}^N u_{ij}^m (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)(\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T + \beta_i (M \text{trace}(\mathbf{E}_i \mathbf{E}_i^T) - \mathbf{1}_{1 \times M} \mathbf{E}_i \mathbf{E}_i^T \mathbf{1}_{M \times 1}) \right]. \quad (3.7)$$

The goal is to identify the optimal endmember sets $\mathbf{E}_i = \{\mathbf{e}_{ik}, k = 1, \dots, M\}$, $i = 1, \dots, C$, the proportion sets $\mathbf{P}_i = \{\mathbf{p}_{ij}, j = 1, \dots, N\}$, $i = 1, \dots, C$, the memberships $\mathbf{U} = [u_{ij}]_{i=1, \dots, C; j=1, \dots, N}$ and the centers $\mathbf{c}_i$, $i = 1, \dots, C$, that minimize (3.7) subject to (3.4) and (3.5).

Lagrange multipliers [83] is a powerful tool for solving optimization problems without the need to explicitly solve the constraints and use them to eliminate extra variables. For the CDSU optimization, we incorporate the constraints in (3.4) and (3.5) into the objective function in (3.7) using Lagrange multipliers and obtain

$$L = \sum_{i=1}^C \sum_{j=1}^N u_{ij}^m (\mathbf{x}_j - \mathbf{c}_i)(\mathbf{x}_j - \mathbf{c}_i)^T + \alpha \sum_{i=1}^C \left[\sum_{j=1}^N u_{ij}^m (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)(\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T + \beta_i (M \text{trace}(\mathbf{E}_i \mathbf{E}_i^T) - \mathbf{1}_{1 \times M} \mathbf{E}_i \mathbf{E}_i^T \mathbf{1}_{M \times 1}) \right] - \sum_{j=1}^N \lambda_j \left(\sum_{i=1}^C u_{ij} - 1 \right) - \sum_{i=1}^C \sum_{j=1}^N \gamma_{ij} (\mathbf{1}_{1 \times M} \mathbf{p}_{ij}^T - 1) - \sum_{i=1}^C \sum_{j=1}^N \xi_{ij} \mathbf{p}_{ij}^T, \quad (3.8)$$

where $\boldsymbol{\Lambda} = [\lambda_1, \dots, \lambda_N]$ is a vector of Lagrange multipliers corresponding to the N constraints on the memberships u_{ij} in (3.4), $\boldsymbol{\Gamma} = [\gamma_{11}, \dots, \gamma_{CN}]$, and $\boldsymbol{\Xi} = [\xi_{11}, \dots, \xi_{CN}]$ are vectors of Lagrange multipliers corresponding to the $C \times N$ constraints on the proportions $\mathbf{p}_{ij}$ in (3.5).

The first-order necessary conditions theorem in [83] states that a local minimizer of the objective function J necessarily satisfies the Karush-Kuhn-Tucker (KKT) conditions. The first-order derivative of the Lagrangian L with respect to the minimizing variable evaluated at the local minimizer is necessary equal to zero. Furthermore, the second-order sufficient conditions theorem in [83] states that for a feasible variable, satisfying the KKT conditions, to be a local minimizer of J , it is sufficient that the second-order derivative of the Lagrangian L with respect to this variable is positive definite.

Theorem 3.2.1, the proof of which is given in Appendix A, gives the update equations for the endmember sets, the proportion sets, the memberships, and the centers that minimize J .

Theorem 3.2.1. *The first and second order conditions yield the following local minimizers of J :*

$$\mathbf{E}_i = \left[\beta_i (M \mathbf{I}_{M \times M} - \mathbf{1}_{M \times M}) + \sum_{j=1}^N u_{ij}^m \mathbf{p}_{ij}^T \mathbf{p}_{ij} \right]^{-1} \left[\sum_{j=1}^N u_{ij}^m \mathbf{p}_{ij}^T \mathbf{x}_j \right], \quad (3.9)$$

$$\mathbf{p}_{ij}^T = \max \left(\left[\mathbf{E}_i \mathbf{E}_i^T \right]^{-1} \left[\mathbf{E}_i \mathbf{x}_j^T + \frac{1 - \mathbf{1}_{1 \times M} (\mathbf{E}_i \mathbf{E}_i^T)^{-1} \mathbf{E}_i \mathbf{x}_j^T}{\mathbf{1}_{1 \times M} (\mathbf{E}_i \mathbf{E}_i^T)^{-1} \mathbf{1}_{M \times 1}} \mathbf{1}_{M \times 1} \right], 0 \right), \quad (3.10)$$

$$u_{ij} = \frac{[(\mathbf{x}_j - \mathbf{c}_i)(\mathbf{x}_j - \mathbf{c}_i)^T + \alpha(\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)(\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T]^{\frac{1}{1-m}}}{\sum_{q=1}^C [(\mathbf{x}_j - \mathbf{c}_q)(\mathbf{x}_j - \mathbf{c}_q)^T + \alpha(\mathbf{x}_j - \mathbf{p}_{qj} \mathbf{E}_q)(\mathbf{x}_j - \mathbf{p}_{qj} \mathbf{E}_q)^T]^{\frac{1}{1-m}}}, \quad (3.11)$$

and

$$\mathbf{c}_i = \frac{\sum_{j=1}^N u_{ij}^m \mathbf{x}_j}{\sum_{j=1}^N u_{ij}^m}. \quad (3.12)$$

By examining (3.11), we notice that a pixel j will have a high membership in cluster i if: (i) its spectra is close to the centroid, $\mathbf{c}_i$, of that cluster in the feature space; and (ii) it fits the endmember model of that cluster (i.e. small error of fit).

CDSU is an iterative algorithm that involves successive updates of the endmember proportion sets $\mathbf{P}_i$, the memberships $\mathbf{U}$, the endmember sets $\mathbf{E}_i$, and the clusters' centers $\mathbf{c}_i$. It is summarized in Algorithm 3.1.

Algorithm 3.1 Context Dependent Spectral Unmixing (CDSU)

Inputs: $\mathbf{X}$: the data points ($N \times d$).

C : the number of contexts.

M : the number of endmembers for each context.

m : the fuzzifier, $m \in (1, +\infty)$.

α : the weight of the second term in the objective function.

β : the vector of weights in the second term of the objective function ($1 \times C$).

Outputs: $\mathbf{U}$: the fuzzy membership matrix of the data samples.

$\mathbf{c}_i$: the cluster centers.

$\mathbf{E}_i$: the sets of endmembers in all clusters.

$\mathbf{P}_i$: the sets of proportions in all clusters.

Initialize $\mathbf{c}_i$ and $\mathbf{E}_i$.

repeat

 Update $\mathbf{P}_i$ using (3.10).

 Update $\mathbf{U}$ using (3.11).

 Update $\mathbf{E}_i$ using (3.9).

 Update $\mathbf{c}_i$ using (3.12).

until parameters do not change significantly

return $\mathbf{U}$, $\mathbf{c}_i$, $\mathbf{E}_i$, $\mathbf{P}_i$

The CDSU algorithm is of linear time per iteration. It runs in $\mathcal{O}(C \times M \times N)$ time per iteration, where C is the number of contexts, M is the number of endmembers per context, and N is the number of data points. In the current implementation of the algorithm, the centers are initialized using the Fuzzy C-Means algorithm [72] (which is, in turn, randomly initialized), and the endmember sets are initialized using the Minimum Volume Simplex Analysis (MVSA) algorithm [70].

Furthermore, the convergence of the algorithm is checked by comparing the values of the objective function from successive iterations. If the difference is below some threshold, the algorithm is stopped.

3.3 Context Dependent Spectral Unmixing Using the Mahalanobis Distance

The current CDSU algorithm uses the Euclidean distance in the clustering component (first term in (3.7)). As a result, it is restricted to identifying spherical clusters. Moreover, clustering in a high dimensional space, as is the case of hyperspectral data, is a challenging task. Features might not be equally important and some of them can be highly correlated. This may result in non-spherical clusters.

In this section, we extend CDSU to use the Mahalanobis distance instead of the Euclidean distance. This would allow more flexibility in the shape of the clusters instead of the traditional spherical shape.

Using the same notation as in Section 3.2, the objective function becomes

$$J_M(\mathbf{E}_i, \mathbf{P}_i, \mathbf{U}, \mathbf{c}_i, \mathbf{A}_i) = \sum_{i=1}^C \sum_{j=1}^N u_{ij}^m (\mathbf{x}_j - \mathbf{c}_i) \mathbf{A}_i (\mathbf{x}_j - \mathbf{c}_i)^T + \alpha \sum_{i=1}^C \left[\sum_{j=1}^N u_{ij}^m (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i) (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T + \beta_i (Mtrace(\mathbf{E}_i \mathbf{E}_i^T) - \mathbf{1}_{1 \times M} \mathbf{E}_i \mathbf{E}_i^T \mathbf{1}_{M \times 1}) \right], \quad (3.13)$$

subject to (3.4), (3.5), and

$$\det(\mathbf{A}_i) = \sigma_i \quad \forall i. \quad (3.14)$$

In (3.13), $\mathbf{A}_i$ is a symmetric and positive definite matrix. Fixing the determinant of the norm matrix $\mathbf{A}_i$ to a positive constant σ_i allows the algorithm to search for a cluster shape that fits the data while preserving the volume of the cluster.

Theorem 3.3.1, the proof of which is given in Appendix B, gives the update equations of the endmember sets, the proportions sets, the memberships, the centers, and the norm matrices that minimize J_M .

Theorem 3.3.1. *Optimizing the objective function in (3.13) using the Lagrange multipliers method leads to the same update equations for the endmember sets, the proportions, and the centers as for CDSU using the Euclidean distance (equations (3.9), (3.10) and (3.12) respectively).*

The update equation for the memberships, u_{ij} , becomes

$$u_{ij} = \frac{\left[(\mathbf{x}_j - \mathbf{c}_i) \mathbf{A}_i (\mathbf{x}_j - \mathbf{c}_i)^T + \alpha (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i) (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T \right]^{\frac{1}{1-m}}}{\sum_{q=1}^C \left[(\mathbf{x}_j - \mathbf{c}_q) \mathbf{A}_q (\mathbf{x}_j - \mathbf{c}_q)^T + \alpha (\mathbf{x}_j - \mathbf{p}_{qj} \mathbf{E}_q) (\mathbf{x}_j - \mathbf{p}_{qj} \mathbf{E}_q)^T \right]^{\frac{1}{1-m}}}. \quad (3.15)$$

Finally, the update equation for the norm matrices, $\mathbf{A}_i$, is

$$\mathbf{A}_i = [\sigma_i \det(\mathbf{C}_i)]^{\frac{1}{d}} \mathbf{C}_i^{-1}, \quad (3.16)$$

where

$$\mathbf{C}_i = \frac{\sum_{j=1}^N u_{ij}^m (\mathbf{x}_j - \mathbf{c}_i)^T (\mathbf{x}_j - \mathbf{c}_i)}{\sum_{j=1}^N u_{ij}^m} \quad (3.17)$$

is the fuzzy covariance matrix of cluster i .

Observing (3.15), we notice that a pixel j will have a high membership in cluster i if: (i) its spectra has a small Mahalanobis distance to the centroid, $\mathbf{c}_i$, of that cluster in the feature space; and (ii) it fits the endmember model of that cluster.

The norm matrices $\mathbf{A}_i$ in (3.16) are a function of the inverse of the fuzzy covariance matrices $\mathbf{C}_i$. This allows CDSU to account for the shape of the data by normalizing by the variances of each dimension.

The resulting Context Dependent Spectral Unmixing using the Mahalanobis distance (CDSU_M) is an iterative algorithm that uses alternating optimization. It involves successive updates of the endmember proportions $\mathbf{p}_{ij}$, the memberships u_{ij} , the endmember sets $\mathbf{E}_i$, the cluster centers $\mathbf{c}_i$, and the norm matrices $\mathbf{A}_i$, until stabilization. It is summarized in Algorithm 3.2.

3.4 Semi-supervised Context Dependent Spectral Unmixing

Spectral unmixing is a challenging, ill-posed, inverse problem that may result in infinitely many solutions, most of which are meaningless. Moreover, context dependent spectral unmixing is based in part on clustering. However, clustering itself is a challenging task, especially in a high dimensional feature space as is the case of hyperspectral data. Many local minima solutions may exist.

To overcome this problem, we propose two semi-supervised algorithms of CDSU_M. The approaches use partial supervision to help guide the search process and narrow the space of possible solutions. The supervision information consists of a small set of pairwise constraints which can be obtained from multiple sources of information.

The first algorithm, called Cluster Constrained Multi-Model Unmixing (CC-MMU) [55], uses constraints on the cluster assignments of the pixels. The second algorithm, called Proportion Constrained Multi-Model Unmixing (PC-MMU), uses constraints on the proportion values of the pixels.

Algorithm 3.2 Context Dependent Spectral Unmixing using the Mahalanobis Distance (CDSU_M)

Inputs: $\mathbf{X}$: the data points ($N \times d$).

C : the number of contexts.

M : the number of endmembers for each context.

m : the fuzzifier, $m \in (1, +\infty)$.

α : the weight of the second term in the objective function.

β : the vector of weights in the second term of the objective function ($1 \times C$).

$\sigma_i > 0$, $i = 1..C$: the determinants of the norm matrices.

Outputs: $\mathbf{U}$: the fuzzy membership matrix of the data samples.

$\mathbf{c}_i$: the cluster centers.

$\mathbf{E}_i$: the sets of endmembers in all clusters.

$\mathbf{P}_i$: the sets of proportions in all clusters.

$\mathbf{A}_i$: the norm matrices for all clusters.

Initialize $\mathbf{c}_i$, $\mathbf{A}_i$ and $\mathbf{E}_i$.

repeat

 Update $\mathbf{P}_i$ using (3.10).

 Update $\mathbf{U}$ using (3.15).

 Update $\mathbf{E}_i$ using (3.9).

 Update $\mathbf{c}_i$ using (3.12).

 Update $\mathbf{A}_i$ using (3.16) and (3.17).

until parameters do not change significantly

return $\mathbf{U}$, $\mathbf{c}_i$, $\mathbf{E}_i$, $\mathbf{P}_i$, $\mathbf{A}_i$

3.4.1 Cluster Constrained Multi-Model Unmixing

CC-MMU uses constraints on the cluster assignments of the pixels. We assume that we dispose of two sets of constraints: a set of *should-link* pairs, $\mathcal{S}$, such that $(j, k) \in \mathcal{S}$ means that $\mathbf{x}_j$ and $\mathbf{x}_k$ should be assigned to the same cluster, and a set of *should not-link* pairs, $\mathcal{N}$, such that $(j, k) \in \mathcal{N}$ means that $\mathbf{x}_j$ and $\mathbf{x}_k$ should not be assigned to the same cluster. Each *should-link* and *should not-link* constraint has a violation cost ρ_{jk} . The constraints and their costs can be obtained by labeling few pixels in the hyperspectral image or by using information extracted from a different sensor.

Following the same notation as in Section 3.3, the proposed CC-MMU minimizes the following objective function:

$$\begin{aligned}
 J_C(\mathbf{E}_i, \mathbf{P}_i, \mathbf{U}, \mathbf{c}_i, \mathbf{A}_i) &= \sum_{i=1}^C \sum_{j=1}^N u_{ij}^m (\mathbf{x}_j - \mathbf{c}_i) \mathbf{A}_i (\mathbf{x}_j - \mathbf{c}_i)^T \\
 &+ \gamma \left(\sum_{(j,k) \in \mathcal{S}} \sum_{i=1}^C \sum_{l=1, l \neq i}^C \rho_{jk} u_{ij}^m u_{lk}^m + \sum_{(j,k) \in \mathcal{N}} \sum_{i=1}^C \rho_{jk} u_{ij}^m u_{ik}^m \right) \\
 &+ \alpha \sum_{i=1}^C \left[\sum_{j=1}^N u_{ij}^m (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i) (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T + \beta_i (M \text{trace}(\mathbf{E}_i \mathbf{E}_i^T) - \mathbf{1}_{1 \times M} \mathbf{E}_i \mathbf{E}_i^T \mathbf{1}_{M \times 1}) \right], \quad (3.18)
 \end{aligned}$$

subject to the constraints in (3.4), (3.5) and (3.14). γ is a positive constant.

In addition to the terms of the CDSU_M objective function described in Section 3.3, the

objective function of CC-MMU contains a semi-supervision term which would equal to zero when none of the defined *should-link* and *should not-link* constraints are violated.

We optimize (3.18) with respect to the centers $\mathbf{c}_i$, the memberships u_{ij} , the norm matrices $\mathbf{A}_i$, the proportions $\mathbf{p}_{ij}$ and the endmembers $\mathbf{E}_i$, using the Lagrange multipliers optimization method.

Theorem 3.4.1, the proof of which is given in Appendix C, gives the update equations of the local minimizers of J_C .

Theorem 3.4.1. *The update equations for the endmember sets, the abundances, the centers and the norm matrices are similar to the ones of CDSU_M (equations (3.9), (3.10), (3.12) and (3.16) respectively).*

The update equation for the memberships becomes

$$u_{ij} = \frac{\left[d_{ij}^2 + \gamma cost_{ij} + \alpha fit_{ij} \right]^{\frac{1}{1-m}}}{\sum_{q=1}^C \left[d_{qj}^2 + \gamma cost_{qj} + \alpha fit_{qj} \right]^{\frac{1}{1-m}}}, \quad (3.19)$$

where

$$d_{ij}^2 = (\mathbf{x}_j - \mathbf{c}_i) \mathbf{A}_i (\mathbf{x}_j - \mathbf{c}_i)^T, \quad (3.20)$$

$$cost_{ij} = \sum_{(j,k) \in \mathcal{S}} \sum_{l=1, l \neq i}^C \rho_{jk} u_{lk}^m + \sum_{(j,k) \in \mathcal{N}} \rho_{jk} u_{ik}^m, \quad (3.21)$$

and

$$fit_{ij} = (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i) (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T. \quad (3.22)$$

In (3.19), d_{ij}^2 represents the squared Mahalanobis distance between pixel j and the center of cluster i , $cost_{ij}$ represents the cost of violating the constraints related to pixel j with respect to cluster i , and fit_{ij} represents the squared error of fit of pixel j to the convex model i . Examining (3.19), we notice that a pixel j will have a high membership in cluster i if: (i) its spectrum is close to the centroid, $\mathbf{c}_i$, of that cluster in the spectral space; (ii) it does not violate many constraints that involve it; and (iii) it fits the model of that cluster.

The resulting CC-MMU is an iterative algorithm that uses alternating optimization. It involves successive updates of the norm matrices $\mathbf{A}_i$, the endmember proportions $\mathbf{p}_{ij}$, the memberships u_{ij} , the endmember sets $\mathbf{E}_i$, and the clusters' centers $\mathbf{c}_i$, until stabilization. It is summarized in Algorithm 3.3.

Algorithm 3.3 Cluster Constrained Multi-Model Unmixing (CC-MMU)

Inputs: $\mathbf{X}$: the data points ($N \times d$).
 C : the number of contexts.
 M : the number of endmembers for each context.
 m : the fuzzifier, $m \in (1, +\infty)$.
 α : the weight of the second term in the objective function.
 β : the vector of weights in the second term of the objective function ($1 \times C$).
 $\sigma_i > 0$, $i = 1..C$: the determinants of the norm matrices.
 $\gamma > 0$: the weight of the semi-supervised term.
 $\mathcal{S}$: the set of *should-link* constraints.
 $\mathcal{N}$: the set of *should not-link* constraints.
 $\rho_{jk} > 0, \forall (j, k) \in \mathcal{S} \text{ or } \mathcal{N}$: the constraints violation costs.

Outputs: $\mathbf{U}$: the fuzzy membership matrix of the data samples.
 $\mathbf{c}_i$: the cluster centers.
 $\mathbf{E}_i$: the sets of endmembers in all clusters.
 $\mathbf{P}_i$: the sets of proportions in all clusters.
 $\mathbf{A}_i$: the norm matrices for all clusters.

Initialize $\mathbf{c}_i$, $\mathbf{A}_i$ and $\mathbf{E}_i$.
repeat
 Update $\mathbf{P}_i$ using (3.10).
 Update $\mathbf{U}$ using (3.19), (3.20), (3.21) and (3.22).
 Update $\mathbf{E}_i$ using (3.9).
 Update $\mathbf{c}_i$ using (3.12).
 Update $\mathbf{A}_i$ using (3.16) and (3.17).
until parameters do not change significantly
return $\mathbf{U}$, $\mathbf{c}_i$, $\mathbf{E}_i$, $\mathbf{P}_i$, $\mathbf{A}_i$

3.4.2 Proportion Constrained Multi-Model Unmixing

PC-MMU uses constraints on the proportion values of the pixels. We assume that we dispose of a set of pixel pairs, $\mathcal{S}$, such that $(j, k) \in \mathcal{S}$ means that $\mathbf{x}_j$ and $\mathbf{x}_k$ should have similar proportion values in the extracted endmembers. Each constraint has a violation cost ρ_{jk} . The constraints and their costs can be obtained through some ground truth knowledge or through consensus unmixing as will be shown in the next chapter.

Following the same notation as in Section 3.3, the proposed PC-MMU minimizes the following objective function:

$$J_P(\mathbf{E}_i, \mathbf{P}_i, \mathbf{U}, \mathbf{c}_i, \mathbf{A}_i) = \sum_{i=1}^C \sum_{j=1}^N u_{ij}^m (\mathbf{x}_j - \mathbf{c}_i) \mathbf{A}_i (\mathbf{x}_j - \mathbf{c}_i)^T + \gamma \sum_{(j,k) \in \mathcal{S}} \rho_{jk} \sum_{i=1}^C \|\mathbf{p}_{ij} - \mathbf{p}_{ik}\|^2 \\ + \alpha \sum_{i=1}^C \left[\sum_{j=1}^N u_{ij}^m (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i) (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T + \beta_i (M \text{trace}(\mathbf{E}_i \mathbf{E}_i^T) - \mathbf{1}_{1 \times M} \mathbf{E}_i \mathbf{E}_i^T \mathbf{1}_{M \times 1}) \right], \quad (3.23)$$

subject to the constraints in (3.4), (3.5) and (3.14). γ is a positive constant.

In addition to the terms of the CDSU_M objective function described in Section 3.3, the objective function of PC-MMU contains a semi-supervision term which would equal to zero when

none of the defined proportion constraints are violated.

We optimize (3.23) with respect to the centers $\mathbf{c}_i$, the memberships u_{ij} , the norm matrices $\mathbf{A}_i$, the proportions $\mathbf{p}_{ij}$ and the endmembers $\mathbf{E}_i$, using the Lagrange multipliers optimization method.

Theorem 3.4.2, the proof of which is given in Appendix D, gives the update equations of the local minimizers of J_P .

Theorem 3.4.2. *The update equations for the endmember sets, the memberships, the centers and the norm matrices are similar to the ones of CDSU_M (equations (3.9), (3.15), (3.12) and (3.16) respectively).*

The update equation for the proportions becomes

$$\mathbf{p}_{ij}^T = \max \left(\left[2\alpha u_{ij}^m \mathbf{E}_i \mathbf{E}_i^T + \gamma \sum_{(j,k) \in \mathcal{S}} \rho_{jk} \right]^{-1} \left[2\alpha u_{ij}^m \mathbf{E}_i \mathbf{x}_j^T + \gamma \sum_{(j,k) \in \mathcal{S}} \rho_{jk} \mathbf{p}_{ik}^T + \xi_{ij} \mathbf{1}_{M \times 1} \right], 0 \right), \quad (3.24)$$

where

$$\xi_{ij} = \frac{1 - \mathbf{1}_{1 \times M} \left[2\alpha u_{ij}^m \mathbf{E}_i \mathbf{E}_i^T + \gamma \sum_{(j,k) \in \mathcal{S}} \rho_{jk} \right]^{-1} \left[2\alpha u_{ij}^m \mathbf{E}_i \mathbf{x}_j^T + \gamma \sum_{(j,k) \in \mathcal{S}} \rho_{jk} \mathbf{p}_{ik}^T \right]}{\mathbf{1}_{1 \times M} \left[2\alpha u_{ij}^m \mathbf{E}_i \mathbf{E}_i^T + \gamma \sum_{(j,k) \in \mathcal{S}} \rho_{jk} \right]^{-1} \mathbf{1}_{M \times 1}}. \quad (3.25)$$

The resulting PC-MMU is an iterative algorithm that uses alternating optimization. It involves successive updates of the norm matrices $\mathbf{A}_i$, the proportions $\mathbf{p}_{ij}$, the memberships u_{ij} , the endmember sets $\mathbf{E}_i$, and the clusters' centers $\mathbf{c}_i$, until stabilization. It is summarized in Algorithm 3.4.

3.5 Robust Context Dependent Spectral Unmixing

One limitation of most multiple model unmixing methods is their sensitivity to noise and outliers due to scene and/or sensor effects [84]. This limitation is inherited from the fuzzy clustering approach used to learn the multiple convex sets. Noise points affect not only the convex sets, but also the estimates of the endmembers and proportions within each set.

Possibilistic clustering [85] has been proposed to overcome the sensitivity of fuzzy clustering to noise. This approach uses possibilistic membership functions to identify and reduce the effect of noise points. Possibilistic memberships may however result in identical clusters. Recent approaches combine fuzzy and possibilistic memberships to avoid such problems [53].

In this section, we propose using both fuzzy and possibilistic membership functions to develop a robust multi-model unmixing algorithm. Fuzzy memberships are used to partition the spectra space

Algorithm 3.4 Proportion Constrained Multi-Model Unmixing (PC-MMU)

Inputs: $\mathbf{X}$: the data points ($N \times d$).

C : the number of contexts.

M : the number of endmembers for each context.

m : the fuzzifier, $m \in (1, +\infty)$.

α : the weight of the second term in the objective function.

β : the vector of weights in the second term of the objective function ($1 \times C$).

$\sigma_i > 0$, $i = 1..C$: the determinants of the norm matrices.

$\gamma > 0$: the weight of the semi-supervised term.

$\mathcal{S}$: the set of proportion constraints.

$\rho_{jk} > 0, \forall (j, k) \in \mathcal{S}$: the constraints violation costs.

Outputs: $\mathbf{U}$: the fuzzy membership matrix of the data samples.

$\mathbf{c}_i$: the cluster centers.

$\mathbf{E}_i$: the sets of endmembers in all clusters.

$\mathbf{P}_i$: the sets of proportions in all clusters.

$\mathbf{A}_i$: the norm matrices for all clusters.

Initialize $\mathbf{c}_i$, $\mathbf{A}_i$ and $\mathbf{E}_i$.

repeat

 Update $\mathbf{P}_i$ using (3.24).

 Update $\mathbf{U}$ using (3.15).

 Update $\mathbf{E}_i$ using (3.9).

 Update $\mathbf{c}_i$ using (3.12).

 Update $\mathbf{A}_i$ using (3.16) and (3.17).

until parameters do not change significantly

return $\mathbf{U}$, $\mathbf{c}_i$, $\mathbf{E}_i$, $\mathbf{P}_i$, $\mathbf{A}_i$

into multiple convex sets that span the entire space and avoid coincident clusters [72]. Possibilistic memberships are used to reduce the effect of noise and obtain robust estimates of the endmembers and proportions within each cluster.

The Robust Context Dependent Spectral Unmixing (RCDSU) combines fuzzy and possibilistic clustering with linear unmixing into a joint objective function. The clustering component is used to partition the data into multiple contexts. The linear unmixing component learns endmembers and abundances within each context. Following the same notation as in Section 3.3, RCDSU minimizes

$$J_R(\mathbf{E}_i, \mathbf{P}_i, \mathbf{U}, \mathbf{T}, \mathbf{c}_i, \mathbf{A}_i) = \sum_{i=1}^C \sum_{j=1}^N (au_{ij}^m + bt_{ij}^n)(\mathbf{x}_j - \mathbf{c}_i)\mathbf{A}_i(\mathbf{x}_j - \mathbf{c}_i)^T + \sum_{i=1}^C \eta_i \sum_{j=1}^N (1 - t_{ij})^n \\ + \alpha \sum_{i=1}^C \left[\sum_{j=1}^N (au_{ij}^m + bt_{ij}^n)(\mathbf{x}_j - \mathbf{p}_{ij}\mathbf{E}_i)(\mathbf{x}_j - \mathbf{p}_{ij}\mathbf{E}_i)^T + \beta_i (Mtrace(\mathbf{E}_i^T \mathbf{E}_i) - \mathbf{1}_{1 \times M} \mathbf{E}_i \mathbf{E}_i^T \mathbf{1}_{M \times 1}) \right], \quad (3.26)$$

subject to the constraints in (3.4), (3.5), (3.14), and

$$t_{ij} \in [0, 1], \forall i, j. \quad (3.27)$$

The first and second terms in (3.26) represent the unsupervised learning component. They seek to partition the N samples into C clusters, and represent each cluster by a center $\mathbf{c}_i$ and a norm matrix $\mathbf{A}_i$. Each sample $\mathbf{x}_j$ will be assigned to each cluster i with two types of memberships. The

fuzzy membership [72], u_{ij} , is the degree with which pixel j belongs to cluster i ; and the possibilistic membership [85], t_{ij} , is the possibility or typicality that pixel j belongs to cluster i . The memberships are weighted by two positive constants, a and b (with $a + b = 1$), that control the prevalence of each membership for the task at hand. The constants m and $n \in (1, +\infty)$ are the fuzzifiers [72] for both memberships. The second term in (3.26) forces the possibilistic memberships, t_{ij} , to be as large as possible, thus avoiding the trivial solution of having them all equal to zero.

The third and fourth terms in (3.26) represent the spectral unmixing component and are similar to the ones of CDSU. Finally, $\boldsymbol{\eta} = [\eta_1, \dots, \eta_C]$ are positive values.

We optimize (3.26) with respect to all parameters using the Lagrange multipliers optimization method. Theorem 3.5.1, the proof of which is given in Appendix E, gives the update equations of the local minimizers of J_R .

Theorem 3.5.1. *The update equations for the proportions and the memberships are similar to the ones of CDSU_M (equations (3.10) and (3.15) respectively).*

The update equation for the endmember sets, $\mathbf{E}_i$, is

$$\mathbf{E}_i = \left[\beta_i (M\mathbf{I}_{M \times M} - \mathbf{1}_{M \times M}) + \sum_{j=1}^N (au_{ij}^m + bt_{ij}^n) \mathbf{p}_{ij}^T \mathbf{p}_{ij} \right]^{-1} \left[\sum_{j=1}^N (au_{ij}^m + bt_{ij}^n) \mathbf{p}_{ij}^T \mathbf{x}_j \right]. \quad (3.28)$$

The update equation for the norm matrices, $\mathbf{A}_i$, is the same as in equation (3.16), but with

$$\mathbf{C}_i = \frac{\sum_{j=1}^N (au_{ij}^m + bt_{ij}^n) (\mathbf{x}_j - \mathbf{c}_i)^T (\mathbf{x}_j - \mathbf{c}_i)}{\sum_{j=1}^N (au_{ij}^m + bt_{ij}^n)}. \quad (3.29)$$

The update equation for the typicalities, t_{ij} , is given by

$$t_{ij} = \frac{1}{1 + \left[b \frac{\text{cost}_{ij}}{\eta_i} \right]^{\frac{1}{n-1}}}, \quad (3.30)$$

where

$$\text{cost}_{ij} = (\mathbf{x}_j - \mathbf{c}_i) \mathbf{A}_i (\mathbf{x}_j - \mathbf{c}_i)^T + \alpha (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i) (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T. \quad (3.31)$$

Finally, the update equation of the cluster centers $\mathbf{c}_i$ is

$$\mathbf{c}_i = \frac{\sum_{j=1}^N (au_{ij}^m + bt_{ij}^n) \mathbf{x}_j}{\sum_{j=1}^N (au_{ij}^m + bt_{ij}^n)}. \quad (3.32)$$

Unlike the fuzzy memberships, the possibilistic memberships are not constrained to sum to one. They provide an intuitive notion of typicality. Note that the larger cost_{ij} is, the smaller the

typicality of pixel j in cluster i . This means that a pixel that is far from any cluster and that does not fit any convex model would have small typicality values in all clusters, and hence can be considered as an outlier or a noise point. The value of η_i determines when the membership of a point in cluster i becomes 0.5 (i.e., the 3 dB point). Thus, it needs to be chosen depending on the desired “bandwidth” of the membership distribution for that cluster.

The resulting RCDSU is an iterative algorithm that uses alternating optimization. It involves successive updates of the endmember proportions $\mathbf{p}_{ij}$, the fuzzy memberships u_{ij} , the possibilistic memberships t_{ij} , the endmember sets $\mathbf{E}_i$, the cluster centers $\mathbf{c}_i$, and the norm matrices $\mathbf{A}_i$, until convergence. It is summarized in Algorithm 3.5.

Algorithm 3.5 Robust Context Dependent Spectral Unmixing (RCDSU)

Inputs: $\mathbf{X}$: the data points ($N \times d$).
 C : the number of contexts.
 M : the number of endmembers for each context.
 m, n : the fuzzifier, $m, n \in (1, +\infty)$.
 α : the weight of the second term in the objective function.
 β : the vector of weights in the second term of the objective function ($1 \times C$).
 $\sigma_i > 0, i = 1..C$: the determinants of the norm matrices.
 a, b : weights of the fuzzy and possibilistic memberships ($a + b = 1$).
Outputs: $\mathbf{U}$: the fuzzy membership matrix of the data samples.
 $\mathbf{T}$: the possibilistic membership matrix of the data samples.
 $\mathbf{c}_i$: the cluster centers.
 $\mathbf{E}_i$: the sets of endmembers in all clusters.
 $\mathbf{P}_i$: the sets of proportions in all clusters.
 $\mathbf{A}_i$: the norm matrices for all clusters.

Initialize $\mathbf{c}_i, \mathbf{A}_i$ and $\mathbf{E}_i$.
repeat
 Update $\mathbf{P}_i$ using (3.10).
 Update $\mathbf{U}$ using (3.15).
 Update $\mathbf{T}$ using (3.30) and (3.31).
 Update $\mathbf{E}_i$ using (3.28).
 Update $\mathbf{c}_i$ using (3.32).
 Update $\mathbf{A}_i$ using (3.16) and (3.29).
until parameters do not change significantly
return $\mathbf{U}, \mathbf{T}, \mathbf{c}_i, \mathbf{E}_i, \mathbf{P}_i, \mathbf{A}_i$

3.6 Unsupervised Robust Context Dependent Spectral Unmixing

The proposed RCDSU algorithm, like other multi-model unmixing algorithms, assumes that the optimal number of endmember sets is known. However, this may not be the case, and it should be learned from the data. In this section, we extend RCDSU to find the “optimal” number of contexts in an unsupervised way.

Our approach is inspired by [86] and explores properties of the possibilistic membership functions. First, we overspecify the number of clusters C and run the RCDSU algorithm. Second, we ignore the fuzzy memberships (by setting a to 0 and b to 1) and run RCDSU to allow clusters to expand over neighboring regions. Since the possibilistic memberships are not constrained, small clusters covering the same dense regions would expand and become similar. Finally, we use a similarity measure to identify similar clusters and merge them. We use

$$S_{ij} = 1 - \frac{\sum_{k=1}^N |t_{ik} - t_{jk}|}{\sum_{k=1}^N t_{ik} + \sum_{k=1}^N t_{jk}}. \quad (3.33)$$

S_{ij} does not depend on the distance measure explicitly, and therefore, it can be used independently of the clusters shape and size. More importantly, it does take into account the model fitting error (used implicitly in the memberships). It can be easily shown that $0 \leq S_{ij} \leq 1$, where $S_{ij} = 1$ when clusters i and j are identical, and $S_{ij} = 0$ when the clusters are disjoint.

The Unsupervised RCDSU (U-RCDSU) is summarized in Algorithm 3.6. We should point out here that even when the number of clusters is known a priori, it is recommended to overspecify the number of clusters. This makes the results less sensitive to initialization and gives tiny clusters (compared to other clusters) a better chance of being detected.

Algorithm 3.6 Unsupervised Robust Context Dependent Spectral Unmixing (U-RCDSU)

Inputs: $\mathbf{X}$: the data points ($N \times d$).

$C = C_{max}$: the over-specified number of clusters.

M : the number of endmembers for each context.

m, n : the fuzzifier, $m, n \in (1, +\infty)$.

α : the weight of the second term in the objective function.

β : the vector of weights in the second term of the objective function ($1 \times C$).

$\sigma_i > 0, i = 1..C$: the determinants of the norm matrices.

a, b : weights of the fuzzy and possibilistic memberships ($a + b = 1$).

Outputs: $\mathbf{U}$: the fuzzy membership matrix of the data samples.

$\mathbf{T}$: the possibilistic membership matrix of the data samples.

$\mathbf{c}_i$: the cluster centers.

$\mathbf{E}_i$: the sets of endmembers in all clusters.

$\mathbf{P}_i$: the sets of proportions in all clusters.

$\mathbf{A}_i$: the norm matrices for all clusters.

Initialize $\mathbf{c}_i, \mathbf{A}_i$ and $\mathbf{E}_i$.

$[\mathbf{U}, \mathbf{T}, \mathbf{c}_i, \mathbf{A}_i, \mathbf{E}_i, \mathbf{P}_i] = \text{RCDSU}(\mathbf{c}_i, \mathbf{A}_i, \mathbf{E}_i)$.

Set a to 0, b to 1.

repeat

$[\mathbf{U}, \mathbf{T}, \mathbf{c}_i, \mathbf{A}_i, \mathbf{E}_i, \mathbf{P}_i] = \text{RCDSU}(\mathbf{c}_i, \mathbf{A}_i, \mathbf{E}_i)$.

for each pair of clusters i and j **do**

 Compute S_{ij} using (3.33).

if $S_{ij} \geq 1 - \epsilon$ **then**

 Merge clusters i and j .

end if

end for

if a merging occurred **then**

 Update $C, \mathbf{c}_i$ and $\mathbf{A}_i$.

 Initialize $\mathbf{E}_i$.

end if

until no merging takes place

Reset a and b .

$[\mathbf{U}, \mathbf{T}, \mathbf{c}_i, \mathbf{A}_i, \mathbf{E}_i, \mathbf{P}_i] = \text{RCDSU}(\mathbf{c}_i, \mathbf{A}_i, \mathbf{E}_i)$.

return $\mathbf{U}, \mathbf{T}, \mathbf{c}_i, \mathbf{E}_i, \mathbf{P}_i, \mathbf{A}_i$

CHAPTER 4

ROBUST UNMIXING USING CONSENSUS ANALYSIS

Spectral unmixing is a challenging, ill-posed, inverse problem. Many algorithms have been proposed for robust, stable, and accurate unmixing solutions. Different algorithms have different modes of operation and usually yield different results. Moreover, most of them require specifying the number of endmembers to be extracted before hand. In Chapter 3, we have proposed various algorithms that can identify endmembers while taking the distribution of the data into account. In this chapter, we propose using consensus analysis on multiple unmixing results to find the “optimal” endmembers in the data. We run different unmixing algorithms, using different numbers of endmembers, and combine the results using consensus analysis. The claim is that actual endmembers will have a consensus among all runs. This chapter is organized as follows. First, we present the motivation behind the proposed idea. Then, we outline the proposed consensus unmixing approach.

4.1 Motivations

Many unmixing algorithms exist in the literature [29] and different algorithms often result in different endmembers for the same data. Moreover, the same algorithm may not result in the same endmembers when run multiple times. This is mainly due to the non-deterministic behavior of the algorithm. In fact, each method has its own assumptions and approach for estimating the endmembers. The goal is to take advantage of this diversity to estimate the “optimal” endmembers using consensus analysis.

The idea has its roots in consensus clustering [57], where different mechanisms are proposed to explore the consensus between multiple data partitions. The idea of consensus analysis has been investigated in [87] to estimate the number of endmembers. However, it did not take into account different unmixing algorithms. In fact, a single algorithm is run multiple times using the same number of endmembers, and the goal is to find the number that achieves the most stable classification among all runs. In this chapter, we investigate using multiple algorithms with multiple parameter settings to find an accurate and consistent set of endmembers in the data. The claim is

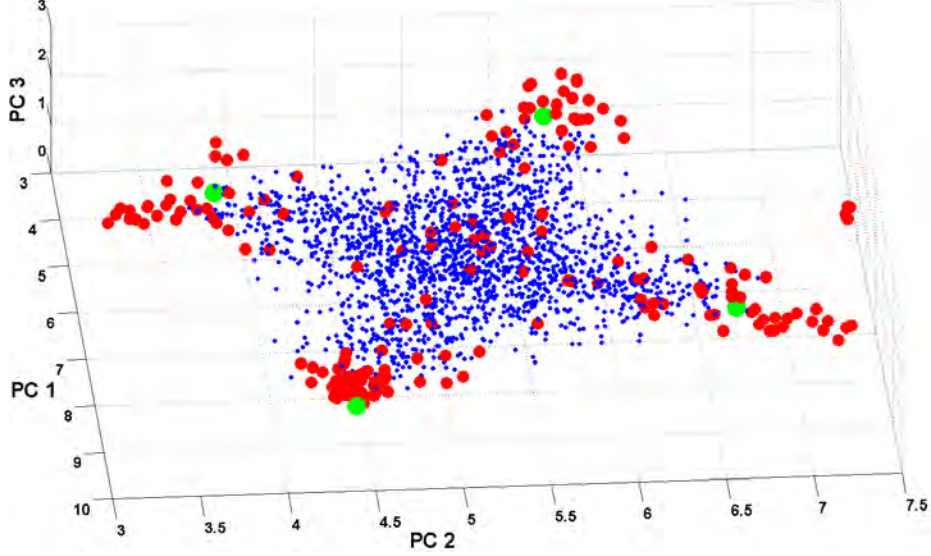


Figure 4.1: First 3 PC of the data (blue), endmembers from all unmixing runs (red), and true endmembers (green).

that “real” endmembers will have a consensus among multiple runs.

To illustrate the diversity of different unmixing results, we use a simulated hyperspectral data of 2000 points generated according to equation (1.1), using 4 endmembers (*Spessartine*, *Halloysite*, *Chlorite* and *Lizardite*) from the USGS digital spectral library [88]. The proportions are linearly generated between 0 and 0.9 for each endmember. A zero-mean Gaussian noise at a signal-to-noise ratio (SNR) of 20 dB was added to the data. For this example, we run ICE [61], VCA [66], PPI [59], and N-FINDR [60] using a number of endmembers varying from 3 to 7, repeating each run 4 times (giving a total of 80 runs). We also run U-RCDSU using $C_{max} = 5$ with $M = 2, 3$, and 4 (we let $m = n = 1.5$, $a = b = 0.5$, $\alpha = 100$, $\beta_i = 4 \forall i$, $\sigma_i = 1 \forall i$, and $\epsilon = 0.2$), leading to 1 context every time with 2, 3 and 4 endmembers each. In total, we have 83 endmember sets.

In figure 4.1, we scatter plot the first 3 principal components (PC) of the data in small blue dots, the resulting endmembers from all runs in large red dots, and the true endmembers in larger green dots. Since most unmixing methods are expected to identify reasonable endmembers on most runs, we observe a concentration of estimated endmembers around the true endmembers. This observation is explored to develop an approach for robust endmembers estimation.

4.2 Consensus Unmixing

Let $\mathbf{X} = \{\mathbf{x}_j \in \mathbb{R}^d, j = 1, \dots, N\}$ be the N spectral signatures obtained from a hyperspectral image having d spectral bands. Let T be the total number of unmixings on $\mathbf{X}$. The T unmixing

results could be obtained by running different unmixing algorithms, one unmixing algorithm with different parameters/initializations, or a combination of the two. Let $\mathbf{E}_i = \{\mathbf{e}_{ik}, k = 1, \dots, M_i\}$ be the set of M_i endmembers resulting from unmixing i , $i = 1, \dots, T$. Let $\mathbf{P}_i = \{p_{jik}, j = 1, \dots, N; k = 1, \dots, M_i\}$ be the set of proportions, p_{jik} , of endmember k in pixel j resulting from unmixing i , $i = 1, \dots, T$.

The first step in our approach is to select a subset, $\mathcal{H}$, of data points that are more likely to inform us about the best locations of the endmembers. For each endmember $\mathbf{e}_{ik}$, we form an α_p -cut set

$$(\mathbf{p}_{ik})_{\alpha_p} = \{j \mid p_{jik} \geq \alpha_p\}. \quad (4.1)$$

$(\mathbf{p}_{ik})_{\alpha_p}$ contains the indices of the points having proportion values, in endmember $\mathbf{e}_{ik}$, that are larger than α_p .

For computational efficiency, we try to keep $\mathcal{H}$ as small as possible while maintaining diversity. Thus, if $|(\mathbf{p}_{ik})_{\alpha_p}| > Q$, we select the top Q points having the largest proportion values. Otherwise, we consider the entire set. We call these points with high proportions the *voters* to $\mathbf{e}_{ik}$. The subset $\mathcal{H}$ is the union of the voters to all endmembers in all runs:

$$\mathcal{H} = \bigcup_{i,k} (\mathbf{p}_{ik})_{\alpha_p}. \quad (4.2)$$

The second step of our approach aims to reduce the size of $\mathcal{H}$ and keep only consistent voters. In fact, $\mathcal{H}$ may include voters to endmembers that are not consistent. In order to filter those out, we construct a co-association matrix [89], $\mathcal{C}$, for the elements of $\mathcal{H}$, and keep those with high association values. The co-association matrix forms a voting mechanism that combines the unmixing results, leading to a new measure of similarity between points. The h points in $\mathcal{H}$ are mapped into an $h \times h$ co-association matrix $\mathcal{C}$ using

$$\mathcal{C}(j, l) = \sum_{i=1}^T \sum_{k=1}^{M_i} p_{jik} p_{lik} \quad (4.3)$$

In other words, if pixels j and l belong to the same “group” of endmembers, then they have high proportion values in some common endmembers of that group. Hence, their co-association value will be high. On the other hand, if pixels j and l do not share any endmember within the group, then their co-association value will be low.

Using an α_c -cut over $\mathcal{C}$, we keep only points having high co-association values. As a result, we obtain a smaller subset, $\mathcal{H}_a$, of voters to mainly consistent endmembers:

$$\mathcal{H}_a = (\mathcal{C})_{\alpha_c} = \{j, l \mid j \neq l \text{ and } \mathcal{C}(j, l) \geq \alpha_c\}. \quad (4.4)$$

Algorithm 4.1 Robust Unmixing Using Consensus Analysis

Input: $\{\mathbf{P}_i = \{p_{jik}, j = 1..N; k = 1..M_i\}\}_{i=1}^T, Q, \alpha_p, \alpha_c$

Output: Optimal endmembers in the data

$$(\mathbf{p}_{ik})_{\alpha_p} = \{j \mid p_{jik} \geq \alpha_p\}, \forall i = 1, \dots, T, k = 1, \dots, M_i$$

$$\mathcal{H} = \bigcup_{i,k} (\mathbf{p}_{ik})_{\alpha_p}$$

$$\mathcal{C}(j, l) = \sum_{i=1}^T \sum_{k=1}^{M_i} p_{jik} p_{lik}, \forall j, l \in \mathcal{H}$$

$$\mathcal{H}_a = (\mathcal{C})_{\alpha_c} = \{j, l \mid j \neq l \text{ and } \mathcal{C}(j, l) \geq \alpha_c\}$$

$$\mathcal{C}_a(j, l) = \mathcal{C}(j, l), \forall j, l \in \mathcal{H}_a$$

$$\text{Convert } \mathcal{C}_a \text{ to a dissimilarity } \mathbf{D} = \exp \left[\frac{-\mathcal{C}_a}{\max(\mathcal{C}_a)} \right]$$

$$\text{diag}(\mathbf{D}) = \mathbf{0}$$

Use a clustering algorithm to identify clusters within $\mathbf{D}$

Select one endmember from each cluster (as the cluster representative)

The subset $\mathcal{H}_a$ can be used as the set of constraints for the Proportion Constrained Multi-Model Unmixing (PC-MMU) algorithm proposed in Section 3.4.2.

The last step of our approach consists of identifying clusters within $\mathcal{H}_a$. Each cluster represents a set of similar endmembers (estimated by the different algorithms/runs). We also estimate the optimal number of clusters. Since $\mathcal{H}_a$ is expected to include a set of well-separated clusters, various clustering algorithms could be used for this task. In our work, we report results using a simple average link hierarchical clustering [90]. Since $\mathcal{C}_a$, sub-matrix of $\mathcal{C}$ corresponding to $\mathcal{H}_a$, is a similarity matrix, we first convert it to a dissimilarity matrix $\mathbf{D}$ using

$$\mathbf{D} = \exp \left[\frac{-\mathcal{C}_a}{\max(\mathcal{C}_a)} \right]. \quad (4.5)$$

Furthermore, the diagonal elements of $\mathbf{D}$ are set to zeros, as they represent the dissimilarities of identical points. Once the voters are clustered, a representative from each cluster is chosen to be an endmember. Another approach would be to run an unmixing algorithm on this reduced subset of voters, using the number of identified clusters as the number of endmembers. Algorithm 4.1 summarizes the proposed approach.

CHAPTER 5

CONTEXT DEPENDENT HYPERSPECTRAL SUBPIXEL TARGET DETECTION

In this chapter, we introduce a new class of subpixel target detection algorithms that use a local structured background model. Our approach, referred to as Context Dependent Target Detectors, extends existing structured detectors to multiple contexts. It is based on the robust context dependent spectral unmixing algorithm, presented in Chapter 3, to model the background variability. The claim is that robust context dependent unmixing provides a better description of the background with minimum target leakage, compared to global unmixing, and hence results in a better target-background separation. The approach is evaluated using the Adaptive Matched Subspace Detector (AMSD) [76], the Orthogonal Subspace Projection (OSP) detector [76] and the Hybrid Subspace Detector (HSD) [79].

5.1 Motivations

As mentioned in Section 2.2, target detection algorithms face the challenge of target leakage into the background, which is the contribution of the target pixels to the background model. This happens due to the presence of targets in the scene. An illustration of this is shown in figure 5.1. The blue dots represent the data and form one convex set. The cyan dots represent noise points. The red cross represents a pixel containing the target. A non robust unmixing algorithm will identify the endmember set displayed by the red dots. In this case, the target will have a good fit in the background model and may not be detected. A robust unmixing algorithm, on the other hand, will ignore noise and target points and can identify the endmember set displayed by green dots. In this case, the target does not fit the background model and can be easily detected.

The proposed robust context dependent spectral unmixing (RCDSU) algorithm, presented in Section 3.5, can be used to solve this problem. In fact, targets can be thought of as noise points or outliers not belonging to the background. They will be assigned low typicalities, meaning that they will not contribute to the estimated endmembers. In other words, there will be no leakage from

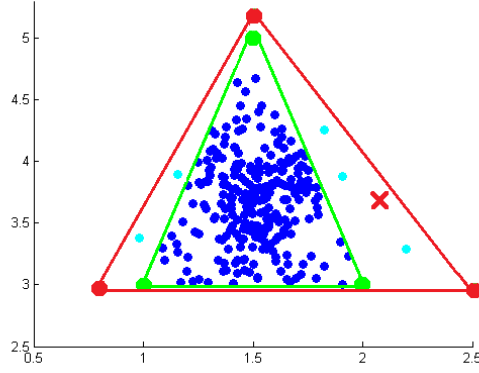


Figure 5.1: Target (red cross) leakage with non robust unmixing (red dots). No target leakage with robust unmixing (green dots).

the target subspace to the background subspace, which is important in target detection [76].

Another challenge with existing target detection algorithms, that are based on the structured background model, is that they usually use global methods to describe the background. Some of them use eigenvector decomposition of the data correlation matrix, while others use unmixing algorithms that find a single set of endmembers. These global methods may not provide a good description of the hyperspectral data, especially when the scene includes multiple regions with distinct materials. Hence, unmixing methods that can find multiple endmember sets are more appropriate.

This problem is illustrated in figure 5.2 using a 2 dimensional synthetic data. The blue points represent the data that form two convex sets. The red cross represents a pixel containing the target. The red dots represent the endmembers resulting from a typical single model unmixing algorithm. If these endmembers were used to describe the background, the target pixel would be a normal background pixel, and hence might not be detected. On the other hand, if a multi-model unmixing algorithm was used, a typical result would be the 2 sets of endmembers displayed as green dots. In this case, the target pixel will not fit the background model as it is located outside both convex hulls.

To summarize, our motivations are two-fold. First, there is the need for a better background description using multiple sets of endmembers. Second, there is the need to limit target leakage into the background subspace. These can be accounted for using the proposed robust context dependent spectral unmixing (RCDSU).

In the following, we assume that the multiple contexts have been identified using RCDSU. Given a hyperspectral data, we will have the background endmember sets $\mathbf{E}_i$, $i = 1..C$; the propor-

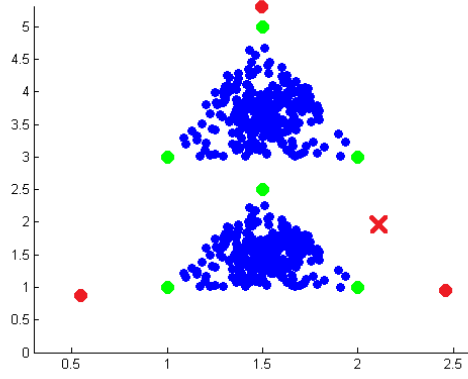


Figure 5.2: Multiple endmember sets (green dots) versus a one global set (ref dots) for target (red cross) detection.

tions sets $\mathbf{P}_i = \{\mathbf{p}_{ij}, j = 1..N\}$, $i = 1..C$; the fuzzy memberships $\mathbf{U} = \{u_{ij}, i = 1..C, j = 1..N\}$, and the cluster centers $\mathbf{c}_i$, $i = 1..C$.

5.2 Context Dependent Target Detectors

In this section, we extend the AMSD [76], OSP [76] and HSD [79] to multiple contexts using RCDSU. A test pixel, $\mathbf{x}_j$, will be assigned a local detection statistic within each context i , and these local statistics will be combined using the pixel's fuzzy memberships resulting from RCDSU, u_{ij} , to produce one confidence value. In the following, $\mathbf{s}_t$ denotes the spectral signature of the target that we are trying to detect.

5.2.1 Context Dependent AMSD

The Context Dependent AMSD (CD-AMSD) extends the AMSD to multiple local endmember sets. Given a test pixel $\mathbf{x}_j$ in the data, the detection confidence value is computed as

$$T_{\text{CD-AMSD}}(\mathbf{x}_j) = \sum_{i=1}^C u_{ij} \frac{\mathbf{x}_j \mathbf{P}_{\mathbf{b}_i}^\perp \mathbf{x}_j^T}{\mathbf{x}_j \mathbf{P}_{\mathbf{S}_i}^\perp \mathbf{x}_j^T}, \quad (5.1)$$

where $\mathbf{P}_{\mathbf{b}_i}^\perp = \mathbf{I} - \mathbf{E}_i^T (\mathbf{E}_i \mathbf{E}_i^T)^{-1} \mathbf{E}_i$ is the orthogonal projection error matrix on the local background subspace $\mathbf{E}_i$ resulting from RCDSU, and $\mathbf{P}_{\mathbf{S}_i}^\perp = \mathbf{I} - \mathbf{S}_i^T (\mathbf{S}_i \mathbf{S}_i^T)^{-1} \mathbf{S}_i$ is the orthogonal projection error matrix on the local composite target and background subspace $\mathbf{S}_i = [\mathbf{s}_t; \mathbf{E}_i]$.

5.2.2 Context Dependent OSP Detector

The Context Dependent OSP (CD-OSP) detector provides a local approach to the traditional OSP detector. The detection statistics are computed within each context, and combined using the

fuzzy memberships of the test pixel in all sets, i.e.,

$$T_{\text{CD-OSP}}(\mathbf{x}_j) = \sum_{i=1}^C u_{ij} \mathbf{s}_t \mathbf{P}_{\mathbf{b}i}^\perp \mathbf{x}_j^T, \quad (5.2)$$

where $\mathbf{P}_{\mathbf{b}i}^\perp$ is the same as in (5.1).

5.2.3 Context Dependent HSD

The Context Dependent HSD (CD-HSD) extends the HSD to multiple local background subspaces using

$$T_{\text{CD-HSD}}(\mathbf{x}_j) = \sum_{i=1}^C u_{ij} \frac{(\mathbf{x}_j - \mathbf{p}_{\mathbf{b}ij} \mathbf{E}_i) \Gamma_i^{-1} (\mathbf{x}_j - \mathbf{p}_{\mathbf{b}ij} \mathbf{E}_i)^T}{(\mathbf{x}_j - \mathbf{p}_{\mathbf{s}ij} \mathbf{S}_i) \Gamma_i^{-1} (\mathbf{x}_j - \mathbf{p}_{\mathbf{s}ij} \mathbf{S}_i)^T}. \quad (5.3)$$

In (5.3), $\mathbf{E}_i$ is the i^{th} endmember set resulting from RCDSU, and $\mathbf{S}_i = [\mathbf{s}_t; \mathbf{E}_i]$ is the i^{th} local composite target and background subspace. $\mathbf{p}_{\mathbf{b}ij}$ and $\mathbf{p}_{\mathbf{s}ij}$ are the proportions of the j^{th} test pixel in the local background subspace $\mathbf{E}_i$, and local composite target and background subspace $\mathbf{S}_i$, respectively. $\mathbf{p}_{\mathbf{b}ij}$ results from RCDSU and $\mathbf{p}_{\mathbf{s}ij}$ is computed by plugging the composite target and background subspace $\mathbf{S}_i$, instead of $\mathbf{E}_i$, in the update equation (3.10) of the proportions of CDSU.

The covariance matrix of context i , Γ_i , is estimated using

$$\Gamma_i = \frac{\sum_{j=1}^N (au_{ij}^m + bt_{ij}^n) (\mathbf{x}_j - \mathbf{c}_i)^T (\mathbf{x}_j - \mathbf{c}_i)}{\sum_{j=1}^N (au_{ij}^m + bt_{ij}^n)}, \quad (5.4)$$

which is already determined by RCDSU to compute the norm matrices $\mathbf{A}_i$.

CHAPTER 6

EXPERIMENTAL RESULTS

This chapter presents the experimental results of the proposed algorithms on synthetic and real hyperspectral data. The experiments were ran using MATLAB R2011a on a computer equipped with a 3.6 GHz Intel Xeon processor and a 24 GB RAM. First, we present the data sets to be used in the experiments, then we present the results of the different algorithms on these data sets.

6.1 Data sets

Two kinds of data sets are used in the experiments: simulated and real data. Simulated data are used to prove the concepts of the proposed algorithms. Real data, on the other hand, are used to test the proposed methods in real case scenarios.

6.1.1 Simulated data

Two categories of simulated data are used: 2-dimensional and synthetic hyperspectral data.

6.1.1.1 Two-dimensional data

Three data sets were used as motivational examples in Section 3.1. These are:

- **D2C1M3** in Section 3.1.1: it forms one convex set with 3 endmembers,
- **D2C2M3** in Section 3.1.2: it forms two convex sets with 3 endmembers each, and
- **D2C3M3** in Section 3.1.3: it forms three convex sets with 3 endmembers each.

Another 2-dimensional data set is generated to prove the concept of the CDSU_M algorithm:

- **D2EC2M3**: it includes 2 elongated convex sets generated using the linear model of equation (1.1) with 3 endmembers for each set. The data is shown in figure 6.1. Each cluster contains 2000 points.

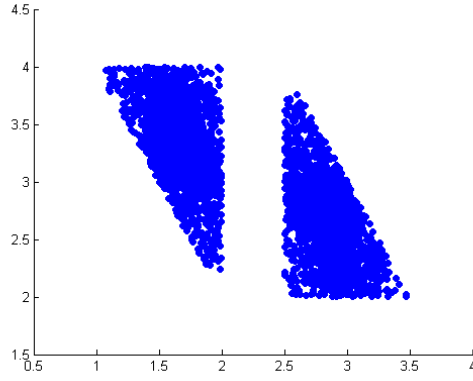


Figure 6.1: The **D2EC2M3** synthetic data with two elongated convex hulls.

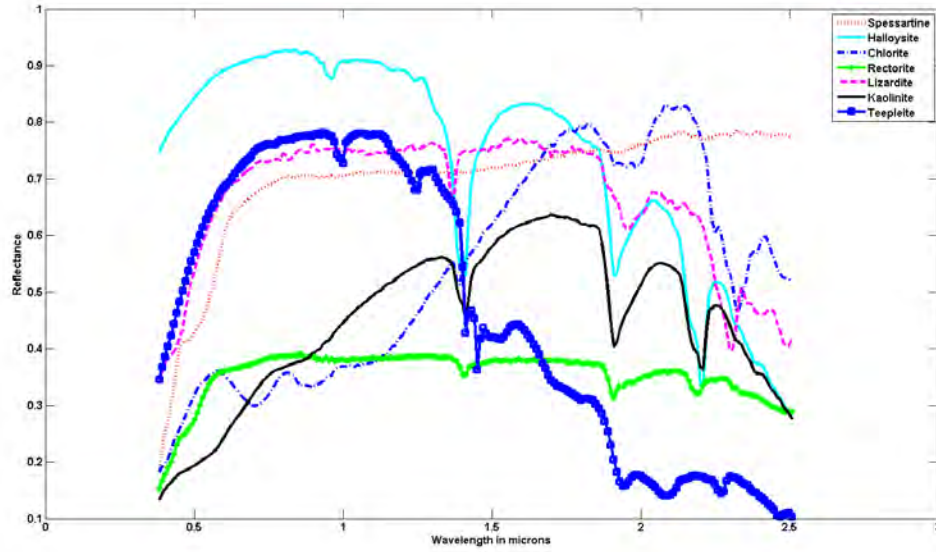


Figure 6.2: USGS spectra used to generate the synthetic hyperspectral data

6.1.1.2 Synthetic hyperspectral data

These are generated using the United States Geological Survey (USGS) digital spectral library [88]. The library contains spectra of 423 minerals, 17 plants and some miscellaneous materials. The spectra have 224 spectral bands, spanning the $0.383 - 2.508 \mu m$ wavelength range. The minerals *Spessartine*, *Halloysite*, *Chlorite*, *Rectorite*, *Lizardite*, *Kaolinite* and *Teepleite*, shown in figure 6.2, are used to generate the data. A total of seven data sets are generated:

- **Usgs1C2M3**: it has two convex sets. Each convex set is generated using the linear model in (1.1), using three different endmembers. Spectra of the minerals *Spessartine*, *Halloysite*,

Chlorite, *Rectorite*, *Lizardite* and *Kaolinite* are used to generate 1000 spectral signatures. The first three endmembers are used to generate the first convex region, and the last three endmembers are used to generate the second one, each having 500 points. The proportions for each data point are generated by sampling from a standard uniform distribution and are normalized to sum to one. Zero-mean Gaussian noise is added to the simulated spectra at three noise levels resulting in 3 data sets. The noise levels are adjusted by changing the variance of the Gaussian to obtain Signal to Noise Ratios (SNR) of 20 dB, 30 dB and 50 dB for the three levels. The SNR is defined using the logarithmic decibel scale as:

$$SNR = 10 \log_{10} \left(\frac{P_{data}}{P_{noise}} \right), \quad (6.1)$$

where P_{data} is the average power of the data, and P_{noise} is the average power of the noise.

- **Usgs2C2M3:** it has two convex sets. Each convex set is generated using the linear model in (1.1), using three different endmembers. The six selected endmembers correspond to the minerals *Spessartine*, *Halloysite*, *Chlorite*, *Rectorite*, *Kaolinite* and *Teepleite*. The first three endmembers are used to generate the first convex region, and the last three endmembers are used to generate the second one, each having 1000 points. The proportions were randomly generated from a standard uniform distribution and were normalized to sum to one. We randomly select few points from each set and add a zero-mean Gaussian noise to them, at a signal to noise ratio (SNR) of 1 dB. We experiment with noise levels of 0%, 5% and 10%.
- **Usgs3C1M4:** it has one convex set of 2000 points generated according to equation (1.1), using 4 endmembers (*Spessartine*, *Halloysite*, *Chlorite* and *Lizardite*). The proportions are linearly generated between 0 and 0.9 for each endmember. A zero-mean Gaussian noise at a signal-to-noise ratio (SNR) of 20 dB was added to the data. This data has been used in Section 4.1 as a motivational example for the robust unmixing using consensus analysis.

6.1.2 Real data

Three real data sets were used in the experiments: the Pavia University data, the University of Southern Mississippi data and the Indian Pines data:

- **Pavia University data:** the data was collected on July 8, 2002 over an urban area around the Pavia University in northern Italy, using the Reflective Optics System Imaging Spectrometer¹

¹Data available at http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes

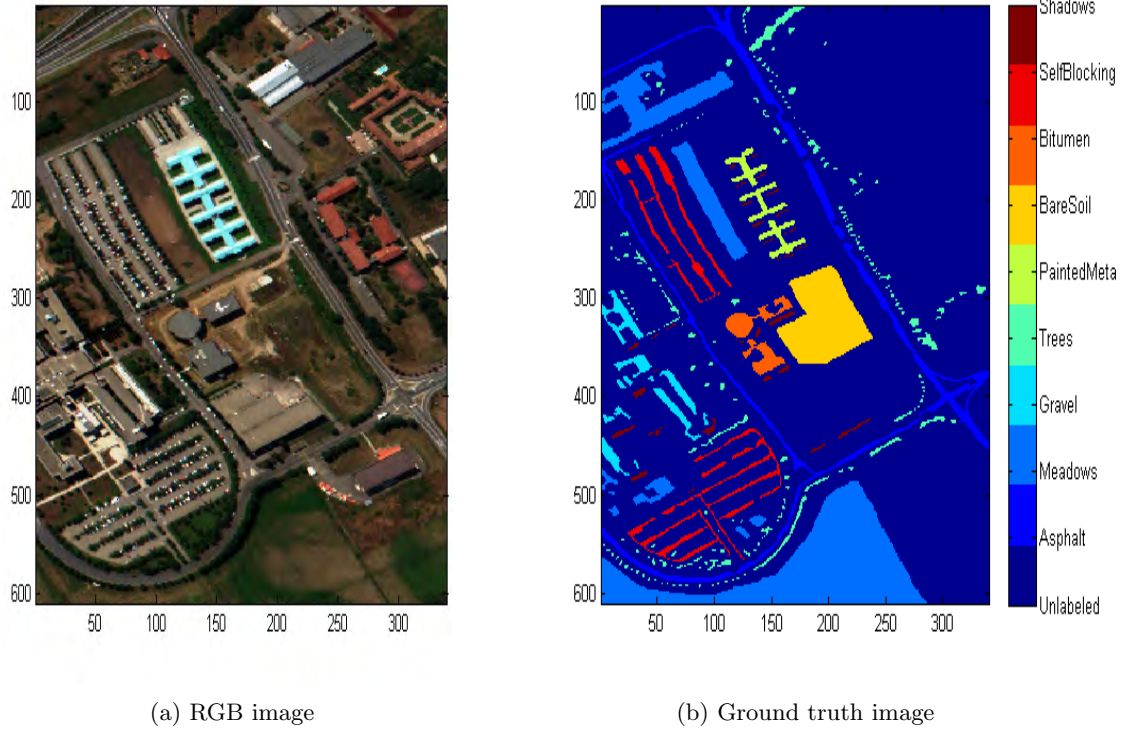


Figure 6.3: **Pavia University** data

(ROSIS). The ROSIS sensor collects data over the 430 - 860 nm wavelength range at a 4 nm spectral sampling interval. The image originally contains 610×340 pixels having 103 spectral bands each (430 - 838 nm), with a spatial resolution of 1.3 meters. The scene consists of both natural and urban regions as shown in figure 6.3(a). Ground truth labels are provided for some areas of the scene. Nine classes are defined: asphalt, meadows, gravel, trees, painted metal sheets, bare soil, bitumen, self-blocking bricks and shadows. Figure 6.3(b) shows the labeled pixels of this image.

- University of Southern Mississippi data:** Two data sets are available for this site [91]. The first one is a LIDAR image acquired using an Optech Inc. Gemini Airborne Topographic LIDAR Mapper (ALTM) system. The second one is a hyperspectral image acquired using an ITRES Inc. hyperspectral Compact Airborne Spectrographic Imager (CASI-1500), which measured reflectance in 72 spectral bands across the Visible and Near-Infrared (VISNIR) wavelengths (375 - 1050 nm) at a 10 nm resolution. The first 4 bands were removed due to the presence of negative values, leaving 68 bands. Both data sets were geometrically corrected to be co-registered with $1m \times 1m$ pixels. The images, originally of size 325×337 pixels, were

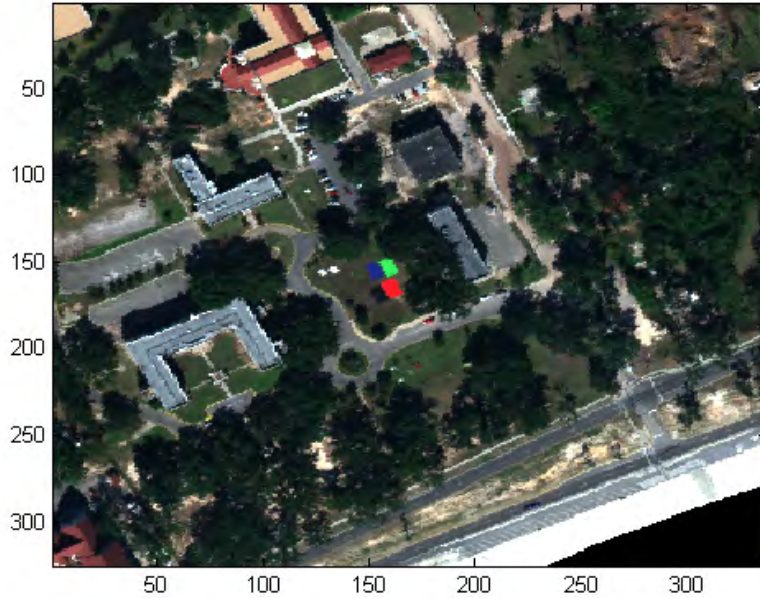


Figure 6.4: RGB image of the **University of Southern Mississippi** data

acquired over the campus of the University of Southern Mississippi in Gulfport, Mississippi, in November 2010. The scene consists of both natural and urban regions as shown in figure 6.4.

- **Indian Pines data:** the data¹ consists of a 145×145 pixels image, having 224 spectral bands covering the $0.4 - 2.5 \mu m$ wavelength range. It was collected by the AVIRIS sensor over the Indian Pines test site in North-western Indiana. The scene, shown in figure 6.5, contains two-thirds agriculture, and one-third forest or other natural perennial vegetation. There are two major dual lane highways, a rail line, as well as some low density housing, other built structures, and smaller roads. The number of bands was reduced to 200 by removing bands covering the region of water absorption (bands 104 - 108, 150 - 163, and 220).

6.2 Context Dependent Spectral Unmixing

In this section, we present the results of the proposed CDSU algorithm on synthetic and real data sets. We also provide a comparison to the results of the P-COMMEND algorithm. For consistency, both algorithms were initialized similarly. The parameters were experimentally determined for both algorithms.

¹Data available at http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes

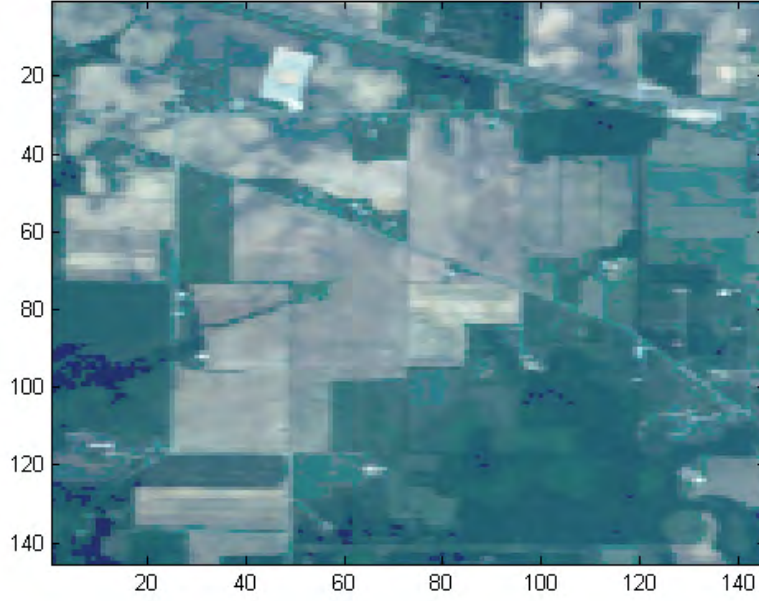


Figure 6.5: RGB image of the **Indian Pines** data

6.2.1 Evaluation using simulated data

First, we present the results of our proposed approach on the three toy data sets (**D2C1M3**, **D2C2M3** and **D2C3M3**) described in Section 3.1. We then apply our method to the **Usgs1C2M3** simulated hyperspectral data sets described in Section 6.1.1.2. Due to the simplicity of the toy data sets, the performance of the algorithms is evaluated visually. However, for the simulated data set, the performance is evaluated quantitatively.

In the first example, we use the **D2C1M3** data set of Section 3.1.1 (displayed in figure 3.1(a)). The result of the CDSU algorithm on this data set, using $C = 1$, $M = 3$, $m = 2$, $\alpha = 1$, and $\beta = 0.01$ is shown in figure 6.6, where the detected endmembers are shown in red X's. As it can be seen, the CDSU algorithm succeeded in identifying endmembers at the vertices of the convex hull enclosing the data points providing a tight fit around them.

As a second example, we use the **D2C2M3** data set of Section 3.1.2 (displayed in figure 3.2(a)). The result of the CDSU algorithm on this data set, using $C = 2$, $M = 3$, $m = 2$, $\alpha = 20$ (α has been increased, compared to **D2C1M3**, in order to account for the greater number of data points in **D2C2M3**), and $\beta = [0.01, 0.01]$ is shown in figure 6.7. The detected endmembers are shown in red X's for one cluster and in green X's for the other cluster. As it can be seen, the CDSU algorithm succeeded in identifying endmember sets at the vertices of the convex hulls enclosing the

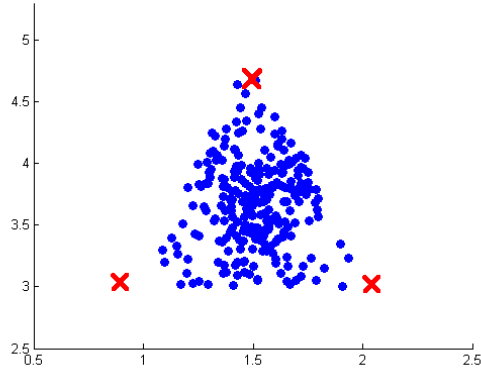


Figure 6.6: Result of the CDSU algorithm on the **D2C1M3** data.

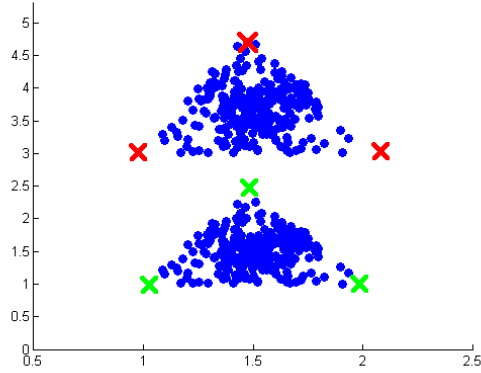


Figure 6.7: Result of the CDSU algorithm on the **D2C2M3** data.

data points and thus, providing a tight fit around them.

For a third example, we use the **D2C3M3** data set of Section 3.1.3 (displayed in figure 3.3(a)). The result of the CDSU algorithm on this data set, using $C = 3$, $M = 3$, $m = 2$, $\alpha = 50$ (α has been increased, compared to **D2C2M3**, in order to account for the greater number of data points in **D2C3M3**), and $\beta = [0.1, 0.1, 0.1]$ (β has been increased, compared to **D2C1M3** and **D2C2M3**, in order to account for the tighter nature of the data in **D2C3M3**) is shown in figure 6.8. The identified clusters are represented using different colors, and the endmembers for each cluster are represented using X's. As it can be seen, the CDSU algorithm succeeded in identifying endmember sets at the vertices of the convex hulls enclosing the data points, providing a tight fit around them. It also provided clusters that match the distribution of the data.

For a fourth example, we consider the **Usgs1C2M3** simulated hyperspectral data set described in Section 6.1.1.2. To evaluate the performance of the CDSU algorithm and compare it to P-COMMEND, the estimated abundance fractions and endmembers are compared to the true

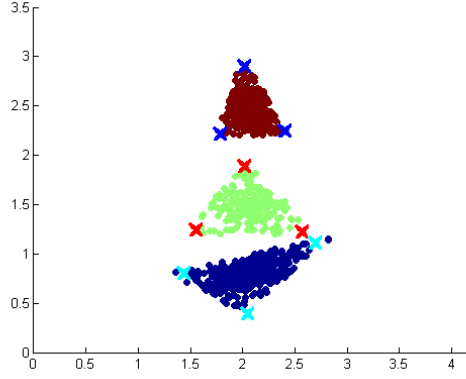


Figure 6.8: Result of the CDSU algorithm on the **D2C3M3** data.

ones. Based on the mean square error (MSE), we define the spectral mean error (SME) and the abundance mean error (AME) as

$$SME \equiv \frac{1}{Md} \|\mathbf{E} - \hat{\mathbf{E}}\|_F^2, \quad (6.2)$$

and

$$AME \equiv \frac{1}{MN} \|\mathbf{P} - \hat{\mathbf{P}}\|_F^2. \quad (6.3)$$

In (6.2) and (6.3), M is the total number of endmembers, d is the number of spectral bands, and N is the number of data points. The rows of $\mathbf{E}$ and $\hat{\mathbf{E}}$ represent the true endmembers and the learned endmembers respectively. $\mathbf{P}$ is a $N \times M$ matrix representing the true endmember abundance fractions. $\hat{\mathbf{P}}$ is a $N \times M$ matrix representing the estimated endmember abundance fractions, where each proportion value is multiplied by the corresponding cluster membership. The notation $\|\cdot\|_F$ stands for the Frobenius norm.

Another common performance metric is the spectral angle distance, which measures the angle between a signature $\mathbf{e}_i$ and its estimate $\hat{\mathbf{e}}_i$ [1]. Based on this metric, we define a spectral mean angle error ($SMAE$) as

$$SMAE \equiv \sqrt{\frac{1}{M} \sum_{i=1}^M \left[\arccos \left(\frac{\mathbf{e}_i \hat{\mathbf{e}}_i^T}{\|\mathbf{e}_i\| \|\hat{\mathbf{e}}_i\|} \right) \right]^2}. \quad (6.4)$$

It is clear that the performance of the algorithms increases as SME , AME , and $SMAE$ approach zero. Notice, however, that the estimates of $\mathbf{E}$ and $\mathbf{P}$ are up to a permutation matrix. Thus, a simple algorithm based on the Hungarian method [92] has been designed and used to infer the permutation matrix.

We run the CDSU and P-COMMEND algorithms using the parameters in table 6.1. We run each

TABLE 6.1

Parameters used for the CDSU and P-COMMEND algorithms on the USGS simulated hyperspectral data

Parameters	CDSU	P-COMMEND
C	2	2
M	3	3
m	1.25	1.25
α, β	$\alpha = 200, \beta = [0.1, 0.1]$	$\alpha = 0.1$
Stopping criterion ($Iter$ = Iteration number)	$abs\left(\frac{J_{CDSU}(Iter+1)-J_{CDSU}(Iter)}{J_{CDSU}(Iter+1)}\right) < 10^{-6}$	$abs\left(\frac{J_{PCOMMEND}(Iter+1)-J_{PCOMMEND}(Iter)}{J_{PCOMMEND}(Iter+1)}\right) < 10^{-6}$

algorithm 25 times at each noise level. For each run, we use the FCM [72] and MVSA [70] algorithms for the initialization of the memberships, centers and endmembers. This experiment is designed to test the sensitivity of CDSU to noise and initialization, and to provide a comparison to P-COMMEND.

Figures 6.9(a), (b) and (c) show a box plot of the different error metrics of both CDSU and P-COMMEND algorithms across the 25 runs and at all noise levels. On each box, the central red mark represents the median value, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme values not considered outliers, and outliers are plotted individually using red crosses. An outlier is a value smaller than the first quartile minus 1.5 times the interquartile range (third minus first quartile), or higher than the third quartile plus 1.5 times the interquartile range.

Examining these results, we can see that, for low noise level cases (SNR = 30 and 50 dB), CDSU and P-COMMEND perform similarly. Both algorithms are also robust to initialization. As we increase the noise level (SNR = 20 dB), CDSU starts to outperform P-COMMEND (p -value=1.4e-09 using the Wilcoxon rank sum test [93]). An explanation for this would be that the clustering term of the CDSU objective function makes it more robust to noise than P-COMMEND, since noise does not affect clustering as much as it would affect spectral unmixing. Furthermore, we see that the error metric increases for both algorithms as the noise level increases, which is expected.

To compare and illustrate the results further, we visualize the estimated endmembers by both algorithms for the case of the highest noise level (SNR = 20 dB). We pick the run in which P-COMMEND gave the highest error. Figure 6.10(a) and 6.10(b) show the true (solid lines) and estimated (dashed lines) endmembers resulting from the CDSU and P-COMMEND algorithms respectively. As it can be seen, CDSU resulted in better estimates of the endmembers.

To analyze the results further, we check the cluster assignments, based on the highest membership values, generated by P-COMMEND and CDSU. As it can be seen in table 6.2, both algorithms

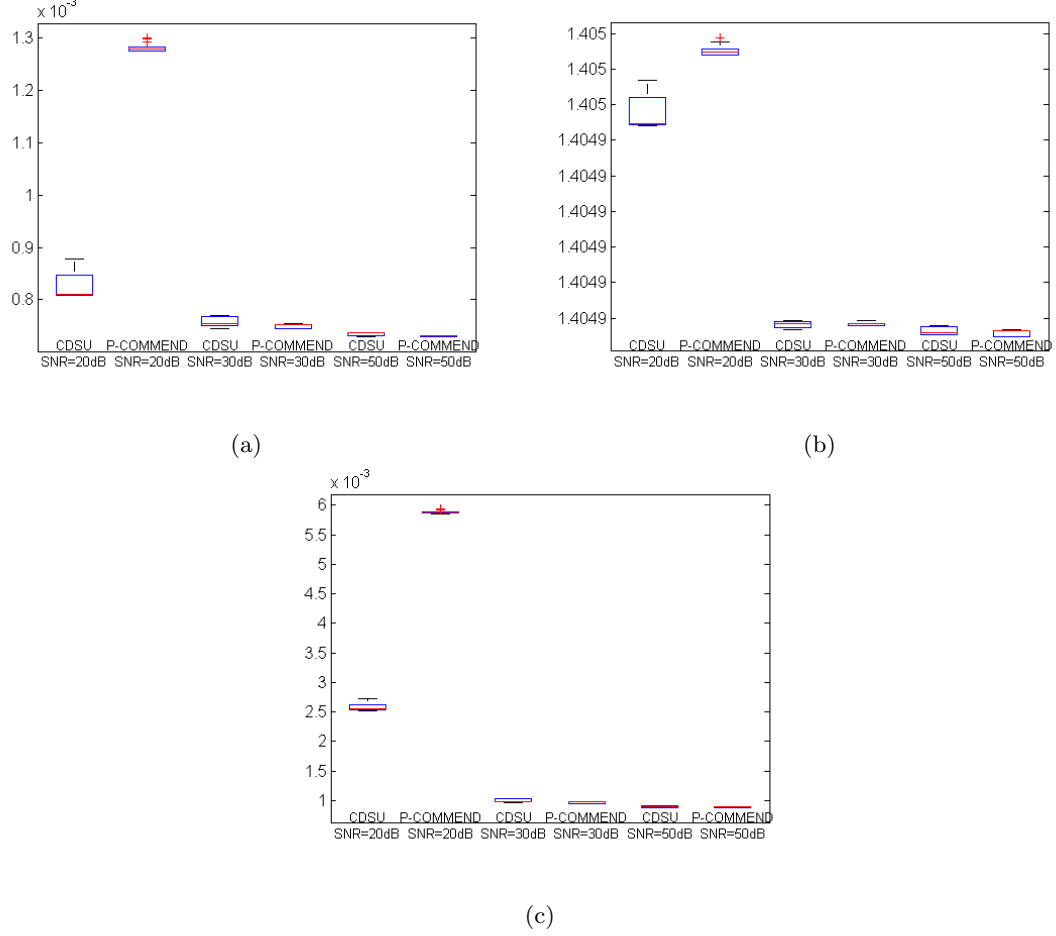


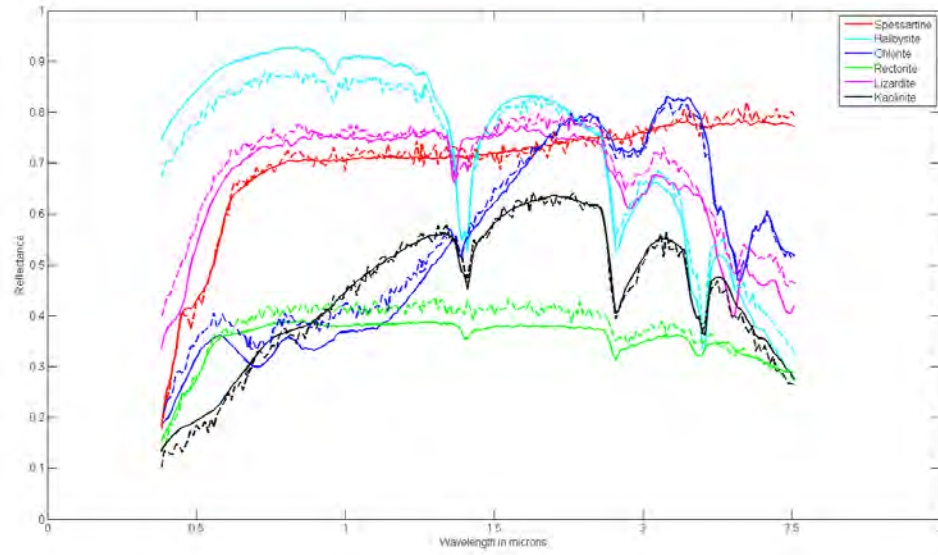
Figure 6.9: Error metrics for CDSU and P-COMMEND on the **Usgs1C2M3** data across the 25 runs and at all noise levels: (a) *SME*, (b) *SMAE*, (c) *AME*.

TABLE 6.2

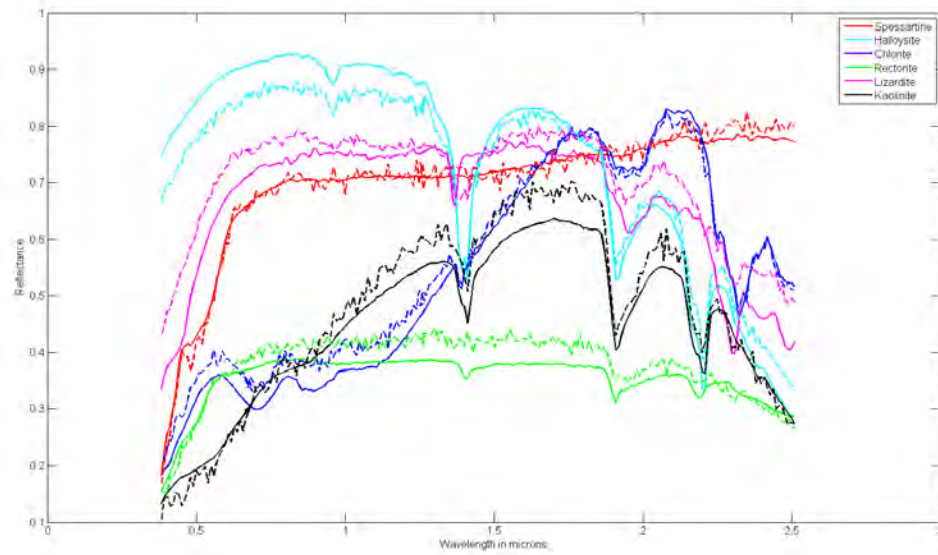
Composition of the clusters generated by CDSU and P-COMMEND on the **Usgs1C2M3** data with $\text{SNR} = 20 \text{ dB}$

	Convex set 1	Convex set 2
Cluster 1	500	0
Cluster 2	0	500

succeeded in generating clusters that are pure and that correspond to the two original convex regions. To analyze the fuzzy partition, we analyze the membership values of the data points in each cluster. We plot the two first principal components of the data and we color each point with a shade corresponding to its membership value in a specific cluster. Figures 6.11(a) and (b) show the membership values in cluster 1 generated by P-COMMEND and CDSU. Similarly, figures 6.12(a) and (b) show the membership values in cluster 2. Compared to CDSU, P-COMMEND assigned higher membership values in cluster 1 to some points from cluster 2 (light blue shades in figure 6.11(b)).



(a) CDSU



(b) P-COMMEND

Figure 6.10: True (solid lines) and estimated (dashed lines) endmembers for the **Usgs1C2M3** data with SNR = 20 dB using CDSU and P-COMMEND.

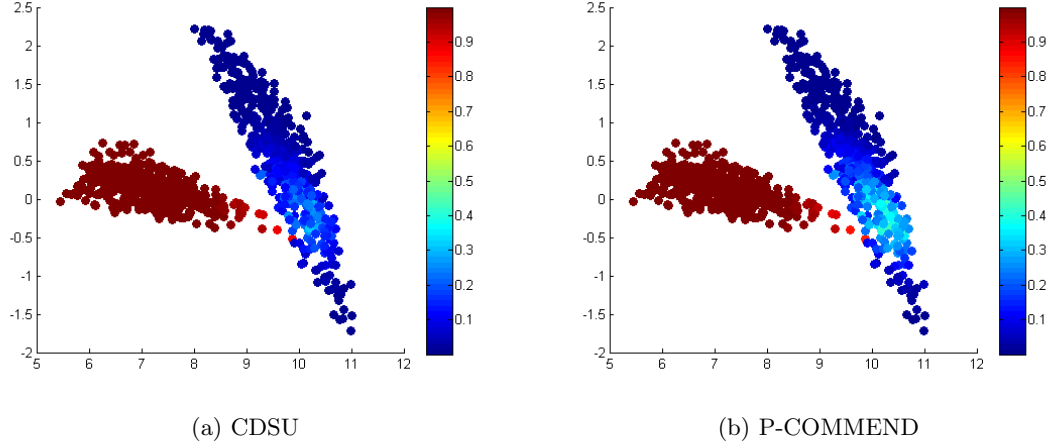


Figure 6.11: Membership values in cluster 1 for the **Usgs1C2M3** data with SNR = 20 dB. Two principal components of the data are scatter plotted.

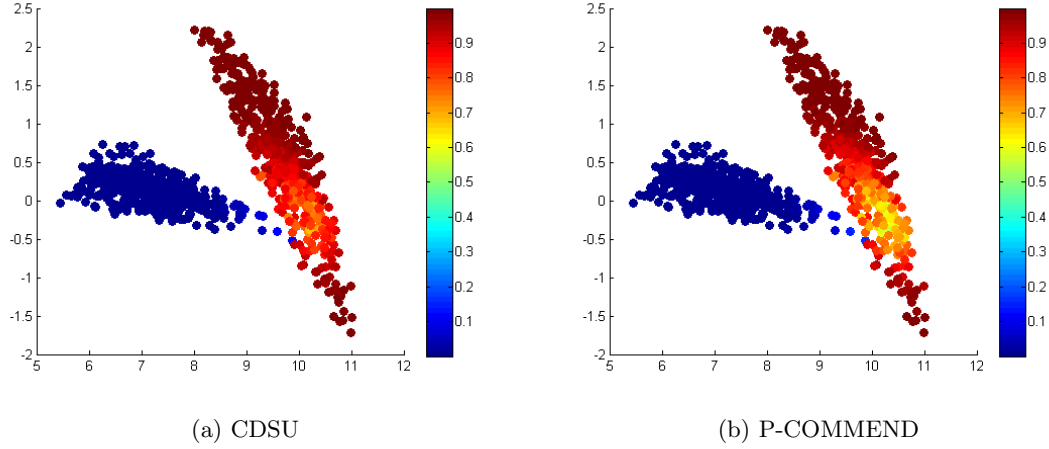


Figure 6.12: Membership values in cluster 2 for the **Usgs1C2M3** data with SNR = 20 dB. Two principal components of the data are scatter plotted.

And due to the sum to one constraint on the memberships, the same points were assigned lower membership values, compared to CDSU, in cluster 2 that they belong to (light green shades in figure 6.12(b)). Since these membership values are used in the update equations of the endmembers and abundances, the parameters estimated using CDSU are more accurate than those estimated using P-COMMEND.

Recall that the update equation of the memberships for P-COMMEND is:

$$u_{ij} = \frac{[(\mathbf{x}_j - \mathbf{p}_{ij}\mathbf{E}_i)(\mathbf{x}_j - \mathbf{p}_{ij}\mathbf{E}_i)^T]^{\frac{1}{1-m}}}{\sum_{q=1}^C [(\mathbf{x}_j - \mathbf{p}_{qj}\mathbf{E}_q)(\mathbf{x}_j - \mathbf{p}_{qj}\mathbf{E}_q)^T]^{\frac{1}{1-m}}}, \quad (6.5)$$

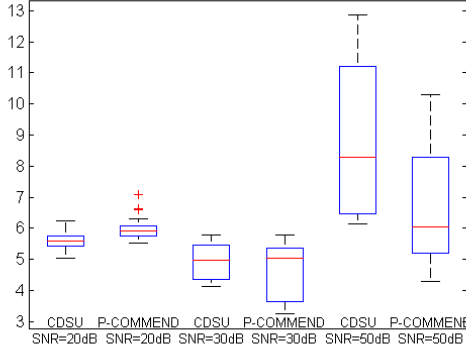


Figure 6.13: Running time (in seconds) of CDSU and P-COMMEND on the USGS simulated data across the 25 runs and at all noise levels.

and the update equation of the memberships for CDSU is:

$$u_{ij} = \frac{[(\mathbf{x}_j - \mathbf{c}_i)(\mathbf{x}_j - \mathbf{c}_i)^T + \alpha(\mathbf{x}_j - \mathbf{p}_{ij}\mathbf{E}_i)(\mathbf{x}_j - \mathbf{p}_{ij}\mathbf{E}_i)^T]^{\frac{1}{1-m}}}{\sum_{q=1}^C [(\mathbf{x}_j - \mathbf{c}_q)(\mathbf{x}_j - \mathbf{c}_q)^T + \alpha(\mathbf{x}_j - \mathbf{p}_{qj}\mathbf{E}_q)(\mathbf{x}_j - \mathbf{p}_{qj}\mathbf{E}_q)^T]^{\frac{1}{1-m}}}. \quad (6.6)$$

By examining (6.6), we notice that a pixel j will have a high membership in cluster i if: (i) its spectrum is close to the centroid, $\mathbf{c}_i$, of that cluster in the feature space; and (ii) it fits the model of that cluster. This is unlike the update equation of the memberships for P-COMMEND in (6.5), where a pixel j will have a high membership in model i only if it fits that model, regardless of the spectral distribution around it. This illustrates the importance of the clustering term in taking the distribution of the data into account and thus leading to better endmember and abundance estimates.

To compare the time complexity of CDSU and P-COMMEND, we generate a box plot of the running time, expressed in seconds, across the 25 runs and at all noise levels. Figure 6.13 shows this box plot. We notice that the CDSU algorithm takes a longer time to run compared to P-COMMEND for the cases where SNR = 30 or 50 dB. The reason for this is that CDSU requires more computations than P-COMMEND. For the SNR = 20 dB case, we see that CDSU becomes slightly faster than P-COMMEND. This can be explained by the clustering term in CDSU which makes it more robust to noise and thus requiring less time to converge.

We use the **Usgs1C2M3** simulated data set, with SNR = 20 dB, to illustrate another advantage of our proposed approach. One might argue why not cluster the data first, then use a single convex region unmixing algorithm, such as ICE, to find the endmembers and abundances of each cluster. In other words, what is the advantage of performing unmixing and clustering

simultaneously.

Figure 6.14 shows the clustering results of applying the FCM algorithm to the simulated data with $\text{SNR} = 20$ dB. We plot the spectral signatures of the two resulting clusters, labeling each point of them with the ground truth, i.e., the convex set it originally came from. We clearly see that the resulting clusters include a mixture of points originally generated from both endmember sets, and hence, any unmixing algorithm applied to each cluster would result in erroneous endmember estimates. We should note here that the erroneous partition is not due to sensitivity of FCM to initialization. In fact, even when we initialize the FCM with the true cluster assignments, it converges to the same result of figure 6.14.

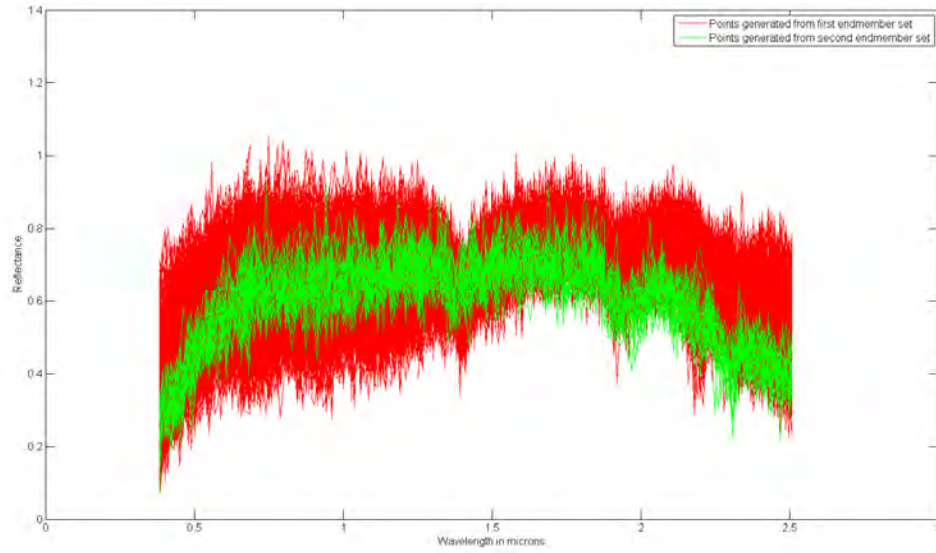
The partition obtained using CDSU is shown in figure 6.15. As it can be seen, unlike the FCM, CDSU partitioned the data into 2 pure clusters even though it was initialized with the FCM. Consequently, using this correct partition, it is more likely to estimate the correct endmembers.

In order to understand why FCM leads to mixed clusters whereas CDSU succeeds to return pure ones, we track the objective function values of both algorithms as they run. Figure 6.16 shows the evolution of the objective function of the FCM until its convergence. As it can be seen, it reaches a local minimum in few iterations. For the CDSU, we track the first term of the objective function in (3.7) (same as the FCM objective function), and the sum of all terms. As it can be seen in figure 6.17, the sum of the terms decreases as the algorithm evolves. However, the first term increases. In other words, the optimal partition is the one that takes into account both data partitioning and model fitting.

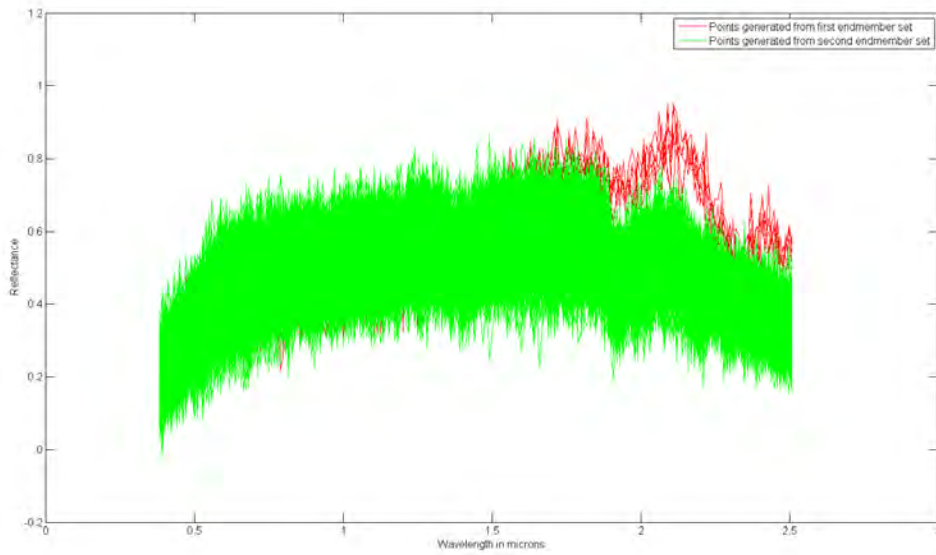
Using two dimensional toy data and simulated hyperspectral data, we have illustrated the effectiveness of the proposed CDSU algorithm in identifying correct abundances and endmember sets. We have also shown that CDSU outperformed P-COMMEND in the case of noisy data. Furthermore, we showed the difference between our approach, that performs spectral unmixing while simultaneously taking into account the distribution of the data in the spectral space, and simply clustering the data and then unmixing each cluster separately. In the next section, we present the results of CDSU on a real hyperspectral data and compare them to those of P-COMMEND.

6.2.2 Evaluation using real data

In this section, we consider a down sampled version, of size 305×170 pixels, of the **Pavia University** real hyperspectral data set described in Section 6.1.2. We run the CDSU and P-COMMEND algorithms on this data set using the parameters in table 6.3.

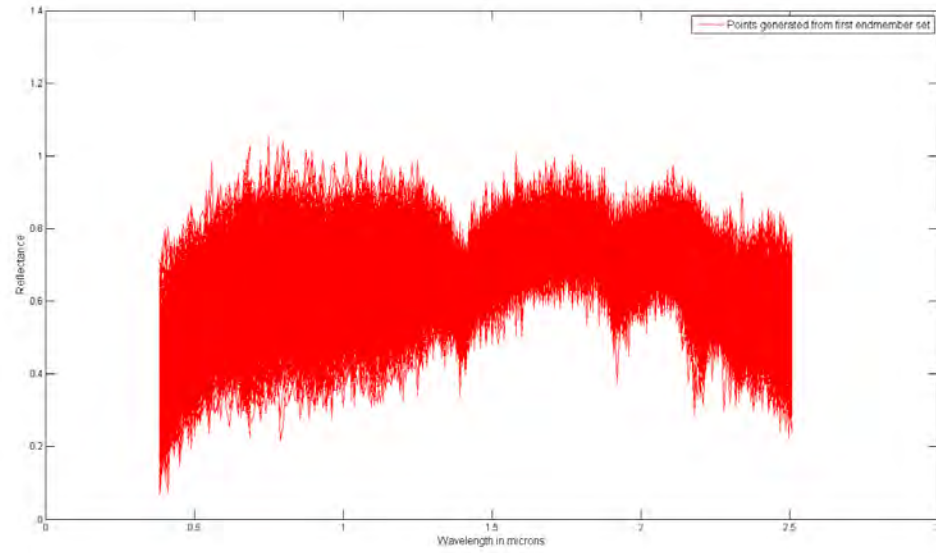


(a) Cluster 1

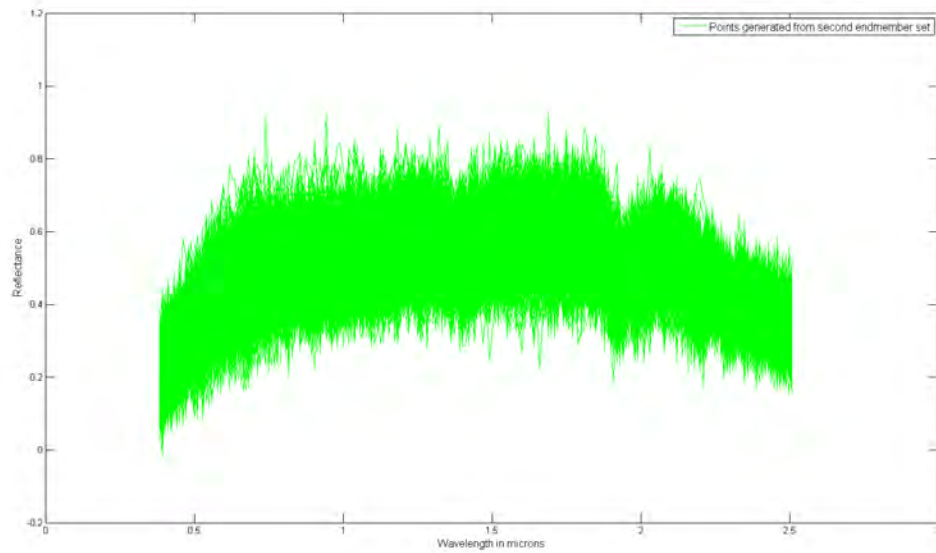


(b) Cluster 2

Figure 6.14: Spectral signatures of the clusters generated by the FCM algorithm on the **Usgs1C2M3** data with SNR = 20 dB.



(a) Cluster 1



(b) Cluster 2

Figure 6.15: Spectral signatures of the clusters generated by the CDSU algorithm on the **Usgs1C2M3** data with $\text{SNR} = 20$ dB.

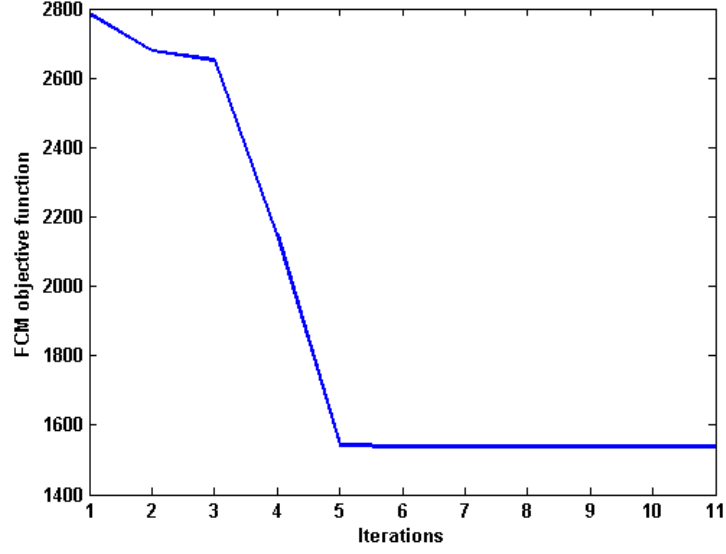


Figure 6.16: Evolution of the FCM objective function on the **Usgs1C2M3** data (SNR = 20 dB).

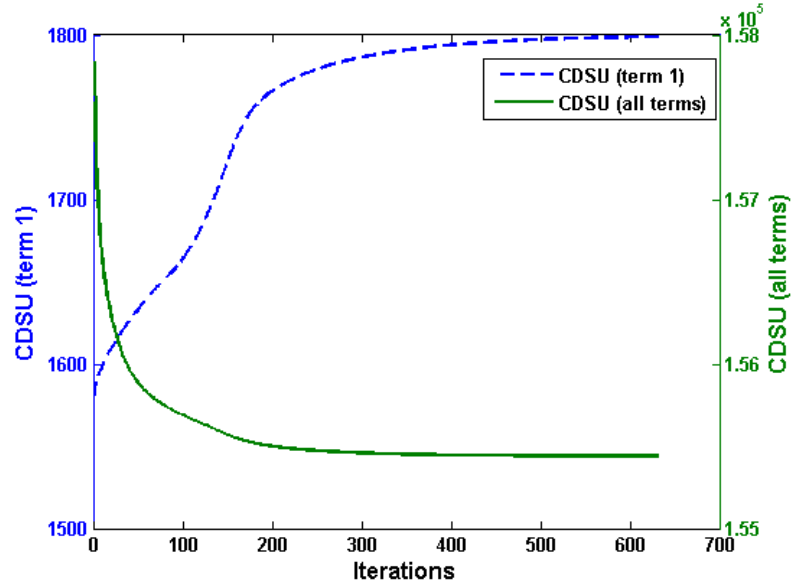


Figure 6.17: Evolution of the objective function of CDSU on the **Usgs1C2M3** data (SNR = 20 dB).

TABLE 6.3

Parameters used for the CDSU and P-COMMEND algorithms on the **Pavia University** data

Parameters	CDSU	P-COMMEND
C	3	3
M	3	3
m	1.25	1.25
α, β	$\alpha = 1, \beta = [5, 5, 5]$	$\alpha = 5$
Stopping criterion ($Iter$ = Iteration number)	$abs\left(\frac{J_{CDSU}(Iter+1) - J_{CDSU}(Iter)}{J_{CDSU}(Iter+1)}\right) < 10^{-6}$	$abs\left(\frac{J_{PCOMMEND}(Iter+1) - J_{PCOMMEND}(Iter)}{J_{PCOMMEND}(Iter+1)}\right) < 10^{-6}$

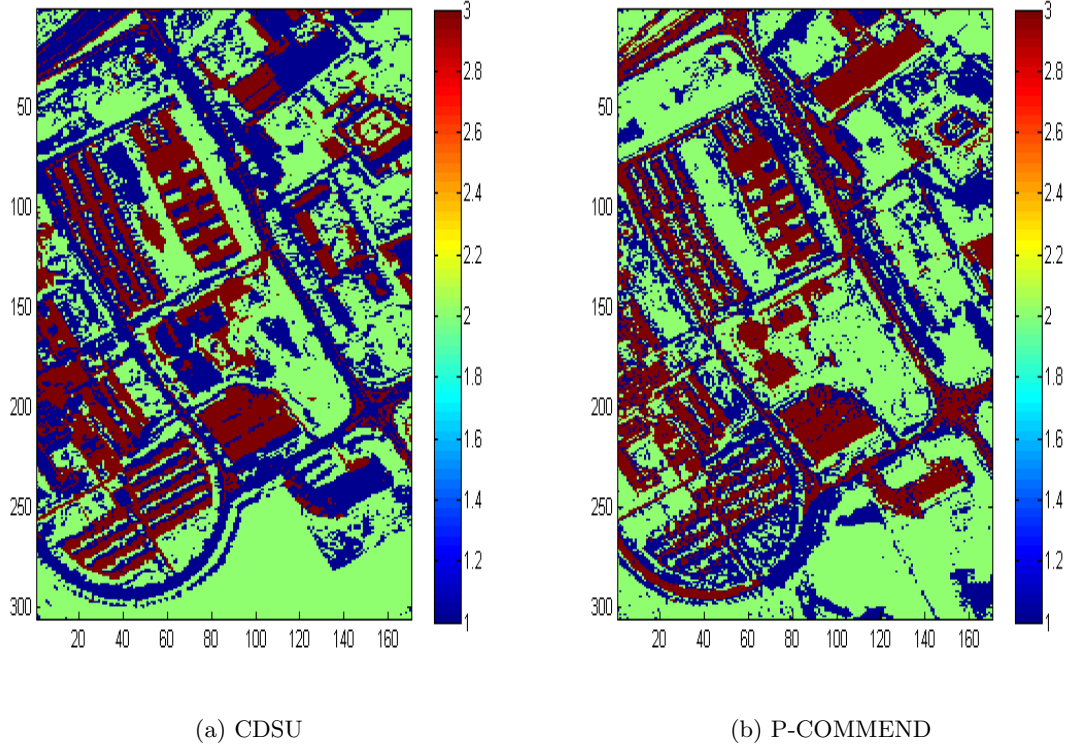


Figure 6.18: Cluster assignments of CDSU and P-COMMEND on the **Pavia University** data

TABLE 6.4

Composition of the clusters generated by P-COMMEND and CDSU, in percentage, for the **Pavia University** data, taking into account the labeled points only

Class	Asphalt	Meadows	Gravel	Trees	Painted metal sheets	Bare soil	Bitumen	Self blocking bricks	Shadows
Cluster									
P-COMMEND									
Cluster 1	16.01	22.78	3.19	35.36	0	7.14	0.19	15.30	0
Cluster 2	0.57	76.74	0.03	0.51	0	17.74	0.01	0.18	4.17
Cluster 3	40.89	0	14.58	0.03	10.66	4.71	10.38	18.66	0.06
CDSU									
Cluster 1	45.54	21.46	3.98	0.24	0.03	10.29	9.23	2.32	6.88
Cluster 2	0.05	73.08	0	14.71	0	12.14	0	0	0
Cluster 3	6.45	6.17	18.32	0.04	15.49	13.36	1.16	38.97	0

Figures 6.18(a) and (b) show the images of cluster assignments for both algorithms. Cluster assignments were based on the highest membership values. As it can be seen, the clusters generated by CDSU are spatially more coherent.

Table 6.4 shows the compositions of the clusters generated by P-COMMEND and CDSU, in percentage, taking into account the labeled points only. The same compositions are illustrated visually in figures 6.19(a) and (b). As it can be seen, the “Trees”, “Painted Metal Sheets”, “Bitumen” and “Shadows” classes were assigned to unique clusters by both algorithms. On the other hand, the “Meadows”, “Gravel” and “Bare soil” classes were divided into more than one cluster by both algo-

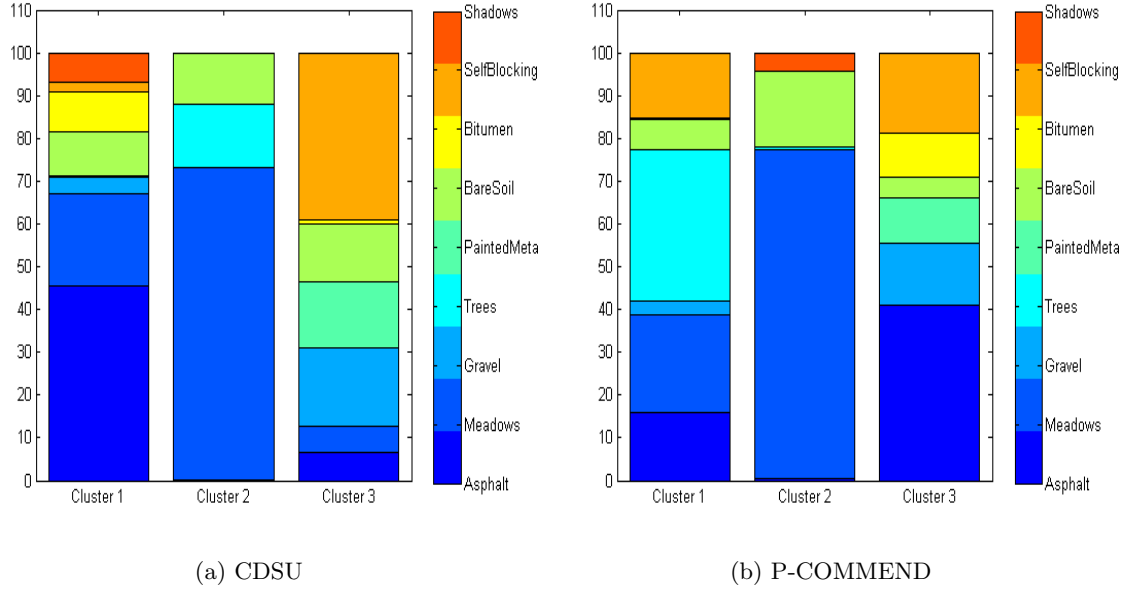


Figure 6.19: Composition, in percentage, of the clusters generated by CDSU and P-COMMEND on the **Pavia University** data, taking into account the labeled points only.

rithms. Finally, the “Asphalt” and “Self blocking bricks” classes were divided into more than one cluster by P-COMMEND and less divided by CDSU. This is mainly due to the spectral similarity that CDSU takes into account when partitioning the data, unlike P-COMMEND.

The proportion maps associated with the three endmembers for each of the three clusters found by P-COMMEND and CDSU are shown in figures 6.20 and 6.21, respectively. Each pixel in the proportion maps was multiplied by the corresponding membership value for each context so that the corresponding endmember with high proportion is highlighted. Since we do not have labels for the entire image, we analyze these proportion maps by comparing them to the RGB image in figure 6.3(a), labeling each map with the dominant material it represents.

It can be seen that the endmembers found by P-COMMEND are duplicated in more than one cluster. For instance, endmembers corresponding to shadows, tall grass and trees are found in cluster 1 and 2. Furthermore, the endmember for cement is found in clusters 1 and 3. Moreover, some endmembers represent non coherent elements, like combining asphalt and bitumen with cement in one endmember. In contrast, the endmembers found by CDSU are not duplicated in the three contexts and they represent coherent elements.

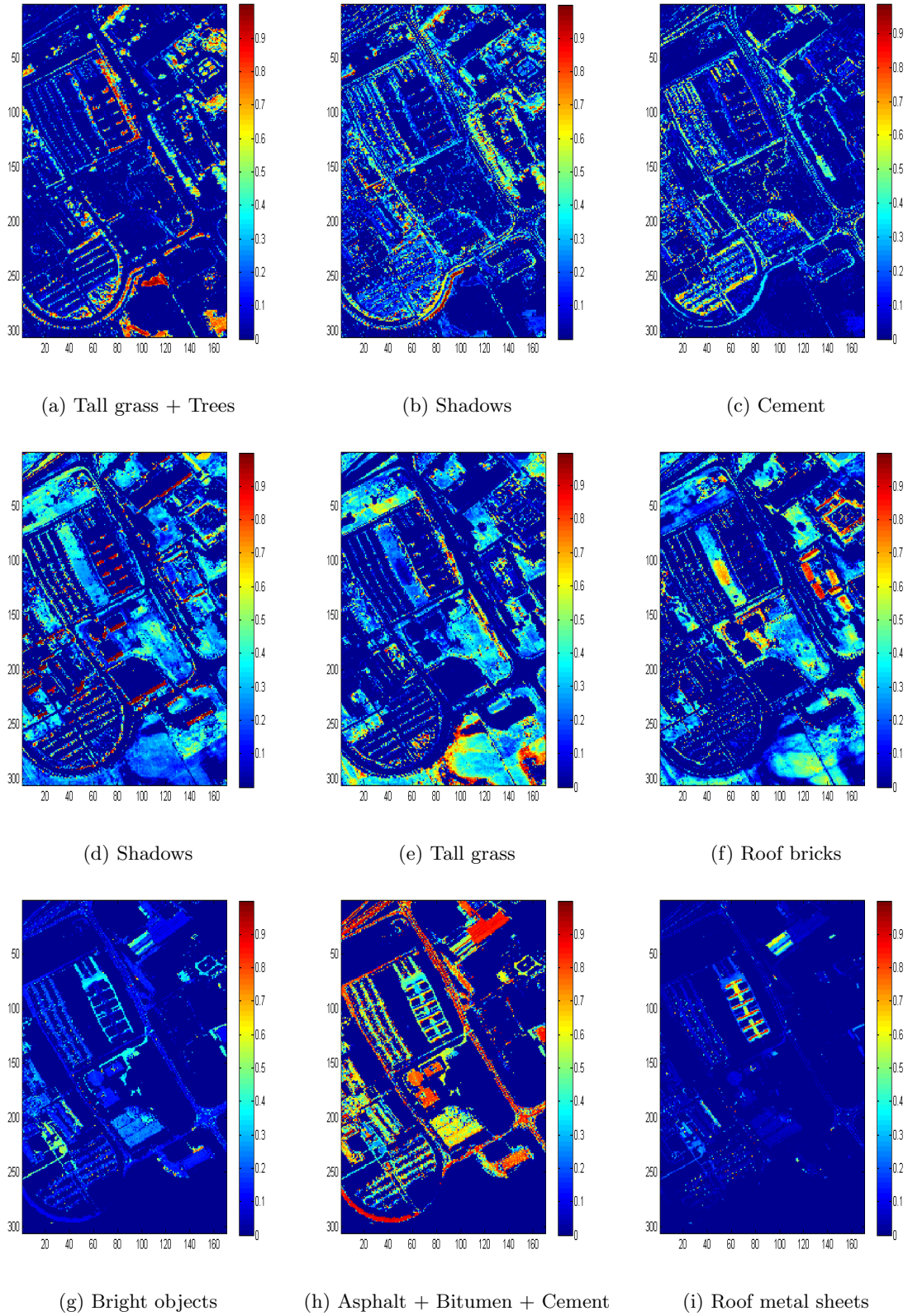


Figure 6.20: Proportion maps estimated by the P-COMMEND algorithm for the **Pavia University** data. Each row of 3 proportion maps represents one context, and each column represents an endmember in that context.

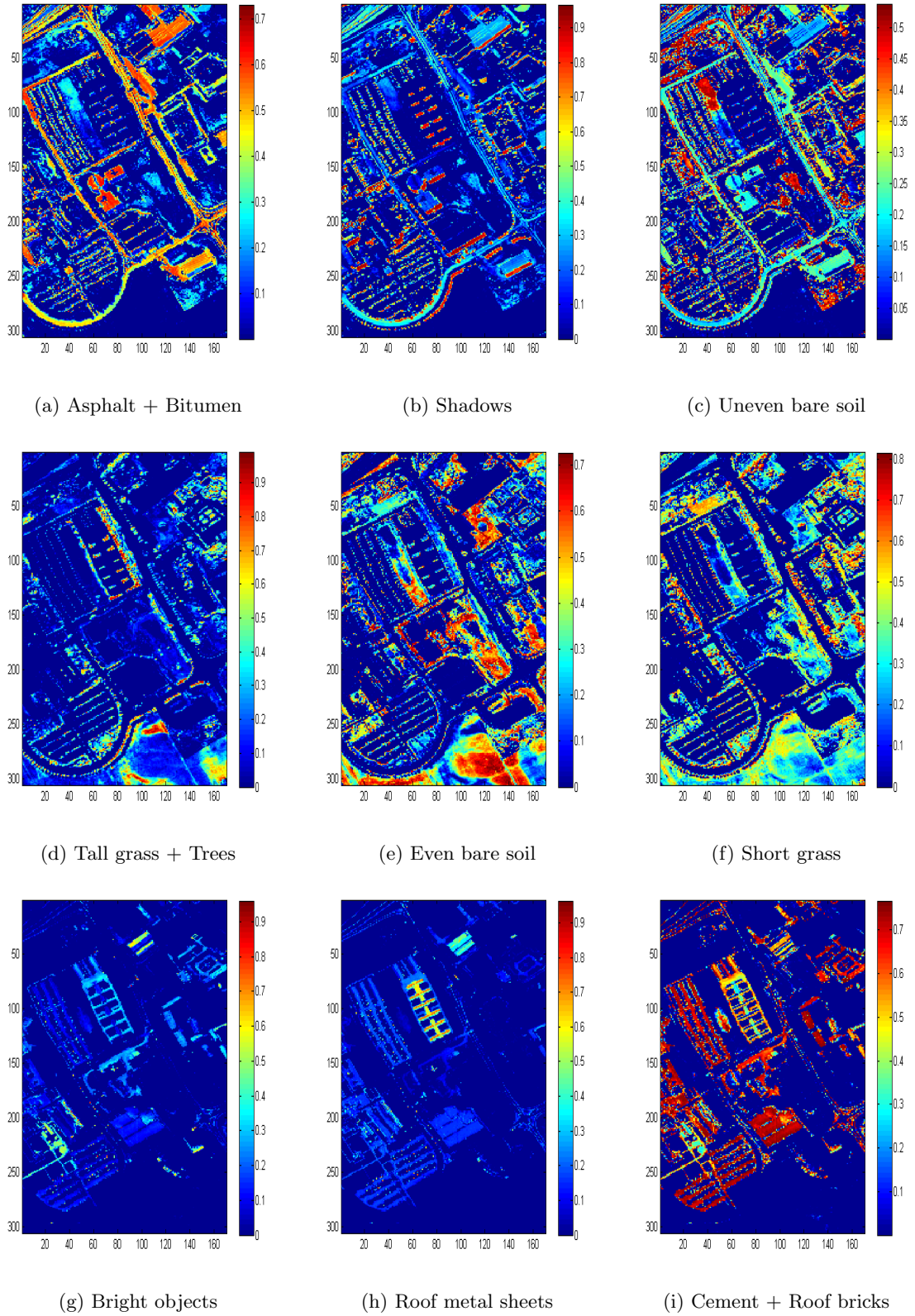
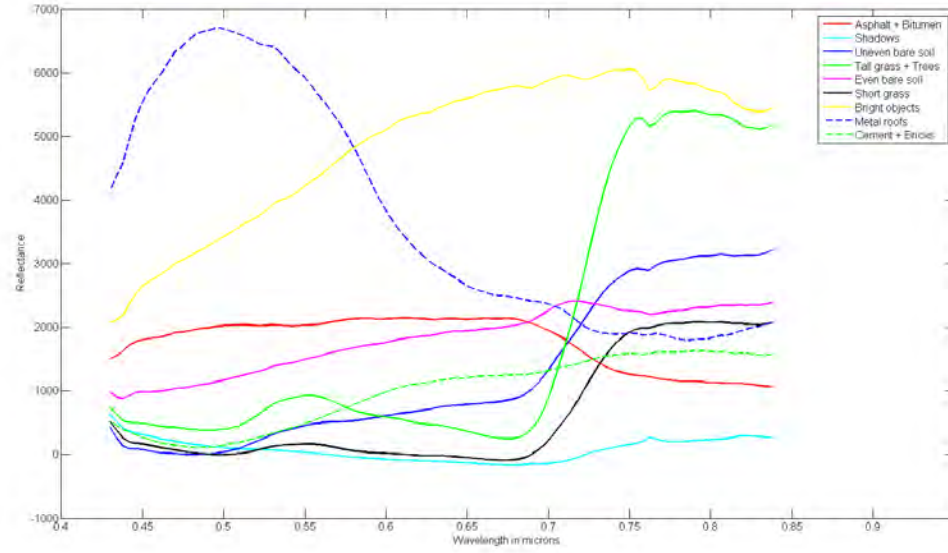
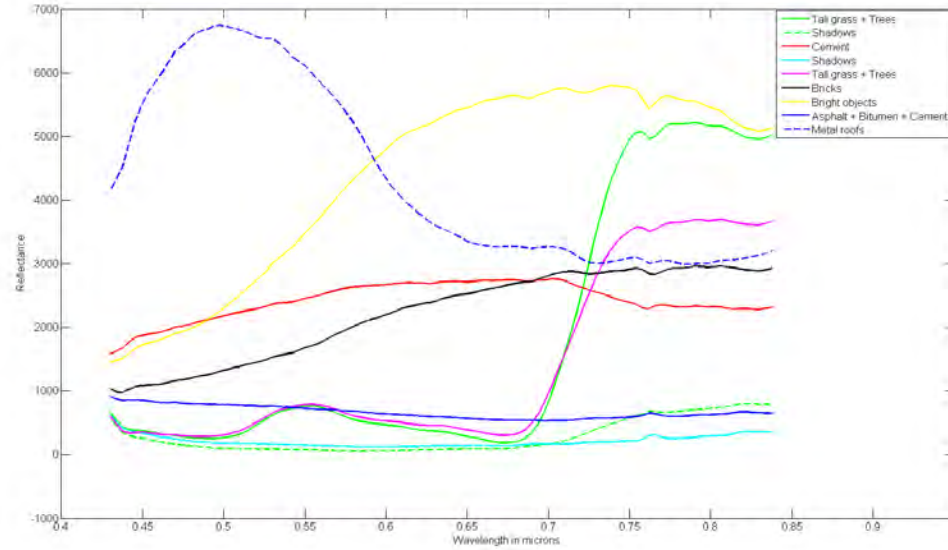


Figure 6.21: Proportion maps estimated by the CDSU algorithm for the **Pavia University** data. Each row of 3 proportion maps represents one context, and each column represents an endmember in that context.

Figures 6.22(a) and (b) show the estimated endmembers by CDSU and P-COMMEND, respectively. As it can be seen, both algorithms identified similar endmembers like the ones corresponding to metal roofs, bright objects, shadows, and tall grass and trees. However, P-COMMEND generated an almost duplicate endmember for shadows, tall grass and trees. P-COMMEND also missed endmembers for bare soil that CDSU was able to identify.



(a) CDSU



(b) P-COMMEND

Figure 6.22: Estimated endmembers for the **Pavia University** data using CDSU and P-COMMEND.

In this section, we presented the results of the proposed CDSU algorithm on real hyperspectral data and compared them to those of P-COMMEND. Both methods did agree on some endmembers. However, while P-COMMEND resulted in some duplicate and non coherent endmembers, CDSU identified coherent and diverse endmembers representing meaningful elements in the scene.

6.3 Context Dependent Spectral Unmixing Using the Mahalanobis Distance

The Context Dependent Spectral Unmixing using the Mahalanobis distance (CDSU_M) is a variation of CDSU that allows more flexibility in the shape of the clusters. It accounts for ellipsoidal shapes beside the traditional spherical shape. The experiment in this section is designed to illustrate the advantage of CDSU_M over CDSU using a simple toy data set.

We use the **D2EC2M3** synthetic data set described in Section 6.1.1.1. CDSU and CDSU_M were ran using the same parameters values: $C = 2$, $M = 3$, $m = 2$, $\alpha = 40$, and $\beta = [0.2, 0.2]$. For CDSU_M, σ_i was set to 1 for $i = 1, 2$.

Figure 6.23 illustrates the results of both algorithms. The retrieved clusters are represented in different colors, and the endmembers are represented in red and green X's. Due to the “non-spherical” nature of the clusters, CDSU failed to identify the correct cluster assignments. This made it fail to identify appropriate endmembers. On the other hand, CDSU_M successfully identified both the clusters and the endmembers.

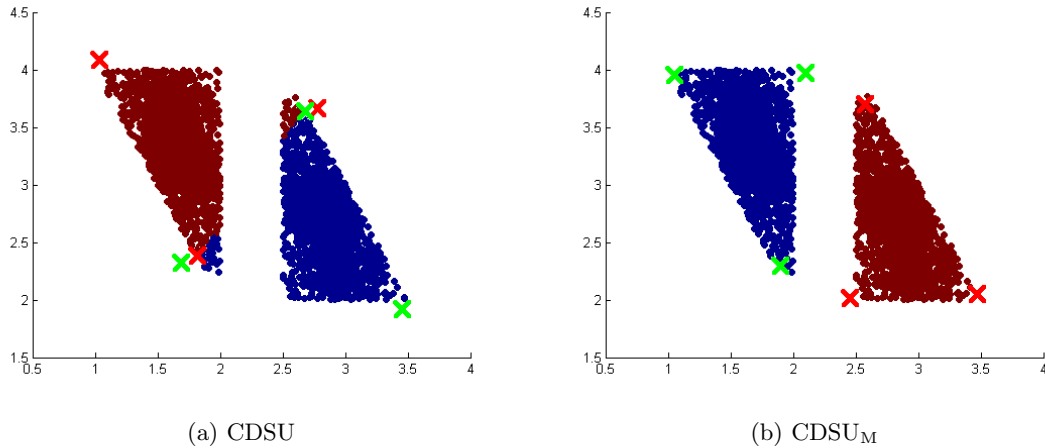


Figure 6.23: Results of CDSU and CDSU_M on the **D2EC2M3** data.

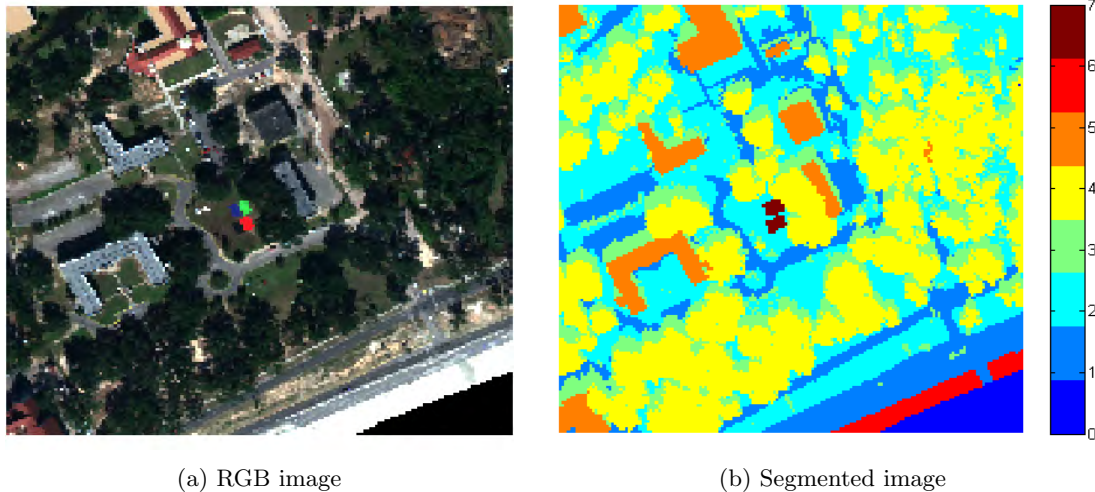


Figure 6.24: The **University of Southern Mississippi** data.

6.4 Cluster Constrained Multi-Model Unmixing

The Cluster Constrained Multi-Model Unmixing algorithm is a semi-supervised variation of CDSU where partial supervision information, in the form of cluster assignment constraints, is used to guide the search process and avoid local minimum solutions. In order to generate such constraints, we needed data collected by two sensors, specifically hyperspectral and LIDAR sensors. We use a down sampled version, of size 163×169 pixels, of the **University of Southern Mississippi** hyperspectral and LIDAR data described in Section 6.1.2.

First, the LIDAR image is segmented using the Digital Elevation Maps (DEM) to partition it into different elevation levels. Then, the vegetation regions are identified using the normalized difference vegetation index NDVI on the co-located hyperspectral image. Finally, the shadowed regions are identified using the altitude of the plane, the position of the sun at the time of the collection, and the DEM [94]. Figure 6.24 illustrates the RGB image of the area and the resulting segmented image. According to [94], the identified segments correspond to: (0) unlabeled, (1) ground/impervious, (2) ground/pervious, (3) ground/shadow, (4) trees, (5) buildings, (6) beach and (7) calibration tarps. The segmented image is used to construct a set of *should-link* constraints by selecting pairs of pixels that belong to the same segment. The image has a total of 27547 pixels, and a subset of pixels were selected randomly to create a set of constraints that includes 6840 *should-link* pairs. The dimensionality of the hyperspectral data was reduced from 68 to 30 using the Ward's linkage strategy with divergence from [95].

To illustrate the advantage of using partial supervision, we compare the performance of

CC-MMU and CDSU. We ran these algorithms using the parameters in table 6.5. Due to the high dimensionality of the feature space, we set the off-diagonal elements of the norm matrices $\mathbf{A}_i$ to zero.

TABLE 6.5

Parameters used for the CC-MMU and CDSU algorithms on the **University of Southern Mississippi** data

Parameters	C	M	m	α	β	γ	$\sigma_i, \forall i$	$\rho_{jk}, \forall j, k$
CC-MMU	3	3	2	4	[10, 10, 10]	0.3	1	1
CDSU	3	3	2	4	[10, 10, 10]	N/A		

Figures 6.25 and 6.26 illustrate the proportion maps resulting from CC-MMU and CDSU respectively. The proportions were multiplied by the corresponding cluster memberships in order to highlight pixels from that cluster. The values are displayed as a heat map where small values are shown in dark blue and large values in dark red. Each proportion map is labeled with the dominant material it represents. In absence of ground truth, endmember labeling was done by comparing the proportion maps to the RGB image in figure 6.24(a). It can be seen that cluster 2 from CC-MMU (2nd row in figure 6.25) and cluster 1 from CDSU (1st row in figure 6.26) are identical. However, the two remaining clusters are different. CDSU combined bare soil and cement in one endmember in cluster 3 (3rd row in figure 6.26), whereas CC-MMU succeeded in identifying different endmembers for those materials (3rd row in figure 6.25). We also notice that CC-MMU combined man-made materials (asphalt, bitumen, cement) in one cluster (3rd row in figure 6.25), as opposed to CDSU which divided them into two different clusters (2nd and 3rd rows in figure 6.26).

We also provide a comparison of running CDSU, CDSU_M and CC-MMU using the Euclidean and the Mahalanobis distances by reporting the number of satisfied constraints in table 6.6. We

TABLE 6.6

Number of satisfied constraints using CC-MMU and CDSU with Euclidean and Mahalanobis distances on the **University of Southern Mississippi** data

Semi-supervised		Unsupervised	
CC-MMU (Euclidean)	CC-MMU (Mahalanobis)	CDSU	CDSU _M
4842 / 6840	6591 / 6840	4755 / 6840	6263 / 6840

notice that running CC-MMU using the Mahalanobis distance yields more satisfied constraints. This is expected since the Mahalanobis distance allows for more degrees of freedom by seeking ellipsoidal clusters instead of spherical ones as with the Euclidean distance. The additional degrees

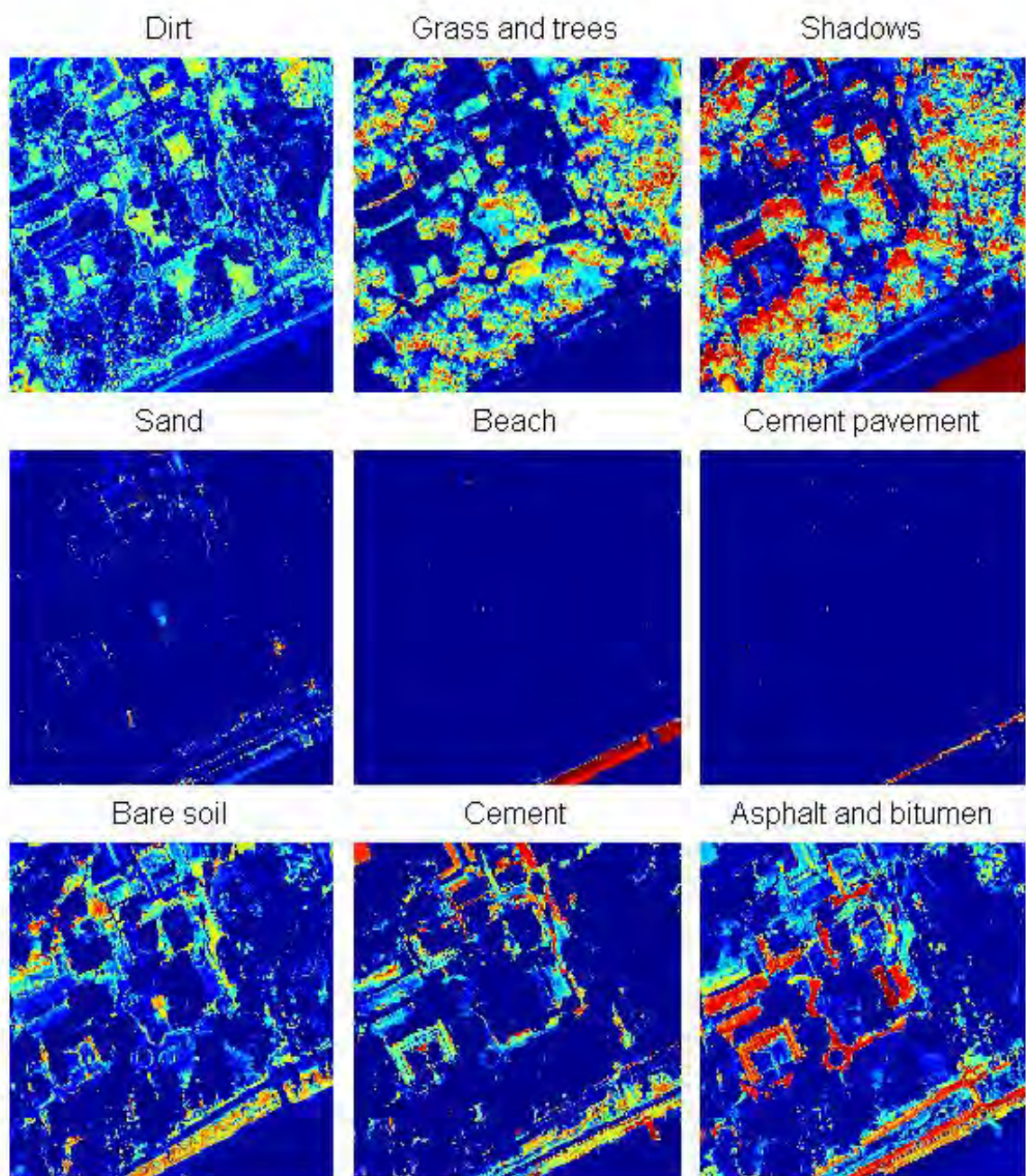


Figure 6.25: Proportion maps from CC-MMU on the **University of Southern Mississippi** data. Rows correspond to clusters and columns correspond to endmembers.

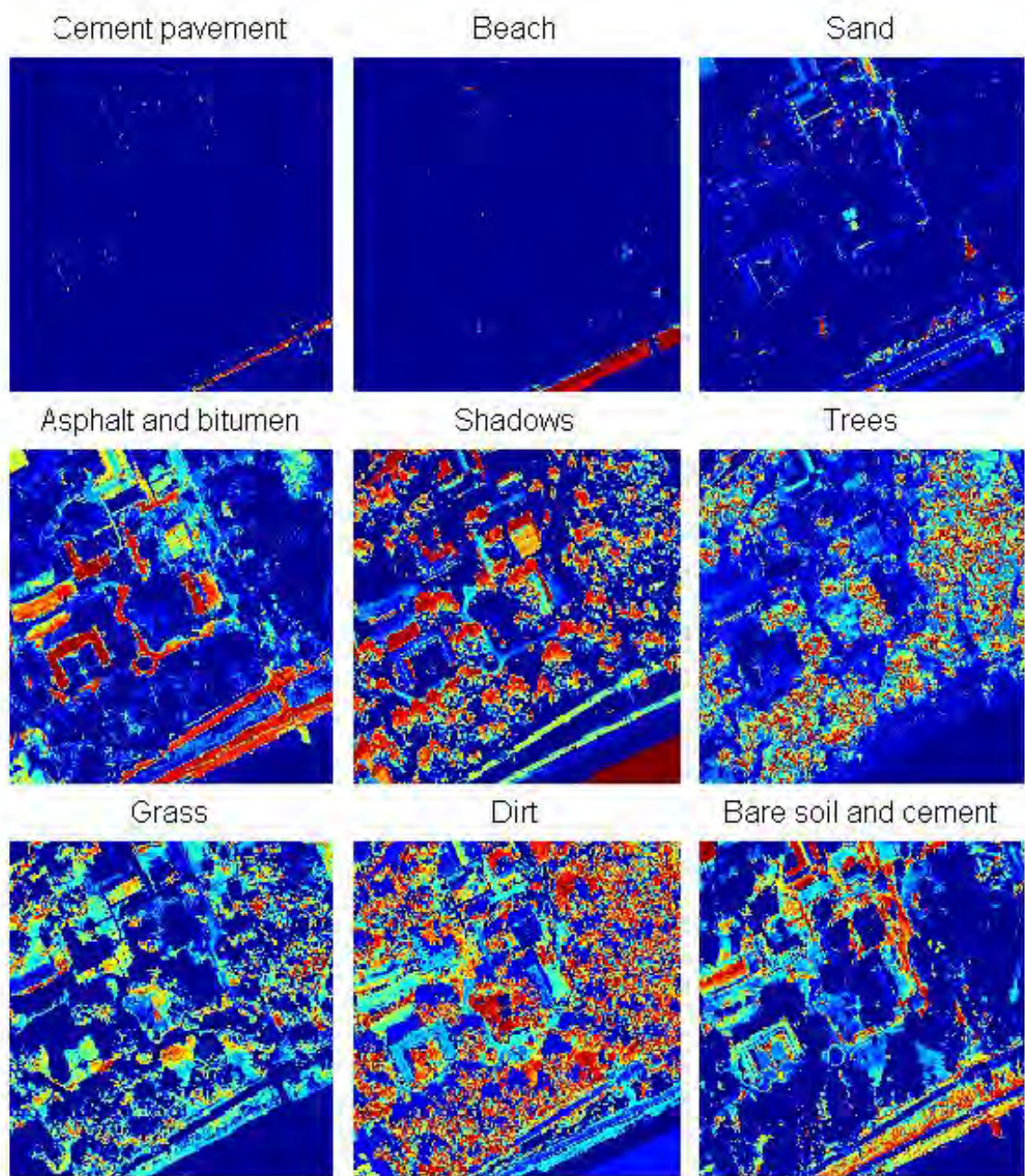


Figure 6.26: Proportion maps from CDSU on the **University of Southern Mississippi** data. Rows correspond to clusters and columns correspond to endmembers.

of freedom enable the supervision constraints to have more influence on the final partition. It is worth mentioning here that the proportion maps for CDSU_M and CC-MMU with the Mahalanobis distance are fairly similar. Moreover, the proportion maps for CDSU and CC-MMU using the Euclidean distance are also fairly similar. This indicates that, for this data, the Mahalanobis distance is more influential than the semi-supervision information for the clusters formation. Nevertheless, CC-MMU can still be helpful for other data sets or with other supervision information as it offers the option to incorporate domain and expert knowledge into the unmixing process.

6.5 Proportion Constrained Multi-Model Unmixing

The Proportion Constrained Multi-Model Unmixing (PC-MMU) algorithm is another semi-supervised version of CDSU where partial supervision information, in the form of constraints based on proportions, is used to guide the search process and avoid local minimum solutions.

We use the same data of the previous section to test PC-MMU. Two cases of generating the constraints are used. In the first case, the set of *should-link* constraints of the previous section is used as the set of points constrained to have similar proportions in the extracted endmembers. In the second case, we use the consensus analysis presented in Section 4.2 to form the set of constraints. We randomly select 5000 pairs of voters, that have high co-association values, from the set $\mathcal{H}_a$. The voters are points having a high proportion value in any extracted endmember of any of the unmixing algorithms. The co-association value of a pair of voters is obtained over the entire unmixing ensemble, forming a new similarity measure based on the proportions. A pair of voters that have a high co-association value is likely to have similar proportions in the extracted endmembers.

For both cases, we use the same parameters as for CC-MMU in Section 6.4 (table 6.5). The resulting proportion maps are fairly similar to the ones resulting from CC-MMU (figure 6.25). Hence, the same conclusions apply here.

6.6 Robust Context Dependent Spectral Unmixing

The Robust Context Dependent Spectral Unmixing (RCDSU) is a variation of CDSU that can handle noisy data. It is based on the use of possibilistic membership functions along with the fuzzy membership functions of CDSU. The fuzzy memberships are used to partition the spectra space into multiple convex sets that span the entire space and avoid coincident clusters, while the possibilistic memberships are used to reduce the effect of noise and obtain robust estimates of the endmembers and proportions within each cluster. In all results reported in this chapter, we update

the parameters η_i every iteration using [85]

$$\eta_i = \text{mean}\{\text{cost}_{ij}, t_{ij} \geq t_i^Q\}, \quad (6.7)$$

where cost_{ij} is given by (3.31), and t_i^Q is the Q^{th} percentile of the typicalities in cluster i . Q can be thought of as the percentage of points not belonging to cluster i . In this experiment, we let $Q = 100 - \frac{100}{C}$. Furthermore, the convergence of the algorithm is checked by comparing the values of the objective function from successive iterations. If the difference is below some threshold, the algorithm is stopped.

We design the following experiment so that we evaluate the ability of RCDSU to handle noisy data. We use the **Usgs2C2M3** simulated hyperspectral data sets described in Section 6.1.1.2.

We run P-COMMEND, CDSU_M and RCDSU 25 times using the parameters in table 6.7. When the data is expected to be noisy, we set a higher weight to the possibilistic memberships ($b = 0.9$) compared to the fuzzy memberships ($a = 0.1$). Otherwise, we set $a = 0.9$ and $b = 0.1$.

TABLE 6.7

Parameters used for P-COMMEND, CDSU_M and RCDSU on the **Usgs2C2M3** data

	C	M	m	n	a	b	$\beta_i, \forall i$	$\sigma_i, \forall i$	α
P-COMMEND	2	3	1.5	N/A					1
CDSU_M	2	3	1.5	N/A			1	1	100
RCDSU	2	3	1.5	1.5	0.1	0.9	1	1	100

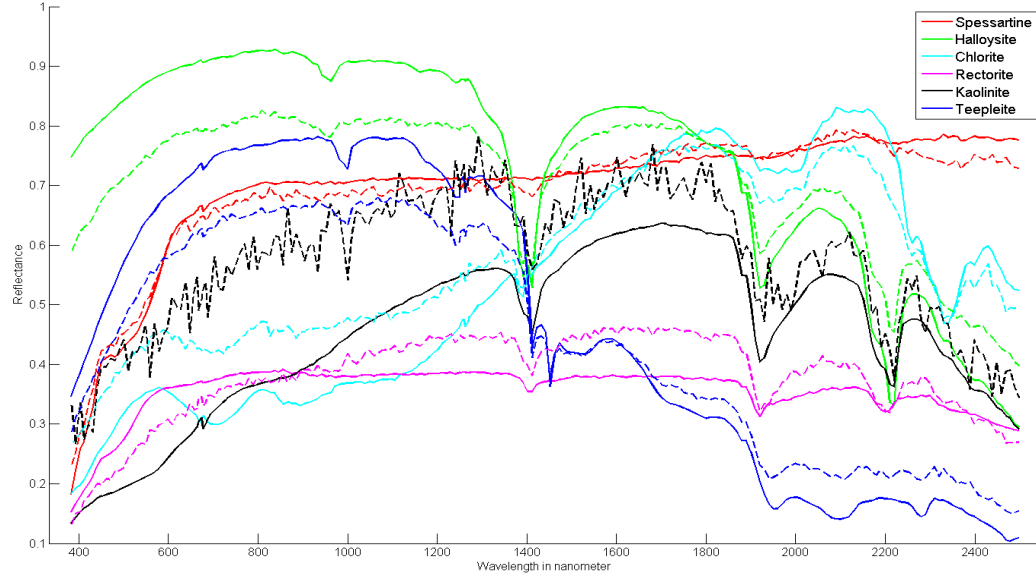
The estimated endmembers from all algorithms are compared to the true endmembers using the spectral mean angle error (SMAE) defined in equation (6.4). We report the average and standard deviation of the SMAE of the resulting endmember estimates from all algorithms in all runs in table 6.8. A two-sample t-test at the 5% significance level shows that, for noisy data, RCDSU provides significantly better endmember estimates than P-COMMEND and CDSU_M (p -value $< 1\text{e-}20$). In

TABLE 6.8

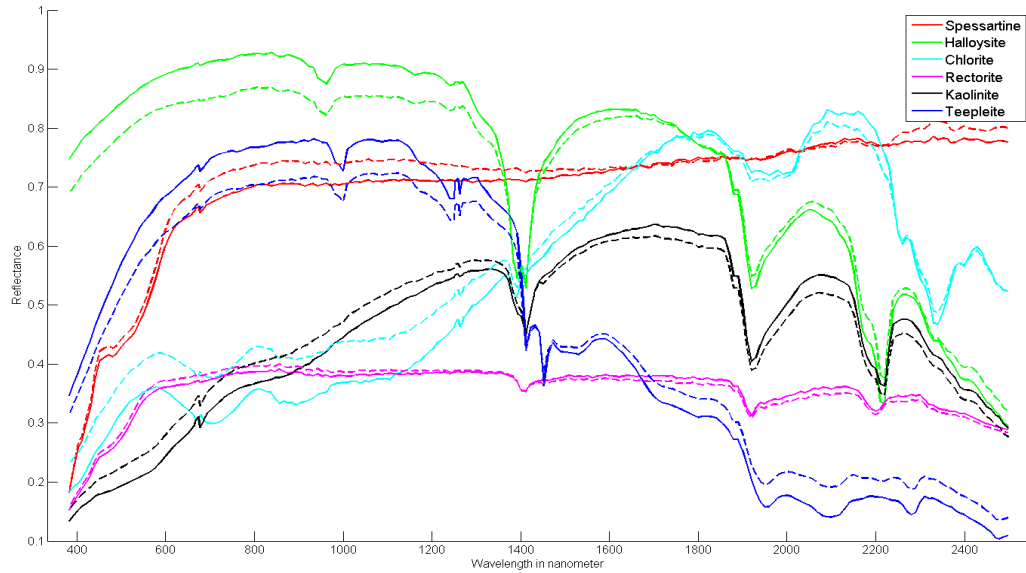
Average ($\pm$ standard deviation) of the SMAE over 25 runs for P-COMMEND, CDSU_M and RCDSU on the **Usgs2C2M3** data

% of noise points	0	5	10
P-COMMEND	0.0630 (± 0)	0.1609 (± 0.0046)	0.1644 (± 0.0135)
CDSU_M	0.0630 (± 0)	0.1609 (± 0.0046)	0.1662 (± 0.0119)
RCDSU	0.0639 (± 0)	0.0681 (± 0.0074)	0.0757 (± 0.0105)

figure 6.27, we illustrate the estimated (dashed line) versus the true (solid line) endmembers for CDSU_M and RCDSU for the **Usgs2C2M3** data with 5% noise points. As it can be seen, the



(a) CDSU_M



(b) RCDSU

Figure 6.27: True (solid line) and estimated (dashed line) endmembers of CDSU_M and RCDSU for the **Usgs2C2M3** data with 5% noise points.

RCDSU was not influenced by the presence of the noise points unlike CDSU_M which got affected and resulted in erroneous estimates.

To verify the ability of RCDSU to identify noise points, after convergence, we identified points with small (< 0.1) possibilistic memberships in all clusters. All of these points correspond to the points with added noise.

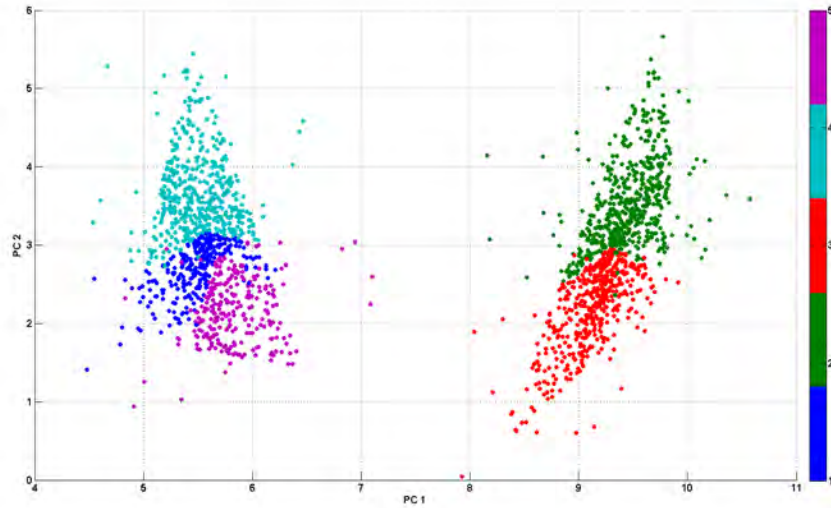


Figure 6.28: Initial cluster assignments on the **Usgs2C2M3** data with 5% noise points using RCDSU with $C_{max} = 5$.

6.7 Unsupervised Robust Context Dependent Spectral Unmixing

So far, we assumed that the number of contexts is known. If this is not the case, we use the proposed Unsupervised Robust Context Dependent Spectral Unmixing (U-RCDSU). U-RCDSU takes advantage of the fact that the possibilistic membership functions are not constrained to sum to one, and hence, small clusters covering the same dense regions would expand and become similar. These are then merged and the number of clusters, initially overspecified, is reduced until it reaches the optimal number. In the following, we evaluate the ability of U-RCDSU to determine the correct number of contexts using simulated and real data.

6.7.1 Evaluation using simulated data

We use the same data (with 5% noise points) and parameters of Section 6.6 and we run U-RCDSU by overspecifying the number of clusters to $C_{max} = 5$. We let $\epsilon = 0.1$. The initial cluster assignments are shown in figure 6.28, where we plot the 2 principal components (PC) of the data, labeling each point with its cluster assignment. Figure 6.29 shows which clusters got merged after each iteration of the U-RCDSU algorithm. Figure 6.30 illustrates the final cluster assignments of the data. As it can be seen, the algorithm converged to $C = 2$. The resulting endmembers are shown in figure 6.31. As it can be seen, the identified endmembers are unaffected by the noise points, and are similar to the results of RCDSU in figure 6.27(b).

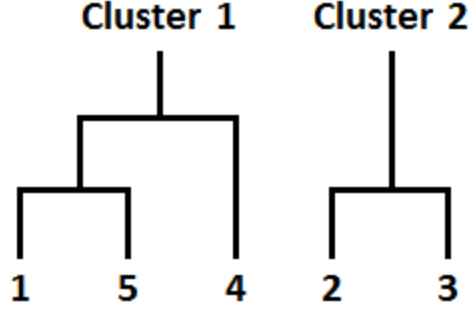


Figure 6.29: Cluster merging on the **Usgs2C2M3** data with 5% noise points using U-RCDSU.

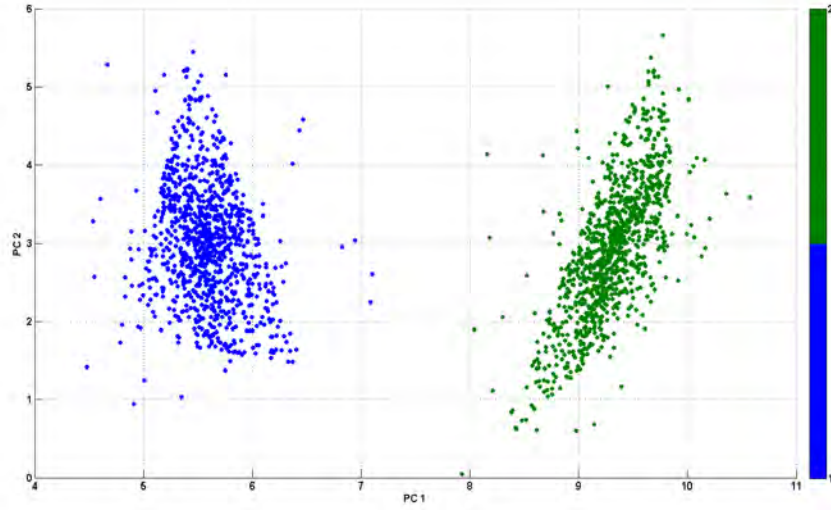


Figure 6.30: Final cluster assignments on the **Usgs2C2M3** data with 5% noise points using U-RCDSU.

6.7.2 Evaluation using real data

We use two real hyperspectral data sets: the **Pavia University** and the **University of Southern Mississippi** data.

6.7.2.1 Pavia University data

We use a down sampled version, of size 204×114 pixels, of the **Pavia University** data described in Section 6.1.2. We run the unsupervised RCDSU algorithm using $C_{max} = 10$, $M = 3$, $m = n = 1.5$, $a = b = 0.5$, $\alpha = 100$, $\beta_i = 4$, $\forall i$, $\sigma_i = 1$, $\forall i$ and $\epsilon = 0.1$. The algorithm converged to $C = 2$. Figure 6.32 shows the evolution of the cluster assignments of the data points as U-RCDSU runs. We plot 2 principal components (PC) of the data, labeling each point with its cluster

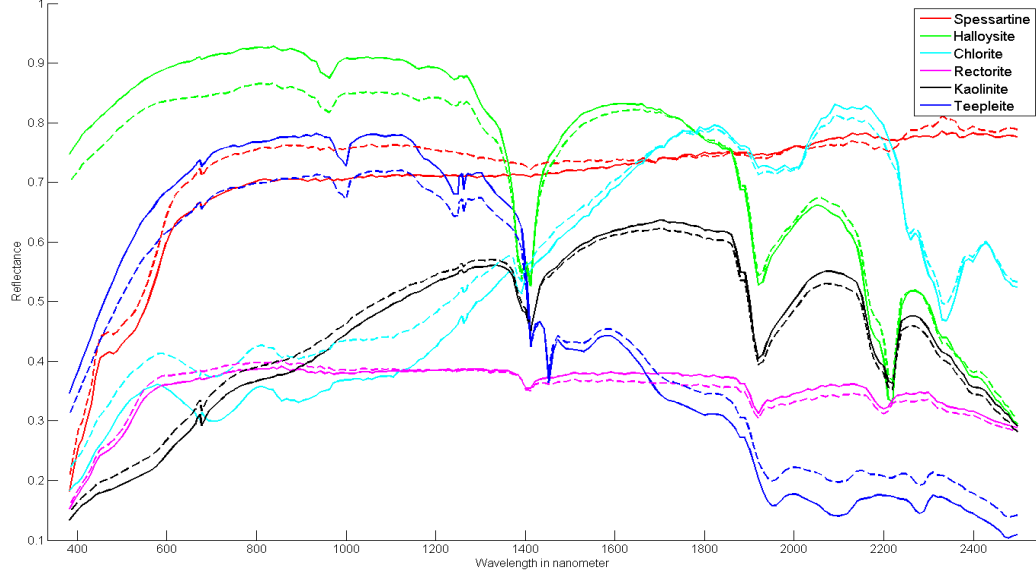
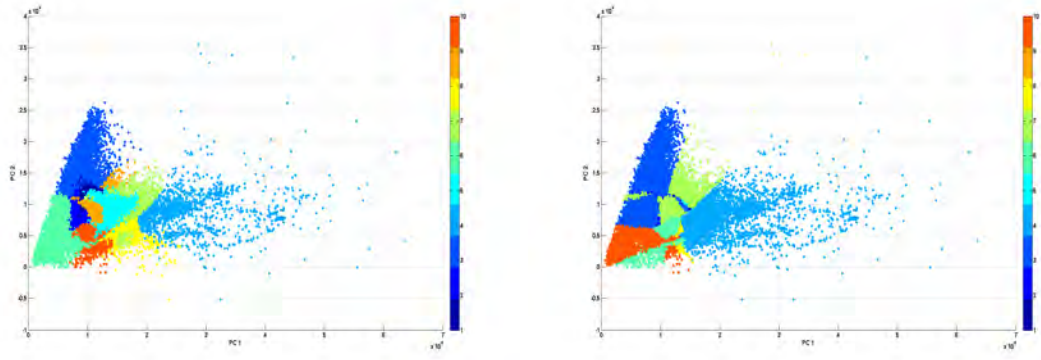


Figure 6.31: True (solid line) and estimated (dashed line) endmembers for the **Usgs2C2M3** data with 5% noise points using U-RCDSU.

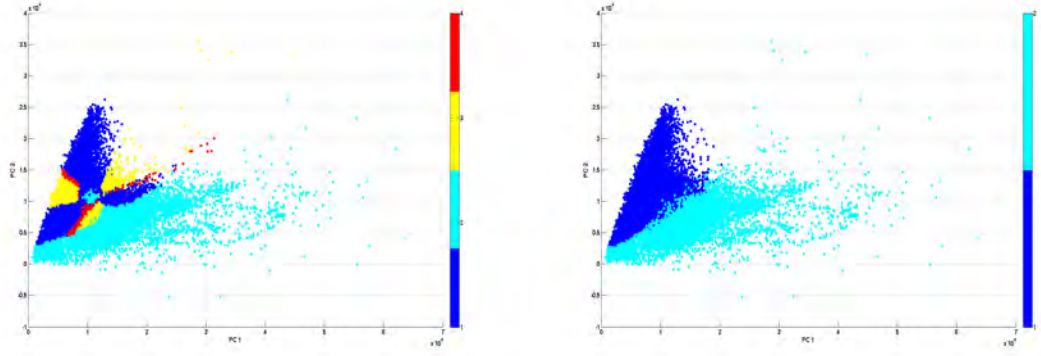
assignment. Figure 6.32(a) shows the cluster assignments after running RCDSU with $a = b = 0.5$ and $C_{max} = 10$. Figure 6.32(b) shows the cluster assignments after running RCDSU with $a = 0$, $b = 1$ and $C_{max} = 10$. Figure 6.32(c) shows the cluster assignments after one more iteration of U-RCDSU. As it can be seen, we went from 10 to 4 clusters. Finally, figure 6.32(d) shows the final cluster assignments after convergence (1 more iteration). We went from 4 to 2 clusters.

Since the ground truth is not available for the entire scene, we evaluate the performance of the algorithm qualitatively by displaying the resulting proportion maps and interpreting them based on the RGB image of the scene shown in figure 6.3(a). The proportion maps associated with the three endmembers for each of the two clusters are shown in figure 6.33. The proportions were multiplied by the corresponding weighted cluster memberships ($au_{ij} + bt_{ij}$) in order to highlight pixels from that cluster. Dark blue represents small values and dark red represents large values. Each proportion map is labeled with the dominant material it represents.

It can be seen that U-RCDSU resulted in two intuitive clusters in the sense that one cluster corresponds to natural materials, and the other corresponds to man-made materials. The three proportion maps of context 1 (figure 6.33(a)) correspond to vegetation, shadows and bare soil, respectively. These materials represent natural regions in the scene. The three proportion maps of context 2 (figure 6.33(b)) correspond to brick roofs, metal roofs, and asphalt and bitumen, respectively. They represent urban regions in the scene. One may conclude that U-RCDSU identified a reasonable number of contexts with coherent content and appropriate endmembers within each



(a) Assignments with $a = b = 0.5$ and $C_{max} = 10$. (b) Assignments with $a = 0, b = 1$ and $C_{max} = 10$.



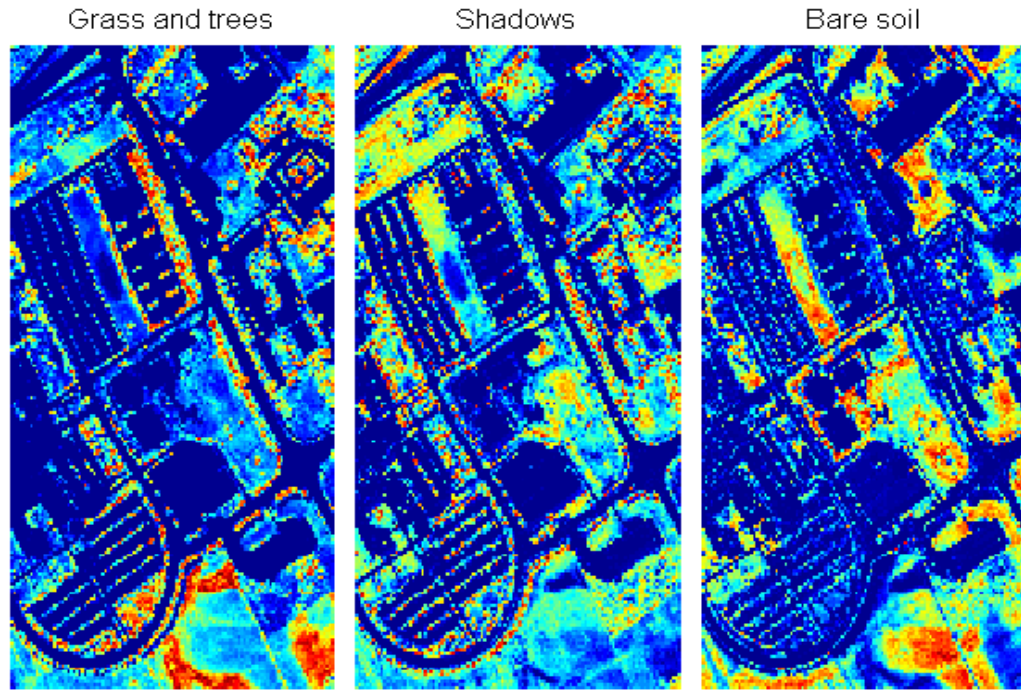
(c) Assignments after 1 iteration. (d) Assignments after convergence (iteration 2).

Figure 6.32: Evolution of the cluster assignments using U-RCDSU with $C_{max} = 10$ on the **Pavia University** data.

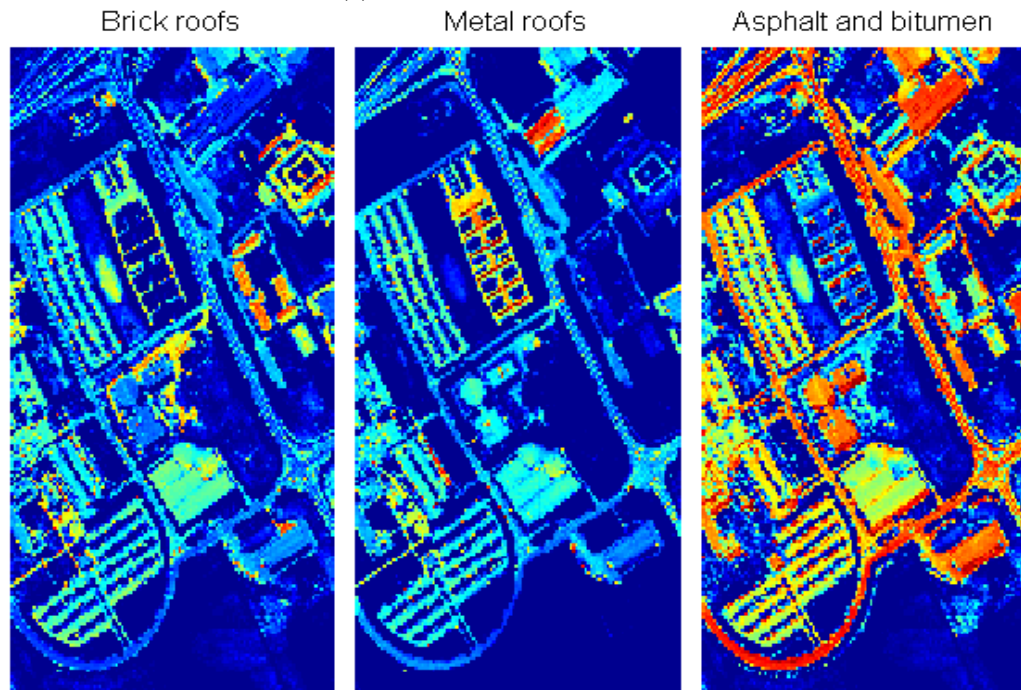
context.

6.7.2.2 University of Southern Mississippi data

We run the unsupervised RCDSU algorithm on the other real data, the **University of Southern Mississippi** hyperspectral data described in Section 6.1.2. We use a down sampled version of size 163×169 pixels. We let $C_{max} = 5$, $M = 3$, $m = n = 1.5$, $a = b = 0.5$, $\alpha = 100$, $\beta_i = 1, \forall i$, $\sigma_i = 1, \forall i$ and $\epsilon = 0.2$. The algorithm converged to $C = 2$. Since the ground truth is not available for this scene, we evaluate the performance of the algorithm qualitatively by displaying the resulting proportion maps and interpreting them based on the RGB image of the scene shown in figure 6.24(a). Each proportion map was multiplied by the corresponding weighted memberships ($au_{ij} + bt_{ij}$). Figure 6.34 illustrates these maps where the title of each one represents the corresponding dominant materials. For this data also, the U-RCDSU identified a reasonable number of contexts with coherent content. The endmembers in cluster 1 correspond to grass, asphalt and



(a) Proportion maps in context 1



(b) Proportion maps in context 2

Figure 6.33: Proportion maps from U-RCDSU on the **Pavia University** data. Rows correspond to clusters and columns correspond to endmembers.

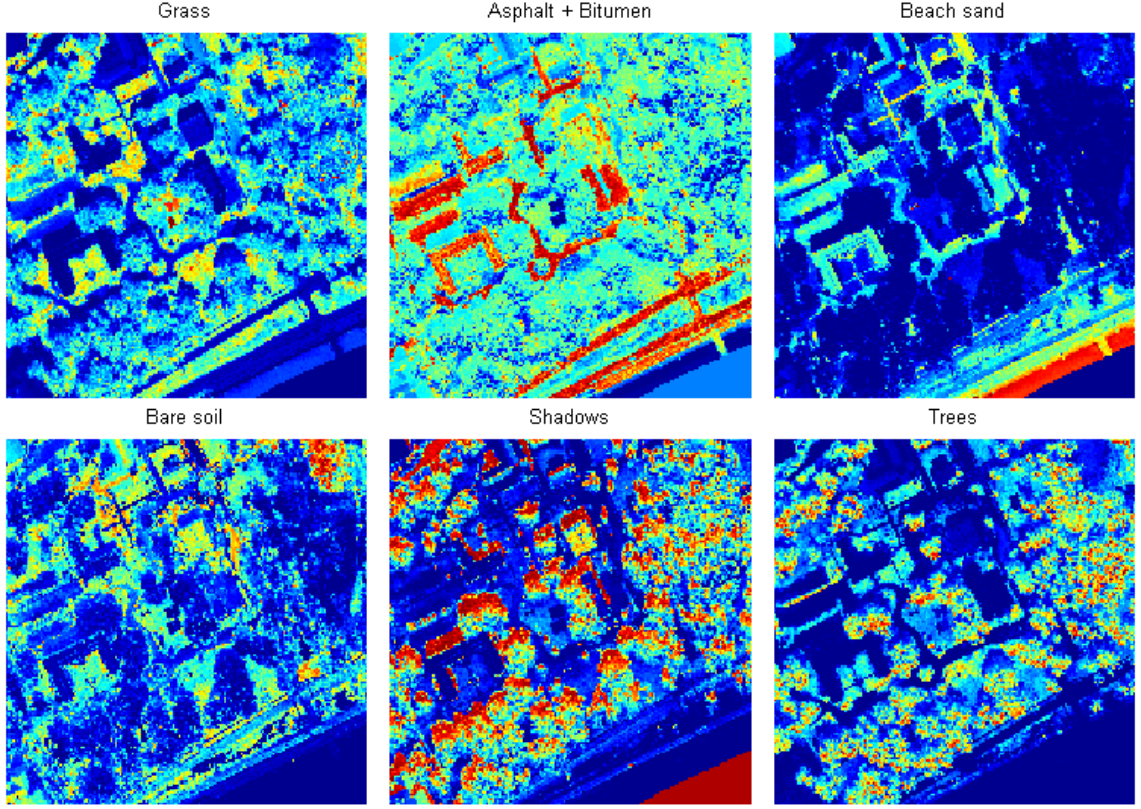


Figure 6.34: Proportion maps from U-RCDSU on the **University of Southern Mississippi** data. Rows correspond to clusters and columns correspond to endmembers.

bitumen, and beach sand. In cluster 2, the endmembers represent bare soil, shadows and trees.

6.8 Robust Unmixing Using Consensus Analysis

Robust Unmixing using Consensus Analysis is based on the idea that “optimal” endmembers in the data will have a consensus among multiple unmixing results. The method combines results from multiple unmixing algorithms with different numbers of endmembers to find the optimal endmembers using consensus analysis. The approach overcomes the weaknesses of individual algorithms and provides a robust alternative to estimating the endmembers in the data. First, we illustrate the results of the proposed approach on the **Usgs3C1M4** simulated data described in Section 4.1. Then, we present the results on a real data. The pixels’ proportion values, for the unmixing algorithms that return endmembers only, are computed through a least squares constrained minimization using Lagrange Multipliers optimization.

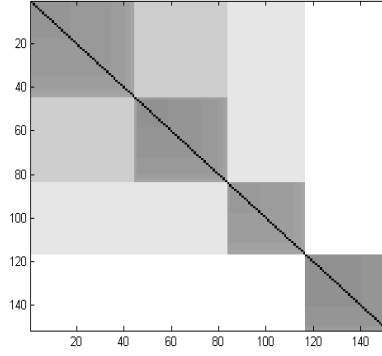


Figure 6.35: iVAT image of $\mathcal{H}_a$ of the **Usgs3C1M4** data.

6.8.1 Evaluation using simulated data

For this data set, we set $Q = 20$, $\alpha_p = 0.85$ and $\alpha_c = 95^{\text{th}}$ percentile of $\{\mathcal{C}(j, l) \mid j \neq l\}$. This resulted in 151 voters ($\sim 8\%$ of the data) in the set $\mathcal{H}_a$. First, to illustrate the fact that $\mathcal{H}_a$ can be easily clustered, we use the Improved Visual Assessment of Cluster Tendency (iVAT) [96] algorithm to visualize it. iVAT takes as input a dissimilarity matrix $\mathbf{D}$ and reorders its elements so that clusters can be visualized. The iVAT image is shown in figure 6.35. We can clearly see 4 dark blocks on the diagonal, indicating the presence of 4 clusters, which is conform to the true number of endmembers used to generate the data. In figure 6.36, we scatter plot the 3 principal components (PC) of the data (gray dots) and the voters of set $\mathcal{H}_a$ in colored dots. Each color represents one cluster of voters. Here, an average link hierarchical clustering [90] is used to identify the 4 clusters. The medoids of the clusters are selected to represent the extracted endmembers. These are illustrated in figure 6.36 using large red 'X's. The large green 'X's represent the true endmembers. As it can be seen, the extracted endmembers are close to the true ones.

6.8.2 Evaluation using real data

In this experiment, we use a down sampled version, of size 204×114 , of the **Pavia University** data described in Section 6.1.2. We run ICE [61], VCA [66], PPI [59], and N-FINDR [60] using a number of endmembers varying from 5 to 10, repeating each run 4 times (giving a total of 96 runs). We also run U-RCDSU using $C_{max} = 5$ with $M = 2$ and $M = 3$ (the other parameters were the same as in Section 6.7.2.1), leading to 4 contexts with 2 endmembers each, and 2 contexts with 3 endmembers each, respectively. In total, we have 102 endmember sets.

We set $Q = 50$, $\alpha_p = 0.85$ and $\alpha_c = 90^{\text{th}}$ percentile of $\{\mathcal{C}(j, l) \mid j \neq l\}$. This resulted in

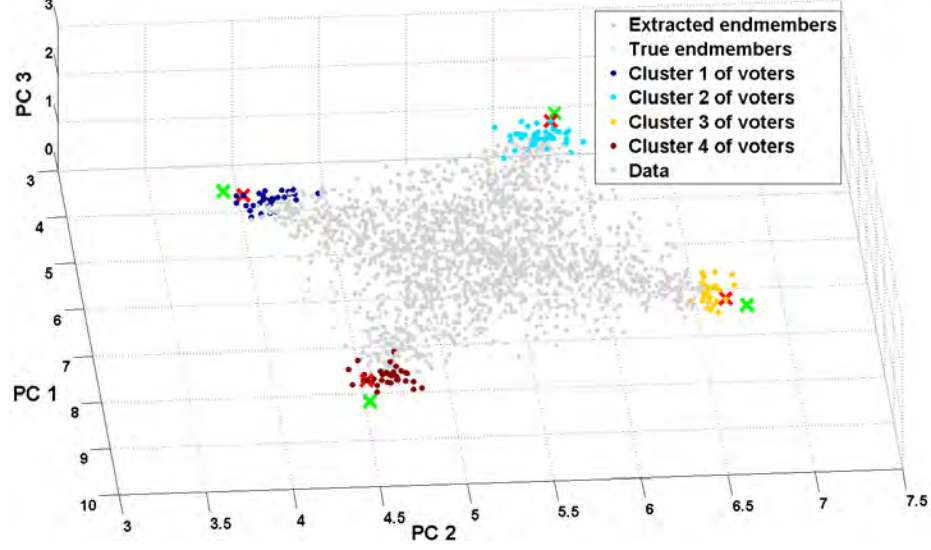


Figure 6.36: The PC of the **Usgs3C1M4** data, the voters, the extracted endmembers and the true endmembers.

1804 voters ($\sim 8\%$ of the data) in the set $\mathcal{H}_a$. The iVAT image for these, illustrated in figure 6.37, shows 5 blocks along the diagonal. An average link hierarchical clustering algorithm identified the 5 clusters. We pick the medoid of each cluster as the representative of the endmember. We compute the proportion values of these endmembers in all pixels and display them as proportion maps in figure 6.38. The values are shown as a heat map (dark blue represents low values and dark red represents high values), and each proportion map is labeled with the dominant material it represents. It can be seen that these proportion maps are similar to the ones in figure 6.33 resulting from U-RCDSU, with the exception that U-RCDSU resulted in separate endmembers for shadows and roads, as opposed to only one by the consensus unmixing.

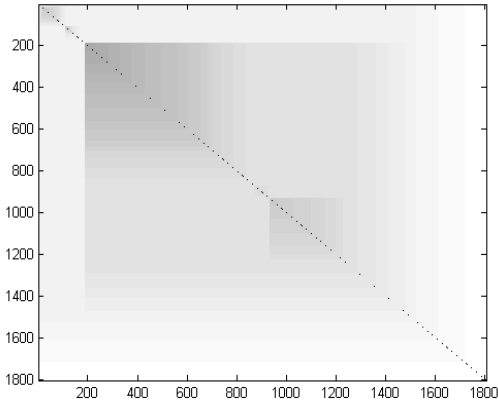


Figure 6.37: iVAT image of $\mathcal{H}_a$ of the **Pavia University** data.

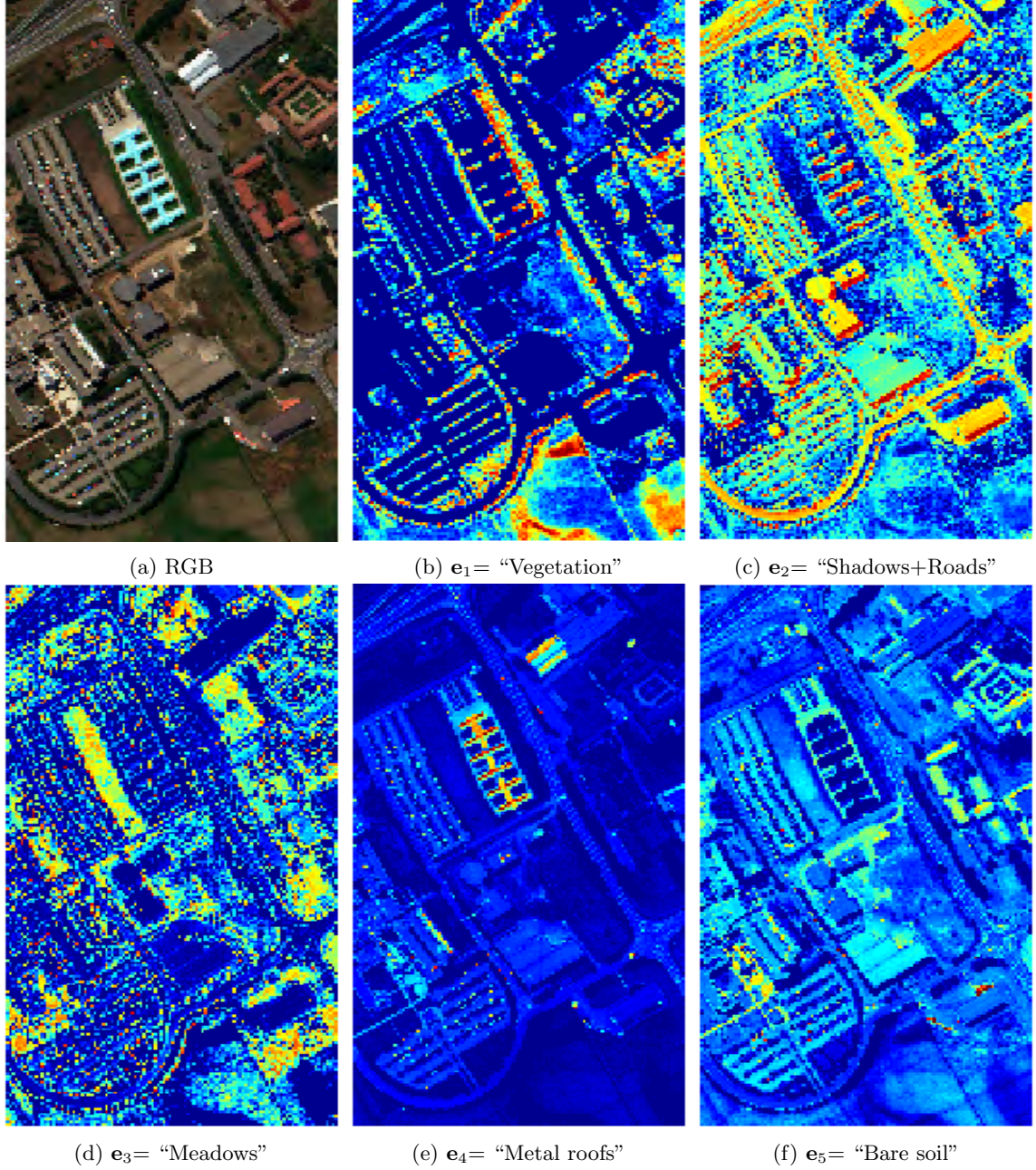


Figure 6.38: RGB and proportion maps of the **Pavia University** data using consensus analysis.

Given the ground truth data of some areas in the scene, in table 6.9, we report the classes of the labeled points having proportion values greater than a threshold t . For each endmember, we show the threshold t and the number n of these points. We compare the proportions generated by the consensus approach to those generated by N-FINDR, VCA, ICE and PPI, using 5 endmembers.

From table 6.9, we notice that some of the selected individual unmixing methods resulted

TABLE 6.9

Comparison of high proportion pixels (in %) among ground truth classes using consensus unmixing, N-FINDR, VCA, ICE and PPI on the **Pavia University** data

Class	Asphalt	Meadows	Trees	Metal sheets	Bare soil	Shadows	Self-blocking bricks	Gravel	Bitumen
Endmembers									
Consensus unmixing									
e_1 : "Vegetation" (Fig. 6.38b), $t=0.5$, $n=638$	0 %	52.66 %	47.18 %	0 %	0.16 %	0 %	0 %	0 %	0 %
e_2 : "Shadows + Roads" (Fig. 6.38c), $t=0.7$, $n=141$	24.82 %	0 %	0 %	0 %	0 %	75.18 %	0 %	0 %	0 %
e_3 : "Meadows" (Fig. 6.38d), $t=0.6$, $n=182$	1.10 %	85.16 %	0 %	0 %	13.19 %	0 %	0 %	0 %	0.55 %
e_4 : "Metal roofs" (Fig. 6.38e), $t=0.5$, $n=135$	0 %	0 %	0 %	100 %	0 %	0 %	0 %	0 %	0 %
e_5 : "Bare soil" (Fig. 6.38f), $t=0.4$, $n=113$	6.19 %	0.88 %	0 %	0 %	84.07 %	0 %	3.54 %	5.31 %	0 %
N-FINDR									
e_1 : "Bare soil", $t=0.2$, $n=266$	6.77 %	18.05 %	0 %	0 %	45.86 %	0 %	15.04 %	14.29 %	0 %
e_2 : "Metal roofs", $t=0.2$, $n=78$	0 %	0 %	0 %	100 %	0 %	0 %	0 %	0 %	0 %
e_3 : "Metal roofs", $t=0.4$, $n=66$	0 %	0 %	0 %	100 %	0 %	0 %	0 %	0 %	0 %
e_4 : "Vegetation", $t=0.6$, $n=454$	0 %	50.88 %	49.12 %	0 %	0 %	0 %	0 %	0 %	0 %
e_5 : "Shadows + Roads", $t=0.8$, $n=166$	35.54 %	0 %	0 %	0 %	0 %	64.46 %	0 %	0 %	0 %
VCA									
e_1 : "Metal roofs", $t=0.5$, $n=144$	0 %	0 %	0 %	100 %	0 %	0 %	0 %	0 %	0 %
e_2 : "Vegetation", $t=0.9$, $n=473$	0 %	39.53 %	60.47 %	0 %	0 %	0 %	0 %	0 %	0 %
e_3 : "Shadows", $t=0.5$, $n=104$	0 %	0 %	0 %	0 %	0 %	100 %	0 %	0 %	0 %
e_4 : "Bricks + Bare soil", $t=0.6$, $n=727$	5.23 %	12.65 %	0.14 %	0.28 %	21.46 %	0 %	42.92 %	17.33 %	0 %
e_5 : "Roads", $t=0.3$, $n=226$	94.69 %	0 %	0 %	0 %	0 %	3.98 %	0.44 %	0 %	0.88 %
ICE									
e_1 : "Shadows + Roads", $t=0.4$, $n=396$	53.03 %	4.04 %	0 %	0 %	6.31 %	26.77 %	2.27 %	2.27 %	5.3 %
e_2 : "Vegetation", $t=0.6$, $n=458$	0 %	51.31 %	48.69 %	0 %	0 %	0 %	0 %	0 %	0 %
e_3 : "Bare soil", $t=0.25$, $n=100$	7 %	8 %	0 %	0 %	75 %	0 %	4 %	6 %	0 %
e_4 : "Shadows + Roads", $t=0.4$, $n=165$	31.52 %	0 %	0 %	4.24 %	0 %	64.24 %	0 %	0 %	0 %
e_5 : "Metal roofs", $t=0.5$, $n=104$	0 %	0 %	0 %	100 %	0 %	0 %	0 %	0 %	0 %
PPI									
e_1 : "Not clear", $t=0.1$, $n=0$	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
e_2 : "Meadows", $t=0.5$, $n=3084$	9.82 %	51.95 %	0.23 %	0.03 %	17.67 %	0.71 %	11.87 %	7.39 %	0.32 %
e_3 : "Not clear", $t=0.1$, $n=0$	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
e_4 : "Vegetation", $t=0.9$, $n=144$	0 %	22.22 %	77.08 %	0 %	0.69 %	0 %	0 %	0 %	0 %
e_5 : "Metal roofs", $t=0.6$, $n=26$	7.69 %	0 %	0 %	92.31 %	0 %	0 %	0 %	0 %	0 %

in duplicated endmembers (N-FINDR and ICE), or missed some endmembers (PPI). VCA, on the other hand, resulted in appropriate endmembers. The consensus unmixing overcame the "erroneous" results of some runs by only considering the endmembers over which there was a consensus among the different runs. As a result, the consensus unmixing identified more robust and consistent endmembers.

6.9 Context Dependent Hyperspectral Subpixel Target Detection

We designed these experiments to evaluate the performance of the proposed context dependent (CD) target detectors and compare them to the traditional detectors that use a single endmember set to describe the background. For the traditional detectors, we use the MVSA, NFINDR, PPI and the eigenvectors of the data correlation matrix method, to find a set of 3 or 6 endmembers (referred to as MVSA3 and MVSA6, NFINDR3 and NFINDR6, PPI3 and PPI6, and EigVect3 and EigVect6 respectively).

6.9.1 Evaluation using implanted targets

We use the **Indian Pines** data described in Section 6.1.2. We implant one hundred spectral signatures of a red tarp target with abundance values ranging from 0.1 to 1. Figure 6.39 shows band



Figure 6.39: Band 115 of the **Indian Pines** data with 100 implanted targets.

115 ($\sim 1.5\mu m$) of the image where we can see the grid of 100 targets. The top row consists of the targets with abundance values of 0.1. These are difficult to see since they have low proportion (90% of the pixel is background). The next row consists of the targets with abundance values of 0.2. This pattern continues till the last row which corresponds to pure targets of abundance values of 1. We run U-RCDSU using $C_{max} = 5$, $M = 3$, $m = n = 1.5$, $a = 0.1$, $b = 0.9$, $\alpha = 100$, $\beta_i = 1$, $\forall i$, $\sigma_i = 1$, $\forall i$ and $\epsilon = 0.1$. The algorithm converges to $C = 2$ clusters.

We start by analyzing the resulting possibilistic memberships of the target pixels in both clusters. Figure 6.40 shows a scatter plot of the maximum possibilistic memberships of the targets in both clusters as a function of their proportions in the pixels. All proportions are very small (< 0.06), which means that the targets did not contribute to the estimated local background subspaces. In other words, there was no leakage of targets into the endmember sets. We can also notice that the larger the proportion of the target, the smaller its possibilistic membership, which is expected. These memberships could be used on their own to detect targets. However, this would result in a large number of false alarms, since the memberships cannot discriminate between targets and non-target outliers.

In figure 6.41, we scatter plot the detection statistic of CD-OSP as a function of the target proportion in the pixel. We notice that the statistic increases as the target proportion increases, which is expected.

The performance of the detectors is evaluated using the receiver operating characteristic

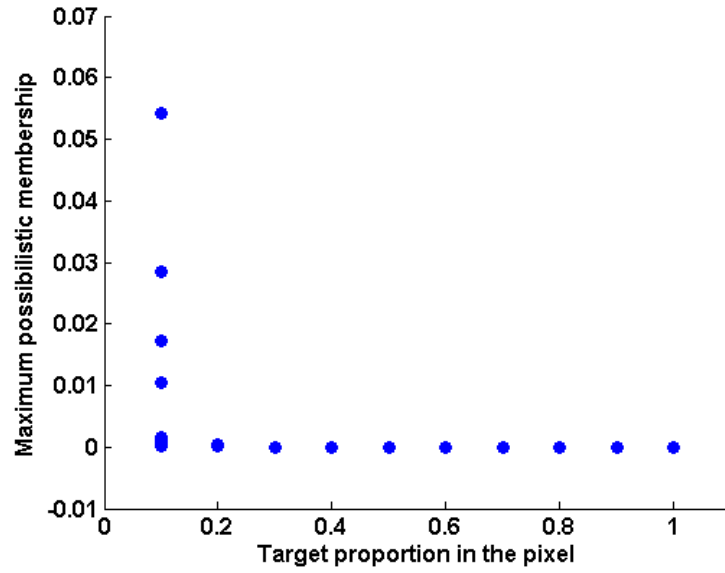


Figure 6.40: Scatter plot of the possibilistic memberships of the targets as a function of their proportions in the pixels (**Indian Pines** data).

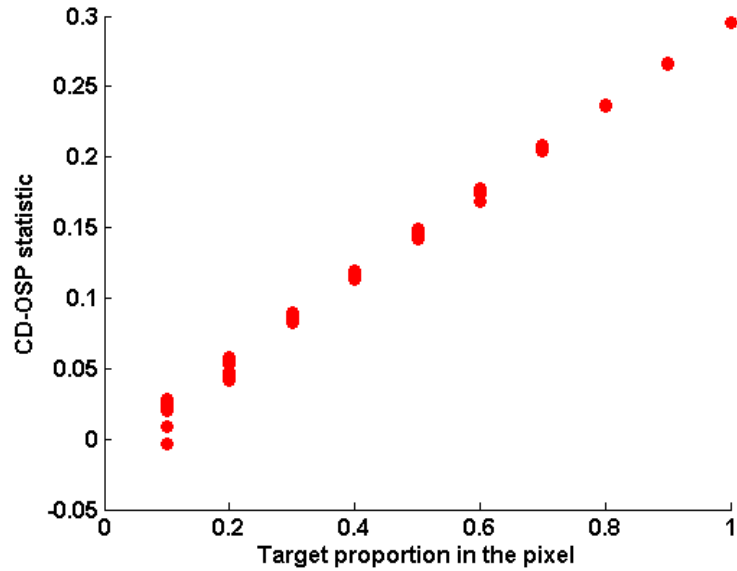


Figure 6.41: Scatter plot of the CD-OSP statistic as a function of the target proportion in the pixel (**Indian Pines** data).

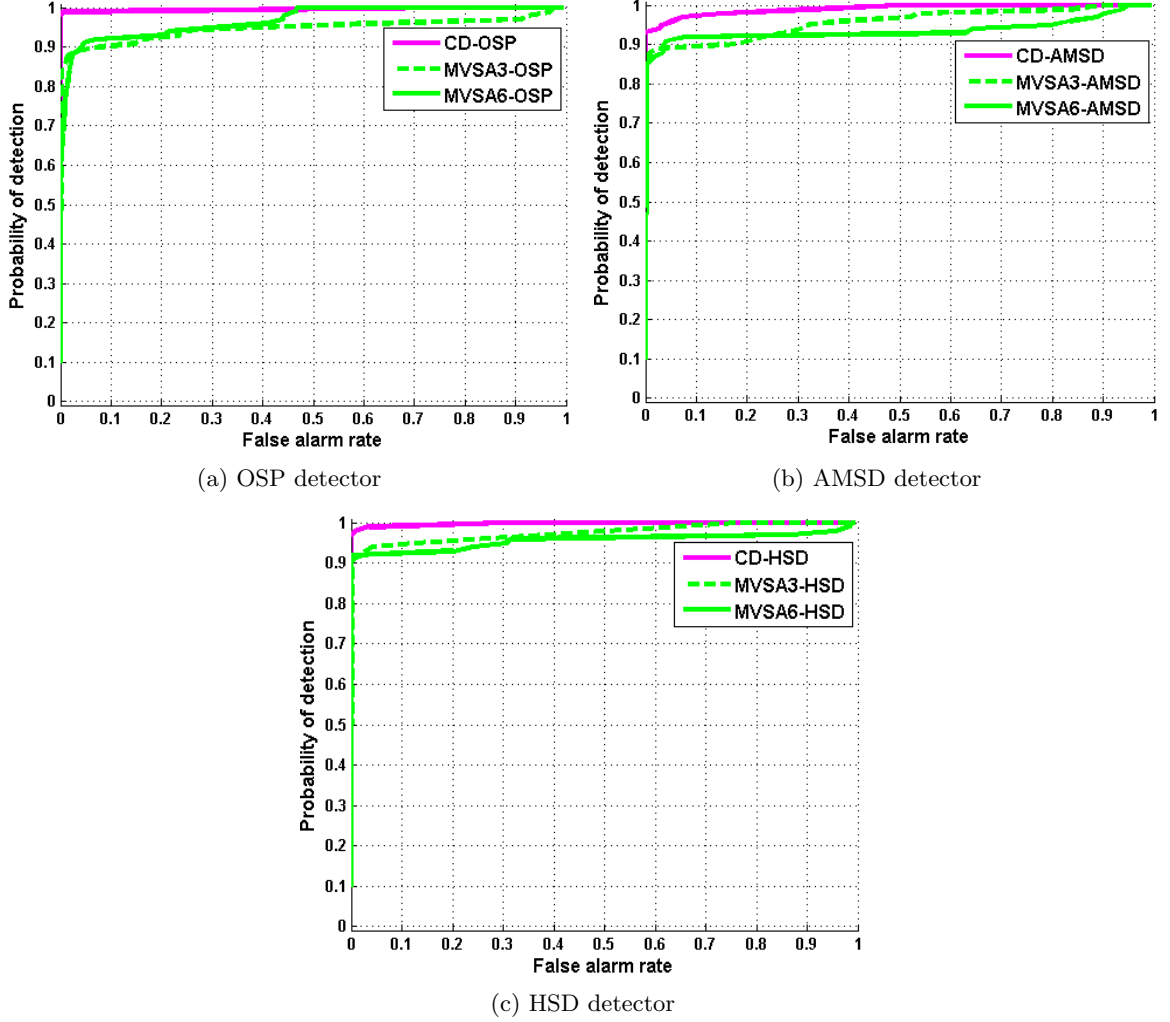


Figure 6.42: Receiver operating characteristic (ROC) curves (**Indian Pines** data).

(ROC) curves. In figure 6.42, we plot the ROC curves of the proposed context dependent detectors versus those of the traditional detectors using MVSA. It can be seen from figure 6.42 that the proposed context dependent target detectors outperform the traditional MVSA single subspace detectors. We also evaluate the performance of the detectors using the Area Under the ROC Curve (AUC). Table 6.10 reports the average and standard deviation of this measure for the CD, EigVect, MVSA, NFINDR and PPI approaches using the OSP, AMSD and HSD detectors over 25 runs.

Using a two-sample t-test at the 5% significance level, we conclude that the proposed context dependent detection approach outperformed the traditional detection approaches when the OSP and AMSD detectors were used (p -value < 0.003). For HSD, the correlation matrix eigen vectors based method outperformed the rest of the methods (p -value $< 1e-10$). However, the context dependent approach still outperformed the other endmember detection methods. This can be explained by the

TABLE 6.10

Average ($\pm$ standard deviation) of the AUC over 25 runs for the CD, EigVect, MVSA, NFINDR and PPI methods (**Indian Pines** data)

Detector	OSP	AMSD	HSD
CD	0.9921 (± 0.0000)	0.9443 (± 0.0507)	0.9977 (± 0.0000)
EigVect3	0.9822 (± 0.0000)	0.8916 (± 0.0000)	0.9978 (± 0.0000)
EigVect6	0.9822 (± 0.0000)	0.8916 (± 0.0000)	0.9978 (± 0.0000)
MVSA3	0.9426 (± 0.0000)	0.8928 (± 0.0500)	0.9725 (± 0.0000)
MVSA6	0.9656 (± 0.0000)	0.8992 (± 0.0476)	0.9537 (± 0.0001)
NFINDR3	0.8270 (± 0.0566)	0.4652 (± 0.1206)	0.6976 (± 0.0006)
NFINDR6	0.6502 (± 0.4482)	0.5316 (± 0.3951)	0.7417 (± 0.0368)
PPI3	0.3911 (± 0.3438)	0.5000 (± 0.0000)	0.1792 (± 0.0590)
PPI6	0.5314 (± 0.4465)	0.5313 (± 0.3947)	0.3357 (± 0.2581)

better description of the background using the RCDSU algorithm, which also ensures that there is no leakage of targets into the endmember sets using the possibilistic memberships. We also notice the robustness of the context dependent approach to initialization (AUC standard deviation < 0.051).

6.9.2 Evaluation using actual targets

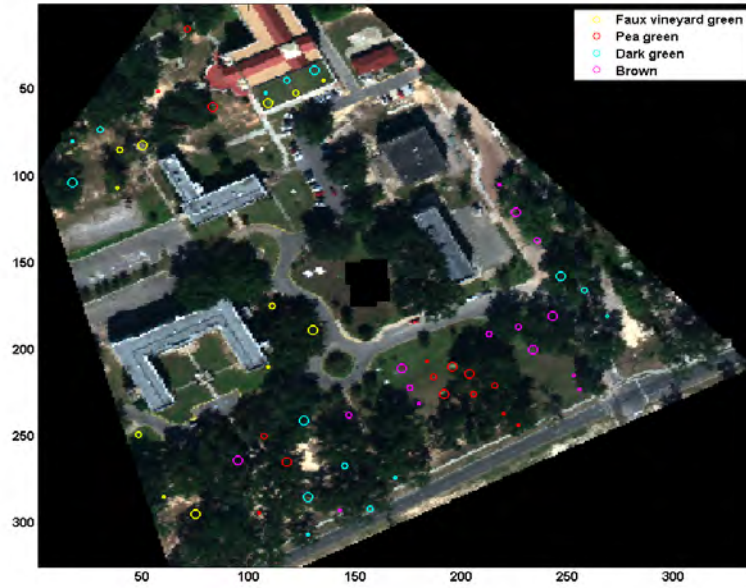
The **University of Southern Mississippi** data described in Section 6.1.2 presents different targets with different sizes. Figure 6.43 shows the subimage (67675 pixels) considered in this experiment, along with the target types, locations and sizes (the circle sizes are relative to the target sizes). The targets were made of 100% cotton fabric and were emplaced so that there would be representatives of each color type completely non-occluded, partially occluded, and almost totally occluded. As a consequence, one can extract pure pixels of most of the colors but it is unlikely that any algorithm will be able to find all the targets. Each target was assigned a confidence number (1- Visible, 2- Probably the target, 3- Possibly the target, 4- Not visible) and a category (0- Target not covered, 1- Covered partly or fully with shadow but no tree, 2- Part or all of the target is covered by a tree branch). Table 6.11 summarizes this information about the targets. We run U-RCDSU using $C_{max} = 5$, $M = 3$, $m = n = 1.5$, $a = 0.1$, $b = 0.9$, $\alpha = 100$, $\beta_i = 4$, $\forall i$, $\sigma_i = 1$, $\forall i$ and $\epsilon = 0.1$. The algorithm converges to $C = 2$ clusters.

We evaluate the performance of the detection methods using the area under the ROC curve measure. Table 6.12 reports this measure for all methods for the “brown” targets. Two cases are analyzed: all target sizes combined and subpixel targets only. For all target sizes combined, we can see that the proposed context dependent approach gave good AUCs, outperforming the other methods when the AMSD and HSD detectors were used (p -value < 0.02). For the subpixel targets,

TABLE 6.11

Targets in the **University of Southern Mississippi** data

Target	Size			Confidence				Category			Total
	0.5	1	3	1	2	3	4	0	1	2	
Brown	5	5	5	2	3	1	9	4	3	8	15
Dark green	5	5	5	2	0	1	12	3	2	10	15
Faux vineyard green	5	5	5	3	3	3	6	4	2	9	15
Pea green	5	5	5	2	2	3	8	4	1	10	15
Total	20	20	20	9	8	8	35	15	8	37	60

Figure 6.43: Subimage of the **University of Southern Mississippi** data with target types, locations and sizes

the context dependent target detectors gave the best AUC when used with HSD (p -value < 0.05).

Table 6.13 reports this measure for all methods for the “dark green” targets, when all target sizes are considered and when subpixel targets only are considered. For all target sizes combined, we can see that the proposed context dependent approach gave good AUCs, outperforming the other methods when the OSP and HSD detectors were used (p -value < 0.006). For the subpixel targets, MVSA3 outperformed all other methods (p -value < 0.003).

Table 6.14 reports this measure for all methods for the “faux vineyard green” targets. For all target sizes combined, we can see that the proposed context dependent approach gave good AUCs, outperforming the other methods when the HSD detector is used (p -value < 0.005). However,

TABLE 6.12

Average ($\pm$ standard deviation) of the AUC over 25 runs for the CD, EigVect, MVSA, NFINDR and PPI methods (**University of Southern Mississippi** data, brown target)

Detector	OSP	AMSD	HSD
AUC for all target sizes			
CD	0.9657 (± 0.0000)	0.9812 (± 0.0000)	0.9778 (± 0.0000)
EigVect3	0.8665 (± 0.0000)	0.9365 (± 0.0000)	0.9107 (± 0.0000)
EigVect6	0.9665 (± 0.0000)	0.9365 (± 0.0000)	0.9107 (± 0.0000)
MVSA3	0.9652 (± 0.0000)	0.9788 (± 0.0000)	0.9671 (± 0.0214)
MVSA6	0.9664 (± 0.0000)	0.9697 (± 0.0000)	0.9429 (± 0.0128)
NFINDR3	0.7892 (± 0.0000)	0.8328 (± 0.0000)	0.7951 (± 0.0037)
NFINDR6	0.9252 (± 0.0257)	0.9359 (± 0.0186)	0.8580 (± 0.0938)
PPI3	0.8505 (± 0.0506)	0.9011 (± 0.0382)	0.8800 (± 0.0086)
PPI6	0.8741 (± 0.0095)	0.9301 (± 0.0058)	0.8709 (± 0.0001)
AUC for target size 0.5			
CD	0.9773 (± 0.0000)	0.9854 (± 0.0000)	0.9891 (± 0.0000)
EigVect3	0.9229 (± 0.0000)	0.9795 (± 0.0000)	0.9727 (± 0.0000)
EigVect6	0.9229 (± 0.0000)	0.9795 (± 0.0000)	0.9727 (± 0.0000)
MVSA3	0.9779 (± 0.0000)	0.9846 (± 0.0000)	0.9654 (± 0.0000)
MVSA6	0.9817 (± 0.0000)	0.9774 (± 0.0000)	0.9659 (± 0.0001)
NFINDR3	0.8656 (± 0.0000)	0.8979 (± 0.0000)	0.9823 (± 0.0000)
NFINDR6	0.9765 (± 0.0173)	0.9884 (± 0.0072)	0.9190 (± 0.1225)
PPI3	0.8973 (± 0.0424)	0.9243 (± 0.0509)	0.9414 (± 0.0070)
PPI6	0.8805 (± 0.0151)	0.9723 (± 0.0052)	0.9318 (± 0.0009)

MVSA6 with AMSD gave the highest AUC. For subpixel targets, MVSA3 with HSD gave the highest AUC (p -value < 0.003).

Table 6.15 reports this measure for all methods for the “pea green” targets. When all target sizes are considered, we can see that the proposed context dependent approach gave good AUCs, however not as good as the MVSA3 with AMSD and HSD, and MVSA6 with OSP (p -value $< 1e-9$). For subpixel targets, the context dependent target detection yielded the highest AUC when HSD was used (p -value < 0.01).

We can also notice that the context dependent target detection approach is robust to initialization (very small standard deviation).

In contrast to the Indian Pines data with implanted targets, the context dependent approach to target detection was not always the best compared to the traditional detectors. The nature of the targets may explain this. In fact, most of the targets (45 out of 60) are covered partly or fully whether by shadows of trees. Moreover, only 9 out of 60 targets are labeled as “visible” in the ground truth. Most of the remaining ones (35 out of 51) are not visible. So, detecting these would be difficult. Actually, if detected, these should be considered as false alarms. Unfortunately, the

TABLE 6.13

Average ($\pm$ standard deviation) of the AUC over 25 runs for the CD, EigVect, MVSA, NFINDR and PPI methods (**University of Southern Mississippi** data, dark green target)

Detector	OSP	AMSD	HSD
AUC for all target sizes			
CD	0.9533 (± 0.0000)	0.9760 (± 0.0000)	0.9781 (± 0.0000)
EigVect3	0.9171 (± 0.0000)	0.9251 (± 0.0000)	0.9341 (± 0.0000)
EigVect6	0.9171 (± 0.0000)	0.9251 (± 0.0000)	0.9341 (± 0.0000)
MVSA3	0.9510 (± 0.0000)	0.9795 (± 0.0000)	0.9558 (± 0.0000)
MVSA6	0.9149 (± 0.0000)	0.9518 (± 0.0000)	0.9165 (± 0.0000)
NFINDR3	0.8874 (± 0.0000)	0.8649 (± 0.0000)	0.8676 (± 0.0000)
NFINDR6	0.9088 (± 0.0202)	0.9224 (± 0.0161)	0.8370 (± 0.0392)
PPI3	0.9418 (± 0.0199)	0.9672 (± 0.0040)	0.9189 (± 0.0173)
PPI6	0.9179 (± 0.0143)	0.9569 (± 0.0089)	0.8978 (± 0.0030)
AUC for target size 0.5			
CD	0.9620 (± 0.0000)	0.9679 (± 0.0000)	0.9829 (± 0.0000)
EigVect3	0.9132 (± 0.0000)	0.9693 (± 0.0000)	0.9765 (± 0.0000)
EigVect6	0.9132 (± 0.0000)	0.9693 (± 0.0000)	0.9765 (± 0.0000)
MVSA3	0.9634 (± 0.0000)	0.9758 (± 0.0000)	0.9858 (± 0.0000)
MVSA6	0.9309 (± 0.0000)	0.9493 (± 0.0000)	0.9453 (± 0.0000)
NFINDR3	0.9342 (± 0.0000)	0.9197 (± 0.0000)	0.9397 (± 0.0000)
NFINDR6	0.9399 (± 0.0065)	0.9467 (± 0.0088)	0.8907 (± 0.1154)
PPI3	0.9629 (± 0.0028)	0.9737 (± 0.0086)	0.9485 (± 0.0019)
PPI6	0.9515 (± 0.0148)	0.9724 (± 0.0012)	0.9468 (± 0.0011)

number of visible targets for each type is not enough to evaluate the methods on them only. This can be seen in figure 6.44 where 3 principal components (PC) of the data points are shown in blue dots while the pixels labeled as targets are shown in red dots. It can be seen that most of the targets are located in the background cloud and that only few of them are “outliers”. This makes them difficult to detect. This illustration should be considered with a grain of salt as it only represents the data in a reduced dimension space.

Another explanation would be the presence of many outliers in the data (as it can be seen in figure 6.44). The robust unmixing did not take these into account in the endmember sets formation. Hence, they do not fit the background model and will result in a high detection score as opposed to the global non robust unmixing methods which consider them in the background modeling, and hence not detecting them as false alarms.

TABLE 6.14

Average ($\pm$ standard deviation) of the AUC over 25 runs for the CD, EigVect, MVSA, NFINDR and PPI methods (**University of Southern Mississippi** data, faux vineyard green target)

Detector	OSP	AMSD	HSD
AUC for all target sizes			
CD	0.9328 (± 0.0000)	0.9439 (± 0.0000)	0.9695 (± 0.0000)
EigVect3	0.9078 (± 0.0000)	0.8917 (± 0.0000)	0.8416 (± 0.0000)
EigVect6	0.9078 (± 0.0000)	0.8917 (± 0.0000)	0.8416 (± 0.0000)
MVSA3	0.9317 (± 0.0000)	0.9378 (± 0.0000)	0.9639 (± 0.0000)
MVSA6	0.9448 (± 0.0000)	0.9791 (± 0.0000)	0.9394 (± 0.0000)
NFINDR3	0.8821 (± 0.0000)	0.8408 (± 0.0000)	0.8825 (± 0.0000)
NFINDR6	0.9493 (± 0.0160)	0.9523 (± 0.0226)	0.9246 (± 0.0714)
PPI3	0.8180 (± 0.0659)	0.9672 (± 0.0040)	0.8134 (± 0.0533)
PPI6	0.9405 (± 0.0034)	0.9569 (± 0.0089)	0.9056 (± 0.0016)
AUC for target size 0.5			
CD	0.9668 (± 0.0000)	0.9708 (± 0.0000)	0.9891 (± 0.0000)
EigVect3	0.9692 (± 0.0000)	0.9390 (± 0.0000)	0.9211 (± 0.0000)
EigVect6	0.9692 (± 0.0000)	0.9390 (± 0.0000)	0.9211 (± 0.0000)
MVSA3	0.9678 (± 0.0000)	0.9704 (± 0.0000)	0.9893 (± 0.0000)
MVSA6	0.9585 (± 0.0000)	0.9792 (± 0.0000)	0.9428 (± 0.0000)
NFINDR3	0.9376 (± 0.0000)	0.9038 (± 0.0000)	0.9679 (± 0.0000)
NFINDR6	0.9677 (± 0.0141)	0.9706 (± 0.0174)	0.9842 (± 0.0072)
PPI3	0.8465 (± 0.0646)	0.8363 (± 0.0242)	0.8937 (± 0.0529)
PPI6	0.9727 (± 0.0012)	0.9186 (± 0.0001)	0.9362 (± 0.0006)

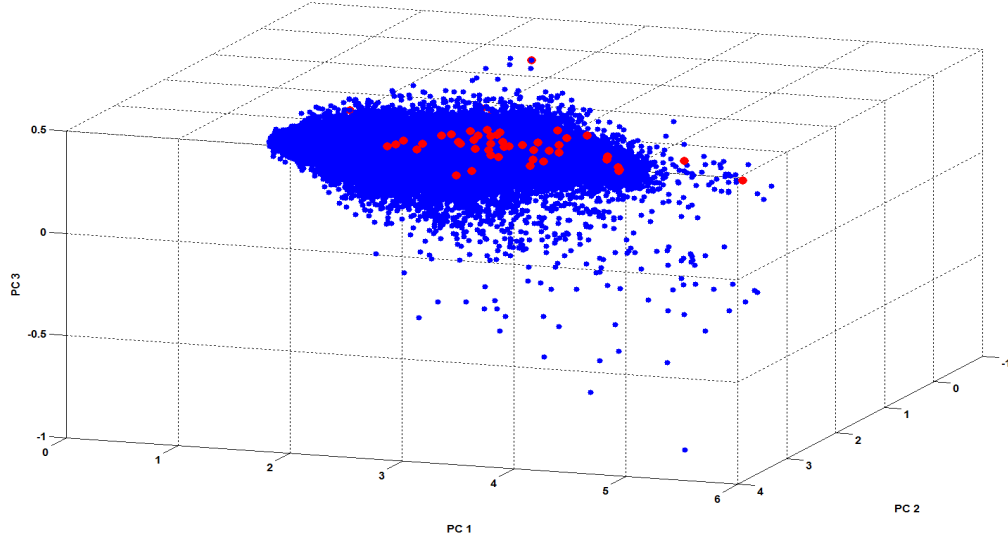


Figure 6.44: Principal components of the **University of Southern Mississippi** data (blue dots) with targets (red dots).

TABLE 6.15

Average ($\pm$ standard deviation) of the AUC over 25 runs for the CD, EigVect, MVSA, NFINDR and PPI methods (**University of Southern Mississippi** data, pea green target)

Detector	OSP	AMSD	HSD
AUC for all target sizes			
CD	0.9563 (± 0.0000)	0.9672 (± 0.0000)	0.9693 (± 0.0000)
EigVect3	0.9177 (± 0.0000)	0.7572 (± 0.0000)	0.7230 (± 0.0000)
EigVect6	0.9177 (± 0.0000)	0.7572 (± 0.0000)	0.7230 (± 0.0000)
MVSA3	0.9532 (± 0.0000)	0.9675 (± 0.0000)	0.9748 (± 0.0000)
MVSA6	0.9663 (± 0.0000)	0.9245 (± 0.0000)	0.9480 (± 0.0000)
NFINDR3	0.8604 (± 0.0000)	0.8264 (± 0.0000)	0.7816 (± 0.0000)
NFINDR6	0.9582 (± 0.0249)	0.9426 (± 0.0113)	0.6097 (± 0.1712)
PPI3	0.8200 (± 0.0480)	0.8519 (± 0.0557)	0.7358 (± 0.0348)
PPI6	0.9657 (± 0.0003)	0.9281 (± 0.0002)	0.7126 (± 0.0000)
AUC for target size 0.5			
CD	0.9741 (± 0.0000)	0.9472 (± 0.0000)	0.9769 (± 0.0000)
EigVect3	0.9524 (± 0.0000)	0.7841 (± 0.0000)	0.7900 (± 0.0000)
EigVect6	0.9524 (± 0.0000)	0.7841 (± 0.0000)	0.7900 (± 0.0000)
MVSA3	0.9729 (± 0.0000)	0.9544 (± 0.0000)	0.9735 (± 0.0000)
MVSA6	0.9663 (± 0.0001)	0.8925 (± 0.0000)	0.9469 (± 0.0001)
NFINDR3	0.9344 (± 0.0000)	0.8400 (± 0.0000)	0.7774 (± 0.0000)
NFINDR6	0.9672 (± 0.0224)	0.9351 (± 0.0223)	0.7299 (± 0.1899)
PPI3	0.8586 (± 0.0790)	0.8628 (± 0.0450)	0.7956 (± 0.0158)
PPI6	0.9766 (± 0.0002)	0.9316 (± 0.0002)	0.7892 (± 0.0001)

CHAPTER 7

CONCLUSIONS AND POTENTIAL FUTURE WORK

7.1 Conclusions

A hyperspectral unmixing algorithm, called Context Dependent Spectral Unmixing (CDSU), that finds multiple sets of endmembers is proposed. Unlike existing hyperspectral unmixing methods, CDSU is a local approach that adapts the unmixing to different regions of the spectral space. Consequently, it finds multiple sets of endmembers that represent semantically meaningful regions of the hyperspectral image.

CDSU is based on defining and optimizing a novel objective function that combines context identification and hyperspectral unmixing into a joint function. This function models contexts as compact clusters and uses the linear mixing model as the basis for unmixing. The unmixing provides optimal endmembers and abundances for each context. The performance of CDSU was evaluated and compared to similar existing methods using synthetic and real data. We showed that the proposed method can identify meaningful and coherent contexts, and appropriate endmembers within each context. We also showed that CDSU is more robust to noise than similar existing methods.

Several extensions of the CDSU approach were also proposed. Due to the high dimensional and correlated nature of the hyperspectral data, the Euclidean distance used in CDSU can be restrictive in the sense that it limits the clusters to the spherical shape. The Context Dependent Spectral Unmixing using the Mahalanobis distance (CDSU_M) was proposed to overcome this limitation and account for non-spherical clusters. This was achieved by using the Mahalanobis distance instead of the Euclidean distance in the clustering component of CDSU. This allows more flexibility in the cluster shapes and can identify ellipsoidal clusters where the variance in each feature dimension is considered.

CDSU is based in part on clustering. However, clustering by itself is a challenging task, especially in a high dimensional space as is the case of hyperspectral data. Many local minima solutions may exist. To overcome this problem, we proposed two semi-supervised algorithms of CDSU_M. These methods use partial supervision information to guide the search process and narrow the space

of possible solutions. The supervision information consists of small sets of pairwise constraints which can be obtained from multiple sources of information, such as labeling few pixels in the hyperspectral image or using information extracted from a different sensor. The first semi-supervised CDSU algorithm, called Cluster Constrained Multi-Model Unmixing (CC-MMU), uses constraints derived from the cluster assignments of the pixels. The second algorithm, called Proportion Constrained Multi-Model Unmixing (PC-MMU), uses constraints derived from the proportion values of the pixels. To validate both proposed semi-supervised CDSU algorithms, we used real data and we constructed a set of constraints using information provided by a LIDAR sensor, as well as information extracted from the consensus of multiple unmixing algorithms.

CDSU is sensitive to noise and outliers present in the hyperspectral data due to scene and/or sensor effects. This is mainly inherited from the fuzzy clustering component of CDSU. Noise points affect not only the resulting partition, but also the estimated endmembers and proportions within each cluster. Possibilistic clustering has been used to overcome the sensitivity of fuzzy clustering to noise. This approach uses possibilistic memberships to identify and reduce the effect of noise points. These unconstrained memberships may however result in identical clusters. In order to avoid this problem, we proposed a Robust Context Dependent Spectral Unmixing (RCDSU) algorithm that uses both fuzzy and possibilistic memberships. Fuzzy memberships are used to partition the spectral space into multiple sets that span the entire space and avoid coincident clusters, while possibilistic memberships are used to reduce the effect of noise and obtain robust estimates of the endmembers and proportions within each set.

The proposed context dependent spectral unmixing algorithms, like other multi-model unmixing algorithms, assume that the optimal number of endmember sets is known a priori. However, this may not be the case, and it should be learned from the data. To address this challenge, we proposed an Unsupervised Robust Context Dependent Spectral Unmixing (U-RCDSU) algorithm that finds the number of contexts in the data in an unsupervised way. U-RCDSU exploits the fact that the possibilistic memberships are not constrained to sum to one. U-RCDSU starts by overspecifying the number of contexts, then, as the algorithm iterates, clusters covering the same dense regions would expand and become similar. These are ultimately merged and the number of clusters is reduced.

Spectral unmixing is a challenging, ill-posed, inverse problem. Many algorithms have been proposed for robust, stable, and accurate unmixing solutions. Moreover, different algorithms have different modes of operation and usually yield different results. Even the same algorithm may not

result in the same endmembers when run multiple times. This is mainly due to the non-deterministic behavior of the algorithm. In order to estimate the “optimal” endmembers for a given data set, we proposed taking advantage of the many existing unmixing algorithms and exploring their similarity and difference using consensus analysis. The idea has its roots in consensus clustering. The proposed consensus unmixing combines the results of multiple unmixing algorithms, run multiple times with different parameters, to find consistent endmembers in the data. The claim is that such endmembers will have a consensus among multiple runs. The approach starts by determining the set of points having high proportion values in any of the endmembers of the unmixing ensemble. The points corresponding to non consistent endmembers are then filtered out using a co-association measure, and only points with high co-association values are kept. Finally, the remaining points are clustered and a representative from each cluster is chosen to be a consistent endmember. The points with high co-association values have also been used to form a set of constraints on the proportions for the semi-supervised PC-MMU algorithm.

Spectral unmixing is used as an initial step for many hyperspectral target detection algorithms. The endmembers are used to model the background and compute a detection statistic for all pixels. This is known as structured background target detection. Most of these detection algorithms suffer from two main problems. First, there is the target leakage problem, which is the contribution of the targets to the background model. This is due to the presence of the targets in the scene being modeled. The second problem is the background modeling itself. Traditional target detectors use global unmixing methods to model the background using a single set of endmembers. These may not provide a good description of the hyperspectral data, especially when the scene includes multiple regions with distinct materials. To overcome these problems, we proposed new context dependent target detectors that are based on the robust context dependent spectral unmixing algorithm (RCDSU) to better describe the background and minimize target leakage. Targets can be thought of as outliers and hence the robust unmixing will not consider them when modeling the background. Moreover, the multi-model unmixing provides a better background description. The proposed detectors were applied to the traditional Orthogonal Subspace Projection detector, the Adaptive Matched Subspace Detector and the Hybrid Subspace Detector algorithms. A local detection statistic is computed for each context and then all scores are combined using the fuzzy memberships of the pixels.

The experimental results showed good performance of the proposed methods on synthetic and real data. The lack of reliable ground truth information was a limiting factor when evaluating

the proposed algorithms on real data sets. Qualitative evaluation was employed in most cases. Data with actual targets were scarce. The University of Southern Mississippi data was the only available data with various targets and various background materials. The only drawback was the occlusion which affected most of the targets present in the scene.

7.2 Potential Future Work

Although our proposed work has shown promising results, there is still room for improvement. The following sections list the various areas that will be explored in the future to build upon the developed algorithms.

7.2.1 Large scale evaluation

Hyperspectral data is not very abundant. Only few public data sets are available. Moreover, ground truth information is usually not accurate if not missing. Even though the results on the few used data sets were promising, more evaluation is still required. In particular, evaluation using very large image where the notion of contexts is more important.

The National Science Foundation funded National Ecological Observatory Network (NEON¹) will be one main source of data. NEON is a continental-scale ecological observation system for examining critical ecological issues. NEON is designed to gather and synthesize data on the impacts of climate change, land use change, and invasive species on natural resources and biodiversity. Data will be collected from 106 sites (60 terrestrial, 36 aquatic and 10 aquatic experimental) across the U.S. using instrument measurements and field sampling. The sites have been strategically selected to represent different regions of vegetation, landforms, climate, and ecosystem performance. NEON will combine site-based data with remotely sensed data and existing continental-scale data sets (e.g. satellite data) to provide a range of scaled data products that can be used to describe changes in the nations ecosystem through space and time.

NEON successfully completed the planning and design phases and entered the construction phase in Spring 2012. It is currently building sites. Constructing the entire NEON network will take approximately five years, so it is expected to be in full operation by approximately 2017. NEON will collect data for 30 years and will have an open-access approach to its data and information products. Once this data is collected and made available, it can be used to compare existing multi-model unmixing methods.

¹See more at: <http://www.neoninc.org/about/overview#sthash.oZ8Jcudt.dpuf>

7.2.2 Target detection

Subpixel target detection is a difficult task. Our experiments showed that no single detector can consistently outperform the others. The fusion of multiple detectors is a potential area of research. This would overcome the weaknesses of some detectors and take advantage of the strength of the others. More importantly, fusion could be adapted to the different contexts identified during the unmixing process.

Currently, in the proposed context dependent detectors, only fuzzy memberships are used. Future work may include investigating the use of the possibilistic memberships in the detectors, as opposed to only using them in estimating the endmember sets.

7.2.3 Multi-sensor fusion

Hyperspectral sensors can be used along with other remote sensors such as LIDAR and Synthetic Aperture Radar (SAR) in order to take advantage of the different information provided by each sensor. For instance, LIDAR provides elevation information, while SAR provides, among other, structural information about the scene. This would help directly or indirectly (e.g. generating supervision constraints) in the unmixing process.

7.2.4 Non-linear unmixing

The proposed work has focused only on the linear mixing model. This assumes that materials are mainly mixed due to the spatial resolution of the sensor, and that the intimate mixing of the materials on the ground is negligible. As a potential future work, the proposed multi-model formulation could be used to generalize non-linear unmixing methods.

REFERENCES

- [1] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 44–57, 2002.
- [2] M. O. Smith, P. E. Johnson, and J. B. Adams, "Quantitative determination of mineral types and abundances from reflectance spectra using principal components analysis," *Journal of Geophysical Research*, vol. 90, pp. 797–804, 1985.
- [3] A. R. Gillespie, M. O. Smith, J. B. Adams, S. C. Willis, A. F. Fischer, and D. E. Sabol, "Interpretation of residual images: spectral mixture analysis of aviris images, owens valley, california," in *Second Airborne Visible Infrared Imaging Spectrometer AVIRIS Workshop*, 1990, vol. 90-54, pp. 243–270.
- [4] J. B. Adams, M. O. Smith, and P. E. Johnson, "Spectral mixture modeling: a new analysis of rock and soil types at the viking lander 1 site," *Journal of Geophysical Research*, vol. 91, no. B8, pp. 8098–8112, 1986.
- [5] G. Swayze, R. N. Clark, F. Kruse, S. Sutley, and A. Gallagher, "Ground-truthing aviris mineral mapping at cuprite, nevada," in *Third Annual JPL Airborne Geoscience Workshop*, 1992, vol. 1, pp. 47–49.
- [6] R. O. Green, M. L. Eastwood, C. M. Sarture, T. G. Chrien, M. Aronsson, B. J. Chippendale, J. A. Faust, B. E. Pavri, C. J. Chovit, M. Solis, M. R. Olah, and O. Williams, "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (aviris)," *Remote Sensing of Environment*, vol. 65, no. 3, pp. 227–248, 1998.
- [7] G. Vane, R. O. Green, T. G. Chrien, H. T. Enmark, E. G. Hansen, and W. M. Porter, "The airborne visible/infrared imaging spectrometer (aviris)," *Remote Sensing of Environment*, vol. 44, no. 23, pp. 127–143, 1993.
- [8] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 29–43, 2002.
- [9] G. Shaw and D. Manolakis, "Signal processing for hyperspectral image exploitation," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 12–16, 2002.
- [10] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 17–28, 2002.
- [11] D. W. J. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 58–69, 2002.
- [12] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sensing of Environment*, vol. 113, pp. S110–S122, 2009.
- [13] M. E. Schaepman, S. L. Ustin, A. J. Plaza, T. H. Painter, J. Verrelst, and S. Liang, "Earth system science related imaging spectroscopy assessment," *Remote Sensing of Environment*, vol. 113, pp. S123–S137, 2009.
- [14] M. Berman, P. M. Connor, L. B. Whitbourn, D. A. Coward, B. G. Osborne, and M. D. Southan, "Classification of sound and stained wheat grains using visible and near infrared hyperspectral image analysis," *Journal of Near Infrared Spectroscopy*, vol. 15, no. 6, pp. 351–358, 2007.

- [15] A. A. Gowen, C. P. O'Donnell, P. J. Cullen, G. Downey, and J. M. Frias, "Hyperspectral imaging an emerging process analytical tool for food quality and safety control," *Trends in Food Science and Technology*, vol. 18, no. 12, pp. 590–598, 2007.
- [16] S. Mahesh, A. Manickavasagan, D. S. Jayas, J. Paliwal, and N. D. G. White, "Feasibility of near-infrared hyperspectral imaging to differentiate canadian wheat classes," *Biosystems Engineering*, vol. 101, no. 1, pp. 50–57, 2008.
- [17] R. Larsen, M. Arngren, P. W. Hansen, and A. A. Nielsen, "Kernel based subspace projection of near infrared hyperspectral images of maize kernels," in *Proceedings of the 16th Scandinavian Conference on Image Analysis*, 2009, vol. 5575, pp. 560–569.
- [18] M. S. Kim, Y. R. Chen, and P. M. Mehl, "Hyperspectral reflectance and fluorescence imaging system for food quality and safety," *Transactions of the American Society of Agricultural Engineers*, vol. 44, no. 3, pp. 721–729, 2001.
- [19] O. Y. Rodionova, L. P. Houmller, A. L. Pomerantsev, P. Geladi, J. Burger, V. L. Dorofeyev, and A. P. Arzamastsev, "Nir spectrometry for counterfeit drug detection: A feasibility study," *Analytica Chimica Acta*, vol. 549, no. 12, pp. 151–158, 2005.
- [20] C. Gendrin, Y. Roggo, and C. Collet, "Pharmaceutical applications of vibrational chemical imaging and chemometrics: A review," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 48, no. 3, pp. 533–553, 2008.
- [21] A. de Juan, M. Maeder, T. Hanczewicz, L. Duponchel, and R. Tauler, *Chemometric tools for image analysis*, pp. 65–109, Wiley-VCH Verlag GmbH and Co. KGaA, 2009.
- [22] M. B. Lopes, J.-C. Wolff, J. Bioucas-Dias, and M. Figueiredo, "Nir hyperspectral unmixing based on a minimum volume criterion for fast and accurate chemical characterization of counterfeit tablets," *Analytical Chemistry*, vol. 82, no. 4, pp. 14621469, 2010.
- [23] G. Begelman, M. Zibulevsky, E. Rivlin, and T. Kolatt, "Blind decomposition of transmission light microscopic hyperspectral cube using sparse representation," *IEEE Transactions on Medical Imaging*, vol. 28, no. 8, pp. 1317–1324, 2009.
- [24] H. Akbari, Y. Kosugi, K. Kojima, and N. Tanaka, "Detection and analysis of the intestinal ischemia using visible and invisible hyperspectral imaging," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 8, pp. 2011–2017, 2010.
- [25] A. Picon, O. Ghita, P. F. Whelan, and P. M. Iriondo, "Fuzzy spectral and spatial feature integration for classification of nonferrous materials in hyperspectral data," *IEEE Transactions on Industrial Informatics*, vol. 5, no. 4, pp. 483–494, 2009.
- [26] I. C. Chein, "Multiparameter receiver operating characteristic analysis for signal detection and classification," *IEEE Sensors Journal*, vol. 10, no. 3, pp. 423–442, 2010.
- [27] L. N. Brewer, J. A. Ohlhausen, P. G. Kotula, and J. R. Michael, "Forensic analysis of bioagents by x-ray and tof-sims hyperspectral imaging," *Forensic science international*, vol. 179, no. 2, pp. 98–106, 2008.
- [28] D. Manolakis, D. Marden, and G. A. Shaw, "Hyperspectral image processing for automatic target detection applications," *Lincoln Laboratory Journal*, vol. 14, no. 1, pp. 79–116, 2003.
- [29] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Du Qian, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 354–379, 2012.
- [30] N. Keshava, J. Kerekes, D. Manolakis, and G. Shaw, "An algorithm taxonomy for hyperspectral unmixing," in *SPIE Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI*, 2000, vol. 4049, pp. 42–63.
- [31] S. Liangrocapart and Maria Petrou, "Mixed pixels classification," in *SPIE Image and Signal Processing for Remote Sensing IV*, 1998, vol. 3500, pp. 72–83.

- [32] R. B. Singer and T. B. McCord, "Mars - large scale mixing of bright and dark surface materials and implications for analysis of spectral reflectance," in *10th Lunar and Planetary Science Conference*, 1979, vol. 2, pp. 1835–1848.
- [33] B. Hapke, "Bidirectional reflectance spectroscopy 1. theory," *Journal of Geophysical Research*, vol. 86, pp. 3039–3054, 1981.
- [34] R. N. Clark and T. L. Roush, "Reflectance spectroscopy: quantitative analysis techniques for remote sensing applications," *Journal of Geophysical Research*, vol. 89, no. 7, pp. 6329–6340, 1984.
- [35] B. Hapke, *Theory of reflectance and emittance spectroscopy*, Cambridge University Press, 1993.
- [36] J. M. Bioucas-Dias and A. Plaza, "Hyperspectral unmixing: geometrical, statistical, and sparse regression-based approaches," in *SPIE Image and Signal Processing for Remote Sensing XVI*, vol. 7830, pp. 1–15.
- [37] A. Plaza, G. Martin, J. Plaza, M. Zortea, and S. Sanchez, *Recent developments in endmember extraction and spectral unmixing*, vol. 3 of *Advances in Signal Processing and Exploitation Techniques*, pp. 235–267, Springer, 2011.
- [38] M. Parente and A. Plaza, "Survey of geometric and statistical unmixing algorithms for hyperspectral images," in *2nd IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2010, pp. 1–4.
- [39] A. Plaza, P. Martinez, R. Perez, and J. Plaza, "A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 3, pp. 650–663, 2004.
- [40] G. A. Shaw and H. K. Burke, "Spectral imaging for remote sensing," *Lincoln Laboratory Journal*, vol. 14, no. 1, pp. 3–28, 2003.
- [41] Y. H. Hu, H. B. Lee, and F. L. Scarpace, "Optimal linear spectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 1, pp. 639–644, 1999.
- [42] M. Petrou and P. G. Foschi, "Confidence in linear spectral unmixing of single pixels," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 1, pp. 624–626, 1999.
- [43] J. J. Settle, "On the relationship between spectral unmixing and subspace projection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, no. 4, pp. 1045–1046, 1996.
- [44] A. S. Mazer, M. Martin, M. Lee, and J. E. Solomon, "Image processing software for imaging spectrometry data analysis," *Remote Sensing of Environment*, vol. 24, no. 1, pp. 201–210, 1988.
- [45] R. Yuhas, A. F. H. Goetz, and J. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (sam) algorithm," in *Third Annual JPL Airborne Geoscience Workshop*, vol. 1, pp. 147–149.
- [46] J. C. Harsanyi and C. I. Chang, "Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 4, pp. 779–785, 1994.
- [47] I. C. Cheng, Z. Xiao-Li, M. L. G. Althouse, and P. Jeng Jong, "Least squares subspace projection approach to mixed pixel classification for hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 3, pp. 898–912, 1998.
- [48] D. Heinz, C. I. Chang, and M. L. G. Althouse, "Fully constrained least-squares based linear unmixing," in *Proceedings IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 1999, vol. 2, pp. 1401–1403.
- [49] A. Zare, P. Gader, O. Bchir, and H. Frigui, "Piecewise convex multiple-model endmember detection and spectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. PP, no. 99, pp. 1–10, 2012.

- [50] O. Bchir, H. Frigui, A. Zare, and P. Gader, "Multiple model endmember detection based on spectral and spatial information," in *2nd IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2010, pp. 1–4.
- [51] A. Zare, O. Bchir, H. Frigui, and P. Gader, "Spatially-smooth piece-wise convex endmember detection," in *2nd IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2010, pp. 1–4.
- [52] A. Zare, O. Bchir, H. Frigui, and P. Gader, "A comparison of deterministic and probabilistic approaches to endmember representation," in *2nd IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2010, pp. 1–4.
- [53] A. Zare and P. Gader, "Piece-wise convex spatial-spectral unmixing of hyperspectral imagery using possibilistic and fuzzy clustering," in *IEEE International Conference on Fuzzy Systems (FUZZ)*, 2011, pp. 741–746.
- [54] H. Jenzri, H. Frigui, and P. Gader, "Context dependent spectral unmixing," in *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, sept. 2012, pp. 1–6.
- [55] H. Jenzri, H. Frigui, and P. Gader, "Graph constrained multi-model unmixing using lidar information," in *Fifth IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, jun. 2013.
- [56] H. Jenzri, H. Frigui, and P. Gader, "Robust context dependent spectral unmixing," in *22nd International Conference on Pattern Recognition (ICPR)*, 2014, Accepted.
- [57] A. Strehl and J. Ghosh, "Cluster ensembles - a knowledge reuse framework for combining multiple partitions," *J. Mach. Learn. Res.*, vol. 3, pp. 583–617, Dec. 2002.
- [58] H. Jenzri, H. Frigui, and P. Gader, "Context dependent hyperspectral subpixel target detection," in *IEEE International Conference on Image Processing 2014 (ICIP 2014)*, Paris, France, oct 2014.
- [59] J. W. Boardman, F. A. Kruse, and R. O. Green, "Mapping target signatures via partial unmixing of AVIRIS data," in *Fifth JPL Airborne Earth Science Workshop*. 1995, pp. 23–26, JPL Publication.
- [60] M. E. Winter, "N-findr: an algorithm for fast autonomous spectral end-member determination in hyperspectral data," in *SPIE Imaging Spectrometry V*, 1999, vol. 3753, pp. 266–275.
- [61] M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Dunne, and J. F. Huntington, "Ice: a statistical approach to identifying endmembers in hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 10, pp. 2085–2095, 2004.
- [62] D. M. Rogge, B. Rivard, J. Zhang, A. Sanchez, J. Harris, and J. Feng, "Integration of spatial-spectral information for the improved extraction of endmembers," *Remote Sensing of Environment*, vol. 110, no. 3, pp. 287–303, 2007.
- [63] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 26, no. 1, pp. 65–74, 1988.
- [64] J. B. Lee, A. S. Woodyatt, and M. Berman, "Enhancement of high spectral resolution remote-sensing data by a noise-adjusted principal components transform," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 28, no. 3, pp. 295–304, 1990.
- [65] R. A. Neville, K. Staenz, T. Szeredi, J. Lefebvre, and P. Hauff, "Automatic endmember extraction from hyperspectral data for mineral exploration," in *Canadian Symposium on Remote Sensing*, 1999, pp. 21–24.
- [66] J. M. P. Nascimento and J. M. B. Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, 2005.

- [67] I. C. Chein, W. Chao-Cheng, L. Wei-min, and O. Yen-Chieh, "A new growing method for simplex-based endmember extraction algorithm," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 10, pp. 2804–2819, 2006.
- [68] J. H. Gruninger, A. J. Ratkowski, and M. L. Hoke, "The sequential maximum angle convex cone (smacc) endmember model," pp. 1–14, 2004, 10.1117/12.543794.
- [69] T. Xuetao, W. Bin, Z. Liming, and Z. Jian Qiu, "A new scheme for decomposition of mixed pixels based on nonnegative matrix factorization," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2007, pp. 1759–1762.
- [70] J. Li and J.M. Bioucas-Dias, "Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2008, vol. 3, pp. III–250–III–253.
- [71] J. M. Bioucas-Dias, "A variable splitting augmented lagrangian approach to linear spectral unmixing," in *1st IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2009, pp. 1–4.
- [72] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Kluwer Academic Publishers, 1981.
- [73] S. Krinidis and V. Chatzis, "A robust fuzzy local information c-means clustering algorithm," *IEEE Transactions on Image Processing*, vol. 19, no. 5, pp. 1328–1337, 2010.
- [74] H. Frigui and R. Krishnapuram, "Clustering by competitive agglomeration," *Pattern Recognition*, vol. 30, no. 7, pp. 1109–1119, 1997.
- [75] N. R. Pal, K. Pal, J. M. Keller, and J. C. Bezdek, "A possibilistic fuzzy c-means clustering algorithm," *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 4, pp. 517–530, 2005.
- [76] D. Manolakis, C. Siracusa, and G. Shaw, "Hyperspectral subpixel target detection using the linear mixing model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 7, pp. 1392–1409, 2001.
- [77] L. L. Scharf and L. T. McWhorter, "Adaptive matched subspace detectors and adaptive coherence estimators," in *Conference Record of the Thirtieth Asilomar Conference on Signals, Systems and Computers*, Nov 1996, vol. 2, pp. 1114–1117.
- [78] E. J. Kelly, "An adaptive detection algorithm," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 22, no. 2, pp. 115–127, march 1986.
- [79] J. Broadwater and R. Chellappa, "Hybrid detectors for subpixel targets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 11, pp. 1891–1903, 2007.
- [80] H. Frigui, L. Zhang, and P. Gader, "Context-dependent multi-sensor fusion for landmine detection," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, july 2008, vol. 2, pp. II–371 –II–374.
- [81] H. Frigui, L. Zhang, and P. Gader, "Context-dependent multisensor fusion and its application to land mine detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 6, pp. 2528–2543, june 2010.
- [82] H. Frigui, P. Gader, and A. C. Ben Abdallah, "A generic framework for context-dependent fusion with application to landmine detection," *Proc. SPIE, Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XIII*, vol. 6953, pp. 69531F–69531F–10, 2008.
- [83] J. Nocedal and S. J. Wright, *Theory of Constrained Optimization, Numerical Optimization*, pp. 304–354, Springer Series in Operations Research and Financial Engineering. Springer New York, 2006.
- [84] J. P. Kerekes and J. E. Baum, "Hyperspectral imaging system modeling," *Lincoln Laboratory Journal*, vol. 14, no. 1, pp. 117–130, 2003.

- [85] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," *IEEE Trans. Fuzzy Syst.*, vol. 1, no. 2, pp. 98–110, 1993.
- [86] H. Frigui and R. Krishnapuram, "A robust algorithm for automatic extraction of an unknown number of clusters from noisy data," *Pattern Recognition Letters*, vol. 17, no. 12, pp. 1223–1232, 1996.
- [87] F. Ayuso, J. Setoain, M. Prieto, C. Tenllado, F. Tirado, J. Plaza, and A. Plaza, "Endmember extraction from hyperspectral imagery using a parallel ensemble approach with consensus analysis," in *IEEE Int. Geosci. Remote Sens. Symp.*, July 2009, vol. 5, pp. V–88–V–91.
- [88] R. N. Clark, G. A. Swayze, A. Gallagher, T. V. King, and W. M. Calvin, "The u.s. geological survey digital spectral library: Version 1: 0.2 to 3.0 μm ," in *U.S. Geological Survey, Sunrise Valley Drive Reston, VA, Open File Report 93-592*, 1993.
- [89] A. L. N. Fred and A. K. Jain, "Combining multiple clusterings using evidence accumulation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 6, pp. 835–850, Jun. 2005.
- [90] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264–323, Sep. 1999.
- [91] P. Gader, A. Zare, R. Close, J. Aitken, and G. Tuell, "Muufi gulfport hyperspectral and lidar airborne data set," Tech. Rep. REP-2013-570, University of Florida, Gainesville, FL, Oct. 2013.
- [92] H. W. Kuhn, "The hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955.
- [93] J. D. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference*, New York: Marcel Dekker, 4th edition, 2003.
- [94] Personal communication with Taylor C. Glenn from the University of Florida, "Lidar image segmentation," jan. 2013.
- [95] A. Martinez-Uso, F. Pla, J. M. Sotoca, and P. Garcia-Sevilla, "Clustering-based hyperspectral band selection using information measures," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 4158–4171, dec. 2007.
- [96] T. C. Havens and J. C. Bezdek, "An efficient formulation of the improved visual assessment of cluster tendency (ivat) algorithm," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 5, pp. 813–822, May 2012.

APPENDIX A

Proof of Theorem 3.2.1

Theorem. *The first and second order conditions yield the following local minimizers of J :*

$$\mathbf{E}_i = \left[\beta_i (M \mathbf{I}_{M \times M} - \mathbf{1}_{M \times M}) + \sum_{j=1}^N u_{ij}^m \mathbf{p}_{ij}^T \mathbf{p}_{ij} \right]^{-1} \left[\sum_{j=1}^N u_{ij}^m \mathbf{p}_{ij}^T \mathbf{x}_j \right], \quad (\text{A.1})$$

$$\mathbf{p}_{ij}^T = \max \left(\left[\mathbf{E}_i \mathbf{E}_i^T \right]^{-1} \left[\mathbf{E}_i \mathbf{x}_j^T + \frac{1 - \mathbf{1}_{1 \times M} (\mathbf{E}_i \mathbf{E}_i^T)^{-1} \mathbf{E}_i \mathbf{x}_j^T}{\mathbf{1}_{1 \times M} (\mathbf{E}_i \mathbf{E}_i^T)^{-1} \mathbf{1}_{M \times 1}} \mathbf{1}_{M \times 1} \right], 0 \right), \quad (\text{A.2})$$

$$u_{ij} = \frac{\left[(\mathbf{x}_j - \mathbf{c}_i)(\mathbf{x}_j - \mathbf{c}_i)^T + \alpha (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)(\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T \right]^{\frac{1}{1-m}}}{\sum_{q=1}^C \left[(\mathbf{x}_j - \mathbf{c}_q)(\mathbf{x}_j - \mathbf{c}_q)^T + \alpha (\mathbf{x}_j - \mathbf{p}_{qj} \mathbf{E}_q)(\mathbf{x}_j - \mathbf{p}_{qj} \mathbf{E}_q)^T \right]^{\frac{1}{1-m}}}, \quad (\text{A.3})$$

and

$$\mathbf{c}_i = \frac{\sum_{j=1}^N u_{ij}^m \mathbf{x}_j}{\sum_{j=1}^N u_{ij}^m}. \quad (\text{A.4})$$

Proof. To obtain the optimal endmember set $\mathbf{E}_i$, we set the derivative of the Lagrangian L in (3.8) with respect to $\mathbf{E}_i$ to zero, i.e.,

$$\frac{\partial L}{\partial \mathbf{E}_i} = 0, \quad (\text{A.5})$$

which leads to

$$-2\alpha \left[\sum_{j=1}^N u_{ij}^m \mathbf{p}_{ij}^T (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i) - \beta_i (M \mathbf{I}_{M \times M} - \mathbf{1}_{M \times M}) \mathbf{E}_i \right] = 0. \quad (\text{A.6})$$

Solving (A.6) for $\mathbf{E}_i$, we obtain

$$\mathbf{E}_i = \left[\beta_i (M \mathbf{I}_{M \times M} - \mathbf{1}_{M \times M}) + \sum_{j=1}^N u_{ij}^m \mathbf{p}_{ij}^T \mathbf{p}_{ij} \right]^{-1} \left[\sum_{j=1}^N u_{ij}^m \mathbf{p}_{ij}^T \mathbf{x}_j \right]. \quad (\text{A.7})$$

We check the second-order sufficient condition by computing the second-order derivative of L in (3.8) with respect to $\mathbf{E}_i$:

$$\frac{\partial^2 L}{\partial \mathbf{E}_i^2} = 2\alpha \left[\sum_{j=1}^N u_{ij}^m \mathbf{p}_{ij}^T \mathbf{p}_{ij} + \beta_i (M \mathbf{I}_{M \times M} - \mathbf{1}_{M \times M}) \right]. \quad (\text{A.8})$$

Since $\frac{\partial^2 L}{\partial \mathbf{E}_i^2}$ is an $M \times M$ matrix, we need to verify whether it is positive definite to make sure that the solution endmember set $\mathbf{E}_i$ is indeed a local minimizer of J . Let $\mathbf{e}$ be a non zero $M \times 1$ vector, we check if $\mathbf{e}^T \frac{\partial^2 L}{\partial \mathbf{E}_i^2} \mathbf{e} > 0$:

$$\begin{aligned}
\mathbf{e}^T \frac{\partial^2 L}{\partial \mathbf{E}_i^2} \mathbf{e} &= 2\alpha \mathbf{e}^T \left[\sum_{j=1}^N u_{ij}^m \mathbf{p}_{ij}^T \mathbf{p}_{ij} + \beta_i (M \mathbf{I}_{M \times M} - \mathbf{1}_{M \times M}) \right] \mathbf{e} \\
&= 2\alpha \left[\sum_{j=1}^N u_{ij}^m \mathbf{e}^T \mathbf{p}_{ij}^T \mathbf{p}_{ij} \mathbf{e} + \beta_i (M \mathbf{e}^T \mathbf{I}_{M \times M} \mathbf{e} - \mathbf{e}^T \mathbf{1}_{M \times M} \mathbf{e}) \right] \\
&= 2\alpha \left[\sum_{j=1}^N u_{ij}^m (\mathbf{p}_{ij} \mathbf{e})^T (\mathbf{p}_{ij} \mathbf{e}) + \beta_i [M \mathbf{e}^T \mathbf{e} - (\mathbf{1}_{1 \times M} \mathbf{e})^T (\mathbf{1}_{1 \times M} \mathbf{e})] \right] \\
&= 2\alpha \left[\sum_{j=1}^N u_{ij}^m \|\mathbf{p}_{ij} \mathbf{e}\|^2 + \beta_i [M \|\mathbf{e}\|^2 - \|\mathbf{1}_{1 \times M} \mathbf{e}\|^2] \right], \tag{A.9}
\end{aligned}$$

where $\|\cdot\|$ denotes the Euclidean norm.

We know that

$$\begin{aligned}
\|\mathbf{1}_{1 \times M} \mathbf{e}\|^2 &\leq \|\mathbf{1}_{1 \times M}\|^2 \|\mathbf{e}\|^2 \\
&= M \|\mathbf{e}\|^2. \tag{A.10}
\end{aligned}$$

Hence,

$$\mathbf{e}^T \frac{\partial^2 L}{\partial \mathbf{E}_i^2} \mathbf{e} > 0. \tag{A.11}$$

Therefore, the solution endmember set $\mathbf{E}_i$ is indeed a local minimizer of the objective function J .

To obtain the optimal proportions $\mathbf{p}_{ij}$, we set the derivative of L in (3.8) with respect to $\mathbf{p}_{ij}$ to zero, i.e.,

$$\frac{\partial L}{\partial \mathbf{p}_{ij}} = 0, \tag{A.12}$$

which leads to

$$-2\alpha(u_{ij}^m \mathbf{E}_i \mathbf{x}_j^T - u_{ij}^m \mathbf{E}_i \mathbf{E}_i^T \mathbf{p}_{ij}^T) - \gamma_{ij} \mathbf{1}_{M \times 1} - \xi_{ij} = 0. \tag{A.13}$$

Solving (A.13) for $\mathbf{p}_{ij}$, we obtain

$$\mathbf{p}_{ij}^T = [\mathbf{E}_i \mathbf{E}_i^T]^{-1} \left[\mathbf{E}_i \mathbf{x}_j^T + \frac{1}{2\alpha u_{ij}^m} (\gamma_{ij} \mathbf{1}_{M \times 1} + \xi_{ij}) \right]. \tag{A.14}$$

The Karush-Kuhn-Tucker (KKT) conditions [83], namely the dual feasibility and the complementary slackness, state that ξ_{ij} should be equal to zero. Finally, we enforce the non-negativity constraint on the proportion values and get

$$\mathbf{p}_{ij}^T = \max \left([\mathbf{E}_i \mathbf{E}_i^T]^{-1} \left[\mathbf{E}_i \mathbf{x}_j^T + \frac{\gamma_{ij}}{2\alpha u_{ij}^m} \mathbf{1}_{M \times 1} \right], 0 \right). \tag{A.15}$$

If the proportion values are clipped at zero, we do renormalize them to ensure that they sum to one.

Using the fact that $\mathbf{1}_{1 \times M} \mathbf{p}_{ij}^T = 1$, we solve for γ_{ij} and obtain

$$\gamma_{ij} = \frac{2\alpha u_{ij}^m [1 - \mathbf{1}_{1 \times M} (\mathbf{E}_i \mathbf{E}_i^T)^{-1} \mathbf{E}_i \mathbf{x}_j^T]}{\mathbf{1}_{1 \times M} (\mathbf{E}_i \mathbf{E}_i^T)^{-1} \mathbf{1}_{M \times 1}}. \quad (\text{A.16})$$

This leads to

$$\mathbf{p}_{ij}^T = \max \left(\left[\mathbf{E}_i \mathbf{E}_i^T \right]^{-1} \left[\mathbf{E}_i \mathbf{x}_j^T + \frac{1 - \mathbf{1}_{1 \times M} (\mathbf{E}_i \mathbf{E}_i^T)^{-1} \mathbf{E}_i \mathbf{x}_j^T}{\mathbf{1}_{1 \times M} (\mathbf{E}_i \mathbf{E}_i^T)^{-1} \mathbf{1}_{M \times 1}} \mathbf{1}_{M \times 1} \right], 0 \right). \quad (\text{A.17})$$

Similarly, we check the second-order sufficient condition by computing the second-order derivative of L in (3.8) with respect to $\mathbf{p}_{ij}$:

$$\frac{\partial^2 L}{\partial \mathbf{p}_{ij}^2} = 2\alpha u_{ij}^m \mathbf{E}_i \mathbf{E}_i^T. \quad (\text{A.18})$$

Since $\frac{\partial^2 L}{\partial \mathbf{p}_{ij}^2}$ is an $M \times M$ matrix, we can verify whether it is positive definite, to make sure that the solution proportions $\mathbf{p}_{ij}$ are indeed a local minimizer of J , by verifying that $\mathbf{p}^T \frac{\partial^2 L}{\partial \mathbf{p}_{ij}^2} \mathbf{p} > 0$ for any non zero $M \times 1$ vector $\mathbf{p}$:

$$\begin{aligned} \mathbf{p}^T \frac{\partial^2 L}{\partial \mathbf{p}_{ij}^2} \mathbf{p} &= 2\alpha u_{ij}^m \mathbf{p}^T \mathbf{E}_i \mathbf{E}_i^T \mathbf{p} \\ &= 2\alpha u_{ij}^m (\mathbf{E}_i^T \mathbf{p})^T (\mathbf{E}_i^T \mathbf{p}) \\ &= 2\alpha u_{ij}^m \|\mathbf{E}_i^T \mathbf{p}\|^2 \\ &> 0. \end{aligned} \quad (\text{A.19})$$

Hence, the solution proportions $\mathbf{p}_{ij}$ are indeed a local minimizer of the objective function J .

To obtain the optimal memberships u_{ij} , we set the derivative of L in (3.8) with respect to u_{ij} to zero, i.e.,

$$\frac{\partial L}{\partial u_{ij}} = 0, \quad (\text{A.20})$$

and obtain:

$$m u_{ij}^{m-1} (\mathbf{x}_j - \mathbf{c}_i)(\mathbf{x}_j - \mathbf{c}_i)^T + \alpha m u_{ij}^{m-1} (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)(\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T - \lambda_j = 0. \quad (\text{A.21})$$

Solving (A.21) for u_{ij} , we obtain

$$u_{ij} = \left[\frac{\lambda_j}{m[(\mathbf{x}_j - \mathbf{c}_i)(\mathbf{x}_j - \mathbf{c}_i)^T + \alpha(\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)(\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T]} \right]^{\frac{1}{m-1}}. \quad (\text{A.22})$$

Using the fact that $\sum_{i=1}^C u_{ij} = 1$, we solve for λ_j , and obtain

$$\lambda_j = \left[\sum_{i=1}^C \left[\frac{1}{m[(\mathbf{x}_j - \mathbf{c}_i)(\mathbf{x}_j - \mathbf{c}_i)^T + \alpha(\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)(\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T]} \right]^{\frac{1}{m-1}} \right]^{1-m}. \quad (\text{A.23})$$

This leads to

$$u_{ij} = \frac{[(\mathbf{x}_j - \mathbf{c}_i)(\mathbf{x}_j - \mathbf{c}_i)^T + \alpha(\mathbf{x}_j - \mathbf{p}_{ij}\mathbf{E}_i)(\mathbf{x}_j - \mathbf{p}_{ij}\mathbf{E}_i)^T]^{\frac{1}{1-m}}}{\sum_{q=1}^C [(\mathbf{x}_j - \mathbf{c}_q)(\mathbf{x}_j - \mathbf{c}_q)^T + \alpha(\mathbf{x}_j - \mathbf{p}_{qj}\mathbf{E}_q)(\mathbf{x}_j - \mathbf{p}_{qj}\mathbf{E}_q)^T]^{\frac{1}{1-m}}}. \quad (\text{A.24})$$

Similarly, we check the second-order sufficient condition by computing the second-order derivative of L in (3.8) with respect to u_{ij} :

$$\begin{aligned} \frac{\partial^2 L}{\partial u_{ij}^2} &= m(m-1)u_{ij}^{m-2}(\mathbf{x}_j - \mathbf{c}_i)(\mathbf{x}_j - \mathbf{c}_i)^T + \alpha m(m-1)u_{ij}^{m-2}(\mathbf{x}_j - \mathbf{p}_{ij}\mathbf{E}_i)(\mathbf{x}_j - \mathbf{p}_{ij}\mathbf{E}_i)^T \\ &= m(m-1)u_{ij}^{m-2}\|\mathbf{x}_j - \mathbf{c}_i\|^2 + \alpha m(m-1)u_{ij}^{m-2}\|\mathbf{x}_j - \mathbf{p}_{ij}\mathbf{E}_i\|^2. \end{aligned} \quad (\text{A.25})$$

Since $m > 1$, $\frac{\partial^2 L}{\partial u_{ij}^2}$ is a positive scalar. Therefore, the solution memberships u_{ij} are indeed a local minimizer of the objective function J .

To obtain the optimal centers $\mathbf{c}_i$, we set the derivative of L in (3.8) with respect to $\mathbf{c}_i$ to zero, i.e.,

$$\frac{\partial L}{\partial \mathbf{c}_i} = 0, \quad (\text{A.26})$$

which leads to

$$-2 \sum_{j=1}^N u_{ij}^m (\mathbf{x}_j - \mathbf{c}_i) = 0. \quad (\text{A.27})$$

Solving (A.27) for $\mathbf{c}_i$, we obtain

$$\mathbf{c}_i = \frac{\sum_{j=1}^N u_{ij}^m \mathbf{x}_j}{\sum_{j=1}^N u_{ij}^m}. \quad (\text{A.28})$$

Similarly, we check the second-order sufficient condition by computing the second-order derivative of L in (3.8) with respect to $\mathbf{c}_i$:

$$\frac{\partial^2 L}{\partial \mathbf{c}_i^2} = 2 \sum_{j=1}^N u_{ij}^m. \quad (\text{A.29})$$

Since $u_{ij} \geq 0$ for all i, j , $\frac{\partial^2 L}{\partial \mathbf{c}_i^2}$ is a positive scalar. Therefore, the solution centers $\mathbf{c}_i$ are indeed a local minimizer of the objective function J . $\square$

APPENDIX B

Proof of Theorem 3.3.1

Theorem. *Optimizing the objective function in (3.13) using the Lagrange multipliers method leads to the same update equations for the endmember sets, the proportions, and the centers as for CDSU using the Euclidean distance (equations (3.9), (3.10) and (3.12) respectively).*

The update equation for the memberships, u_{ij} , becomes

$$u_{ij} = \frac{\left[(\mathbf{x}_j - \mathbf{c}_i) \mathbf{A}_i (\mathbf{x}_j - \mathbf{c}_i)^T + \alpha (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i) (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T \right]^{\frac{1}{1-m}}}{\sum_{q=1}^C \left[(\mathbf{x}_j - \mathbf{c}_q) \mathbf{A}_q (\mathbf{x}_j - \mathbf{c}_q)^T + \alpha (\mathbf{x}_j - \mathbf{p}_{qj} \mathbf{E}_q) (\mathbf{x}_j - \mathbf{p}_{qj} \mathbf{E}_q)^T \right]^{\frac{1}{1-m}}}. \quad (\text{B.1})$$

Finally, the update equation for the norm matrices, $\mathbf{A}_i$, is

$$\mathbf{A}_i = [\sigma_i \det(\mathbf{C}_i)]^{\frac{1}{d}} \mathbf{C}_i^{-1}, \quad (\text{B.2})$$

where

$$\mathbf{C}_i = \frac{\sum_{j=1}^N u_{ij}^m (\mathbf{x}_j - \mathbf{c}_i)^T (\mathbf{x}_j - \mathbf{c}_i)}{\sum_{j=1}^N u_{ij}^m} \quad (\text{B.3})$$

is the fuzzy covariance matrix of cluster i .

Proof. We incorporate the constraints in (3.4), (3.5), and (3.14) into the objective function in (3.13) using Lagrange multipliers and obtain

$$\begin{aligned} L = & \sum_{i=1}^C \sum_{j=1}^N u_{ij}^m (\mathbf{x}_j - \mathbf{c}_i) \mathbf{A}_i (\mathbf{x}_j - \mathbf{c}_i)^T \\ & + \alpha \sum_{i=1}^C \left[\sum_{j=1}^N u_{ij}^m (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i) (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T + \beta_i (M \text{trace}(\mathbf{E}_i \mathbf{E}_i^T) - \mathbf{1}_{1 \times M} \mathbf{E}_i \mathbf{E}_i^T \mathbf{1}_{M \times 1}) \right] \\ & - \sum_{j=1}^N \lambda_j \left(\sum_{i=1}^C u_{ij} - 1 \right) - \sum_{i=1}^C \sum_{j=1}^N \gamma_{ij} (\mathbf{1}_{1 \times M} \mathbf{p}_{ij}^T - 1) - \sum_{i=1}^C \sum_{j=1}^N \xi_{ij} \mathbf{p}_{ij}^T - \sum_{i=1}^C \delta_i (\det(\mathbf{A}_i) - \sigma_i), \end{aligned} \quad (\text{B.4})$$

where $\boldsymbol{\Lambda} = [\lambda_1, \dots, \lambda_N]$ is a vector of Lagrange multipliers corresponding to the N constraints on the memberships u_{ij} in (3.4), $\boldsymbol{\Gamma} = [\gamma_{11}, \dots, \gamma_{CN}]$ and $\boldsymbol{\Xi} = [\xi_{11}, \dots, \xi_{CN}]$ are vectors of Lagrange multipliers corresponding to the $C \times N$ constraints on the proportions $\mathbf{p}_{ij}$ in (3.5), and $\boldsymbol{\Delta} = [\delta_1, \dots, \delta_C]$ is a vector of Lagrange multipliers corresponding to the C constraints on the norm matrices $\mathbf{A}_i$ in (3.14).

The proof of the first part of the theorem, concerning the endmember sets, the proportion sets and the centers, is similar to the one in Appendix A.

To obtain the optimal memberships u_{ij} , we set the derivative of L in (B.4) with respect to u_{ij} to zero, i.e.,

$$\frac{\partial L}{\partial u_{ij}} = 0, \quad (\text{B.5})$$

and obtain:

$$m u_{ij}^{m-1} (\mathbf{x}_j - \mathbf{c}_i) \mathbf{A}_i (\mathbf{x}_j - \mathbf{c}_i)^T + \alpha m u_{ij}^{m-1} (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i) (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T - \lambda_j = 0. \quad (\text{B.6})$$

Solving (B.6) for u_{ij} , we obtain

$$u_{ij} = \left[\frac{\lambda_j}{m [(\mathbf{x}_j - \mathbf{c}_i) \mathbf{A}_i (\mathbf{x}_j - \mathbf{c}_i)^T + \alpha (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i) (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T]} \right]^{\frac{1}{m-1}}. \quad (\text{B.7})$$

Using the fact that $\sum_{i=1}^C u_{ij} = 1$, we solve for λ_j , and obtain

$$\lambda_j = \left[\sum_{i=1}^C \left[\frac{1}{m [(\mathbf{x}_j - \mathbf{c}_i) \mathbf{A}_i (\mathbf{x}_j - \mathbf{c}_i)^T + \alpha (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i) (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T]} \right]^{\frac{1}{m-1}} \right]^{1-m}. \quad (\text{B.8})$$

This leads to

$$u_{ij} = \frac{[(\mathbf{x}_j - \mathbf{c}_i) \mathbf{A}_i (\mathbf{x}_j - \mathbf{c}_i)^T + \alpha (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i) (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T]^{\frac{1}{1-m}}}{\sum_{q=1}^C [(\mathbf{x}_j - \mathbf{c}_q) \mathbf{A}_q (\mathbf{x}_j - \mathbf{c}_q)^T + \alpha (\mathbf{x}_j - \mathbf{p}_{jq} \mathbf{E}_q) (\mathbf{x}_j - \mathbf{p}_{jq} \mathbf{E}_q)^T]^{\frac{1}{1-m}}}. \quad (\text{B.9})$$

We check the second-order sufficient condition by computing the second-order derivative of L in (B.4) with respect to u_{ij} :

$$\frac{\partial^2 L}{\partial u_{ij}^2} = m(m-1) u_{ij}^{m-2} (\mathbf{x}_j - \mathbf{c}_i) \mathbf{A}_i (\mathbf{x}_j - \mathbf{c}_i)^T + \alpha m(m-1) u_{ij}^{m-2} (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i) (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T. \quad (\text{B.10})$$

Since $m > 1$, $\frac{\partial^2 L}{\partial u_{ij}^2}$ is a positive scalar. Therefore, the solution memberships u_{ij} are indeed a local minimizer of the objective function J_M .

To obtain the optimal norm matrices, $\mathbf{A}_i$, we set the derivative of L in (B.4) with respect to $\mathbf{A}_i$ to zero, i.e.,

$$\frac{\partial L}{\partial \mathbf{A}_i} = 0, \quad (\text{B.11})$$

and obtain:

$$\sum_{j=1}^N u_{ij}^m (\mathbf{x}_j - \mathbf{c}_i)^T (\mathbf{x}_j - \mathbf{c}_i) - \delta_i \det(\mathbf{A}_i) \mathbf{A}_i^{-1} = 0. \quad (\text{B.12})$$

Solving (B.12) for $\mathbf{A}_i$, we obtain

$$\mathbf{A}_i = \left[\frac{\sum_{j=1}^N u_{ij}^m (\mathbf{x}_j - \mathbf{c}_i)^T (\mathbf{x}_j - \mathbf{c}_i)}{\delta_i \det(\mathbf{A}_i)} \right]^{-1}. \quad (\text{B.13})$$

We define the fuzzy covariance matrix

$$\mathbf{C}_i = \frac{\sum_{j=1}^N u_{ij}^m (\mathbf{x}_j - \mathbf{c}_i)^T (\mathbf{x}_j - \mathbf{c}_i)}{\sum_{j=1}^N u_{ij}^m}. \quad (\text{B.14})$$

Using (B.14) gives

$$\mathbf{A}_i = \frac{\delta_i \det(\mathbf{A}_i)}{\sum_{j=1}^N u_{ij}^m} \mathbf{C}_i^{-1}. \quad (\text{B.15})$$

Using the fact that $\det(\mathbf{A}_i) = \sigma_i$, we solve for δ_i , and obtain

$$\delta_i = \frac{\det(\mathbf{C}_i)^{\frac{1}{d}} \sum_{j=1}^N u_{ij}^m}{\sigma_i^{1-\frac{1}{d}}}. \quad (\text{B.16})$$

This leads to

$$\mathbf{A}_i = [\sigma_i \det(\mathbf{C}_i)]^{\frac{1}{d}} \mathbf{C}_i^{-1}. \quad (\text{B.17})$$

Similarly, we check the second-order sufficient condition by computing the second-order derivative of L in (B.4) with respect to $\mathbf{A}_i$:

$$\begin{aligned} \frac{\partial^2 L}{\partial \mathbf{A}_i^2} &= -\delta_i [\det(\mathbf{A}_i) \mathbf{A}_i^{-2} - \det(\mathbf{A}_i) \mathbf{A}_i^{-2}] \\ &= 0. \end{aligned} \quad (\text{B.18})$$

Therefore, the test is inconclusive. We, hence, move to a higher derivative test. We compute the third order derivative of L in (B.4) with respect to $\mathbf{A}_i$:

$$\frac{\partial^3 L}{\partial \mathbf{A}_i^3} = -3\delta_i \det(\mathbf{A}_i) \mathbf{A}_i^{-3}. \quad (\text{B.19})$$

According to (B.16), $\delta_i > 0$. Hence $\frac{\partial^3 L}{\partial \mathbf{A}_i^3} < 0$. An odd derivative order of negative sign indicates that the critical solution $\mathbf{A}_i$ is neither local maximum nor local minimum. It is instead a point of decrease for the function J_M . In fact, it is also a point of inflection, though that is not of relevance here. $\square$

APPENDIX C

Proof of Theorem 3.4.1

Theorem. *The update equations for the endmember sets, the abundances, the centers and the norm matrices are similar to the ones of CDSU_M (equations (3.9), (3.10), (3.12) and (3.16) respectively). The update equation for the memberships becomes*

$$u_{ij} = \frac{\left[d_{ij}^2 + \gamma \text{cost}_{ij} + \alpha \text{fit}_{ij} \right]^{\frac{1}{1-m}}}{\sum_{q=1}^C \left[d_{qj}^2 + \gamma \text{cost}_{qj} + \alpha \text{fit}_{qj} \right]^{\frac{1}{1-m}}}, \quad (\text{C.1})$$

where

$$d_{ij}^2 = (\mathbf{x}_j - \mathbf{c}_i) \mathbf{A}_i (\mathbf{x}_j - \mathbf{c}_i)^T, \quad (\text{C.2})$$

$$\text{cost}_{ij} = \sum_{(j,k) \in \mathcal{S}} \sum_{l=1, l \neq i}^C \rho_{jk} u_{lk}^m + \sum_{(j,k) \in \mathcal{N}} \rho_{jk} u_{ik}^m, \quad (\text{C.3})$$

and

$$\text{fit}_{ij} = (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i) (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T. \quad (\text{C.4})$$

Proof. We incorporate the constraints in (3.4), (3.5), and (3.14) into the objective function in (3.18) using Lagrange multipliers and obtain

$$\begin{aligned} L = & \sum_{i=1}^C \sum_{j=1}^N u_{ij}^m (\mathbf{x}_j - \mathbf{c}_i) \mathbf{A}_i (\mathbf{x}_j - \mathbf{c}_i)^T \\ & + \gamma \left(\sum_{(j,k) \in \mathcal{S}} \sum_{i=1}^C \sum_{l=1, l \neq i}^C \rho_{jk} u_{ij}^m u_{lk}^m + \sum_{(j,k) \in \mathcal{N}} \sum_{i=1}^C \rho_{jk} u_{ij}^m u_{ik}^m \right) \\ & + \alpha \sum_{i=1}^C \left[\sum_{j=1}^N u_{ij}^m (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i) (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T + \beta_i (M\text{trace}(\mathbf{E}_i \mathbf{E}_i^T) - \mathbf{1}_{1 \times M} \mathbf{E}_i \mathbf{E}_i^T \mathbf{1}_{M \times 1}) \right] \\ & - \sum_{j=1}^N \lambda_j \left(\sum_{i=1}^C u_{ij} - 1 \right) - \sum_{i=1}^C \sum_{j=1}^N \varphi_{ij} (\mathbf{1}_{1 \times M} \mathbf{p}_{ij}^T - 1) - \sum_{i=1}^C \sum_{j=1}^N \xi_{ij} \mathbf{p}_{ij}^T - \sum_{i=1}^C \delta_i (\det(\mathbf{A}_i) - \sigma_i), \quad (\text{C.5}) \end{aligned}$$

where $\mathbf{\Lambda} = [\lambda_1, \dots, \lambda_N]$ is a vector of Lagrange multipliers corresponding to the N constraints on the memberships u_{ij} in (3.4), $\mathbf{\Phi} = [\varphi_{11}, \dots, \varphi_{CN}]$ and $\mathbf{\Xi} = [\xi_{11}, \dots, \xi_{CN}]$ are vectors of Lagrange multipliers corresponding to the $C \times N$ constraints on the proportions $\mathbf{p}_{ij}$ in (3.5), and $\mathbf{\Delta} = [\delta_1, \dots, \delta_C]$ is a vector of Lagrange multipliers corresponding to the C constraints on the norm matrices $\mathbf{A}_i$ in (3.14).

The proof of the first part of the theorem, concerning the endmember sets, the proportion sets and the centers, is similar to the one in Appendix A. The proof concerning the norm matrices is similar to the one in Appendix B.

To obtain the optimal memberships u_{ij} , we set the derivative of L in (C.5) with respect to u_{ij} to zero, i.e.,

$$\frac{\partial L}{\partial u_{ij}} = 0, \quad (\text{C.6})$$

and obtain:

$$\begin{aligned} m u_{ij}^{m-1} (\mathbf{x}_j - \mathbf{c}_i) \mathbf{A}_i (\mathbf{x}_j - \mathbf{c}_i)^T + \gamma \left(\sum_{(j,k) \in \mathcal{S}} \sum_{l=1, l \neq i}^C m \rho_{jk} u_{ij}^{m-1} u_{lk}^m + \sum_{(j,k) \in \mathcal{N}} m \rho_{jk} u_{ij}^{m-1} u_{ik}^m \right) \\ + \alpha m u_{ij}^{m-1} (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i) (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T - \lambda_j = 0. \end{aligned} \quad (\text{C.7})$$

We rewrite (C.7) as

$$m u_{ij}^{m-1} d_{ij}^2 + \gamma m u_{ij}^{m-1} \text{cost}_{ij} + \alpha m u_{ij}^{m-1} \text{fit}_{ij} - \lambda_j = 0, \quad (\text{C.8})$$

where

$$d_{ij}^2 = (\mathbf{x}_j - \mathbf{c}_i) \mathbf{A}_i (\mathbf{x}_j - \mathbf{c}_i)^T, \quad (\text{C.9})$$

$$\text{cost}_{ij} = \sum_{(j,k) \in \mathcal{S}} \sum_{l=1, l \neq i}^C \rho_{jk} u_{lk}^m + \sum_{(j,k) \in \mathcal{N}} \rho_{jk} u_{ik}^m, \quad (\text{C.10})$$

and

$$\text{fit}_{ij} = (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i) (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T. \quad (\text{C.11})$$

Solving (C.8) for u_{ij} , we obtain

$$u_{ij} = \left[\frac{\lambda_j}{m[d_{ij}^2 + \gamma \text{cost}_{ij} + \alpha \text{fit}_{ij}]} \right]^{\frac{1}{m-1}}. \quad (\text{C.12})$$

Using the fact that $\sum_{i=1}^C u_{ij} = 1$, we solve for λ_j , and obtain

$$\lambda_j = \left[\sum_{i=1}^C \left[\frac{1}{m[d_{ij}^2 + \gamma \text{cost}_{ij} + \alpha \text{fit}_{ij}]} \right]^{\frac{1}{m-1}} \right]^{1-m}. \quad (\text{C.13})$$

This leads to

$$u_{ij} = \frac{\left[d_{ij}^2 + \gamma \text{cost}_{ij} + \alpha \text{fit}_{ij} \right]^{\frac{1}{1-m}}}{\sum_{q=1}^C \left[d_{qj}^2 + \gamma \text{cost}_{qj} + \alpha \text{fit}_{qj} \right]^{\frac{1}{1-m}}}, \quad (\text{C.14})$$

We check the second-order sufficient condition by computing the second-order derivative of L in (C.5) with respect to u_{ij} :

$$m(m-1)u_{ij}^{m-2}d_{ij}^2 + \gamma m(m-1)u_{ij}^{m-2}\text{cost}_{ij} + \alpha m(m-1)u_{ij}^{m-2}\text{fit}_{ij}. \quad (\text{C.15})$$

Since $m > 1$, $\frac{\partial^2 L}{\partial u_{ij}^2}$ is a positive scalar. Therefore, the solution memberships u_{ij} are indeed a local minimizer of the objective function J_C . $\square$

APPENDIX D

Proof of Theorem 3.4.2

Theorem. *The update equations for the endmember sets, the memberships, the centers and the norm matrices are similar to the ones of CDSU_M (equations (3.9), (3.15), (3.12) and (3.16) respectively).*

The update equation for the proportions becomes

$$\mathbf{p}_{ij}^T = \max \left(\left[2\alpha u_{ij}^m \mathbf{E}_i \mathbf{E}_i^T + \gamma \sum_{(j,k) \in \mathcal{S}} \rho_{jk} \right]^{-1} \left[2\alpha u_{ij}^m \mathbf{E}_i \mathbf{x}_j^T + \gamma \sum_{(j,k) \in \mathcal{S}} \rho_{jk} \mathbf{p}_{ik}^T + \xi_{ij} \mathbf{1}_{M \times 1} \right], 0 \right), \quad (\text{D.1})$$

where

$$\xi_{ij} = \frac{1 - \mathbf{1}_{1 \times M} \left[2\alpha u_{ij}^m \mathbf{E}_i \mathbf{E}_i^T + \gamma \sum_{(j,k) \in \mathcal{S}} \rho_{jk} \right]^{-1} \left[2\alpha u_{ij}^m \mathbf{E}_i \mathbf{x}_j^T + \gamma \sum_{(j,k) \in \mathcal{S}} \rho_{jk} \mathbf{p}_{ik}^T \right]}{\mathbf{1}_{1 \times M} \left[2\alpha u_{ij}^m \mathbf{E}_i \mathbf{E}_i^T + \gamma \sum_{(j,k) \in \mathcal{S}} \rho_{jk} \right]^{-1} \mathbf{1}_{M \times 1}}. \quad (\text{D.2})$$

Proof. We incorporate the constraints in (3.4), (3.5), and (3.14) into the objective function in (3.23) using Lagrange multipliers and obtain

$$\begin{aligned} L = & \sum_{i=1}^C \sum_{j=1}^N u_{ij}^m (\mathbf{x}_j - \mathbf{c}_i) \mathbf{A}_i (\mathbf{x}_j - \mathbf{c}_i)^T + \gamma \sum_{(j,k) \in \mathcal{S}} \rho_{jk} \sum_{i=1}^C \|\mathbf{p}_{ij} - \mathbf{p}_{ik}\|^2 \\ & + \alpha \sum_{i=1}^C \left[\sum_{j=1}^N u_{ij}^m (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i) (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T + \beta_i (M \text{trace}(\mathbf{E}_i \mathbf{E}_i^T) - \mathbf{1}_{1 \times M} \mathbf{E}_i \mathbf{E}_i^T \mathbf{1}_{M \times 1}) \right] \\ & - \sum_{j=1}^N \lambda_j \left(\sum_{i=1}^C u_{ij} - 1 \right) - \sum_{i=1}^C \sum_{j=1}^N \xi_{ij} (\mathbf{1}_{1 \times M} \mathbf{p}_{ij}^T - 1) - \sum_{i=1}^C \sum_{j=1}^N \varphi_{ij} \mathbf{p}_{ij}^T - \sum_{i=1}^C \delta_i (\det(\mathbf{A}_i) - \sigma_i), \quad (\text{D.3}) \end{aligned}$$

where $\mathbf{\Lambda} = [\lambda_1, \dots, \lambda_N]$ is a vector of Lagrange multipliers corresponding to the N constraints on the memberships u_{ij} in (3.4), $\mathbf{\Xi} = [\xi_{11}, \dots, \xi_{CN}]$ and $\mathbf{\Phi} = [\varphi_{11}, \dots, \varphi_{CN}]$ are vectors of Lagrange multipliers corresponding to the $C \times N$ constraints on the proportions $\mathbf{p}_{ij}$ in (3.5), and $\mathbf{\Delta} = [\delta_1, \dots, \delta_C]$ is a vector of Lagrange multipliers corresponding to the C constraints on the norm matrices $\mathbf{A}_i$ in (3.14).

The proof of the first part of the theorem, concerning the endmember sets, the memberships, the centers, and the norm matrices is similar to the one in Appendix B.

To obtain the optimal proportions $\mathbf{p}_{ij}$, we set the derivative of L in (D.3) with respect to $\mathbf{p}_{ij}$ to zero, i.e.,

$$\frac{\partial L}{\partial \mathbf{p}_{ij}} = 0, \quad (\text{D.4})$$

which leads to

$$-2\alpha(u_{ij}^m \mathbf{E}_i \mathbf{x}_j^T - u_{ij}^m \mathbf{E}_i \mathbf{E}_i^T \mathbf{p}_{ij}^T) + \gamma \sum_{(j,k) \in \mathcal{S}} \rho_{jk} (\mathbf{p}_{ij} - \mathbf{p}_{ik})^T - \xi_{ij} \mathbf{1}_{M \times 1} - \varphi_{ij} = 0. \quad (\text{D.5})$$

Solving (D.5) for $\mathbf{p}_{ij}$, we obtain

$$\mathbf{p}_{ij}^T = \left[2\alpha u_{ij}^m \mathbf{E}_i \mathbf{E}_i^T + \gamma \sum_{(j,k) \in \mathcal{S}} \rho_{jk} \right]^{-1} \left[2\alpha u_{ij}^m \mathbf{E}_i \mathbf{x}_j^T + \gamma \sum_{(j,k) \in \mathcal{S}} \rho_{jk} \mathbf{p}_{ik}^T + \xi_{ij} \mathbf{1}_{M \times 1} + \varphi_{ij} \right], \quad (\text{D.6})$$

The Karush-Kuhn-Tucker (KKT) conditions [83], namely the dual feasibility and the complementary slackness, state that φ_{ij} should be equal to zero. Finally, we enforce the non-negativity constraint on the proportion values and get

$$\mathbf{p}_{ij}^T = \max \left(\left[2\alpha u_{ij}^m \mathbf{E}_i \mathbf{E}_i^T + \gamma \sum_{(j,k) \in \mathcal{S}} \rho_{jk} \right]^{-1} \left[2\alpha u_{ij}^m \mathbf{E}_i \mathbf{x}_j^T + \gamma \sum_{(j,k) \in \mathcal{S}} \rho_{jk} \mathbf{p}_{ik}^T + \xi_{ij} \mathbf{1}_{M \times 1} \right], 0 \right), \quad (\text{D.7})$$

If the proportion values are clipped at zero, we do renormalize them to ensure that they sum to one.

Using the fact that $\mathbf{1}_{1 \times M} \mathbf{p}_{ij}^T = 1$, we solve for ξ_{ij} and obtain

$$\xi_{ij} = \frac{1 - \mathbf{1}_{1 \times M} \left[2\alpha u_{ij}^m \mathbf{E}_i \mathbf{E}_i^T + \gamma \sum_{(j,k) \in \mathcal{S}} \rho_{jk} \right]^{-1} \left[2\alpha u_{ij}^m \mathbf{E}_i \mathbf{x}_j^T + \gamma \sum_{(j,k) \in \mathcal{S}} \rho_{jk} \mathbf{p}_{ik}^T \right]}{\mathbf{1}_{1 \times M} \left[2\alpha u_{ij}^m \mathbf{E}_i \mathbf{E}_i^T + \gamma \sum_{(j,k) \in \mathcal{S}} \rho_{jk} \right]^{-1} \mathbf{1}_{M \times 1}}. \quad (\text{D.8})$$

We check the second-order sufficient condition by computing the second-order derivative of L in (D.3) with respect to $\mathbf{p}_{ij}$:

$$\frac{\partial^2 L}{\partial \mathbf{p}_{ij}^2} = 2\alpha u_{ij}^m \mathbf{E}_i \mathbf{E}_i^T + \gamma \sum_{(j,k) \in \mathcal{S}} \rho_{jk}. \quad (\text{D.9})$$

Since $\frac{\partial^2 L}{\partial \mathbf{p}_{ij}^2}$ is an $M \times M$ matrix, we can verify whether it is positive definite, to make sure that the solution proportions $\mathbf{p}_{ij}$ are indeed a local minimizer of J_P , by verifying that $\mathbf{p}^T \frac{\partial^2 L}{\partial \mathbf{p}_{ij}^2} \mathbf{p} > 0$ for any non zero $M \times 1$ vector $\mathbf{p}$:

$$\begin{aligned} \mathbf{p}^T \frac{\partial^2 L}{\partial \mathbf{p}_{ij}^2} \mathbf{p} &= 2\alpha u_{ij}^m \mathbf{p}^T \mathbf{E}_i \mathbf{E}_i^T \mathbf{p} + \gamma \sum_{(j,k) \in \mathcal{S}} \rho_{jk} \mathbf{p}^T \mathbf{p} \\ &= 2\alpha u_{ij}^m (\mathbf{E}_i^T \mathbf{p})^T (\mathbf{E}_i^T \mathbf{p}) + \gamma \sum_{(j,k) \in \mathcal{S}} \rho_{jk} \|\mathbf{p}\|^2 \\ &= 2\alpha u_{ij}^m \|\mathbf{E}_i^T \mathbf{p}\|^2 + \gamma \sum_{(j,k) \in \mathcal{S}} \rho_{jk} \|\mathbf{p}\|^2 \\ &> 0. \end{aligned} \quad (\text{D.10})$$

Hence, the solution proportions $\mathbf{p}_{ij}$ are indeed a local minimizer of the objective function J_P .

□

APPENDIX E

Proof of Theorem 3.5.1

Theorem. *The update equations for the proportions and the memberships are similar to the ones of CDSU_M (equations (3.10) and (3.15) respectively).*

The update equation for the endmember sets, $\mathbf{E}_i$, is

$$\mathbf{E}_i = \left[\beta_i (M\mathbf{I}_{M \times M} - \mathbf{1}_{M \times M}) + \sum_{j=1}^N (au_{ij}^m + bt_{ij}^n) \mathbf{p}_{ij}^T \mathbf{p}_{ij} \right]^{-1} \left[\sum_{j=1}^N (au_{ij}^m + bt_{ij}^n) \mathbf{p}_{ij}^T \mathbf{x}_j \right]. \quad (\text{E.1})$$

The update equation for the norm matrices, $\mathbf{A}_i$, is the same as in equation (3.16), but with

$$\mathbf{C}_i = \frac{\sum_{j=1}^N (au_{ij}^m + bt_{ij}^n) (\mathbf{x}_j - \mathbf{c}_i)^T (\mathbf{x}_j - \mathbf{c}_i)}{\sum_{j=1}^N (au_{ij}^m + bt_{ij}^n)}. \quad (\text{E.2})$$

The update equation for the typicalities, t_{ij} , is given by

$$t_{ij} = \frac{1}{1 + \left[b \frac{\text{cost}_{ij}}{\eta_i} \right]^{\frac{1}{n-1}}}, \quad (\text{E.3})$$

where

$$\text{cost}_{ij} = (\mathbf{x}_j - \mathbf{c}_i) \mathbf{A}_i (\mathbf{x}_j - \mathbf{c}_i)^T + \alpha (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i) (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T. \quad (\text{E.4})$$

Finally, the update equation of the cluster centers $\mathbf{c}_i$ is

$$\mathbf{c}_i = \frac{\sum_{j=1}^N (au_{ij}^m + bt_{ij}^n) \mathbf{x}_j}{\sum_{j=1}^N (au_{ij}^m + bt_{ij}^n)}. \quad (\text{E.5})$$

Proof. We incorporate the constraints in (3.4), (3.5), and (3.14) into the objective function in (3.26) using Lagrange multipliers and obtain

$$\begin{aligned} L = & \sum_{i=1}^C \sum_{j=1}^N (au_{ij}^m + bt_{ij}^n) (\mathbf{x}_j - \mathbf{c}_i) \mathbf{A}_i (\mathbf{x}_j - \mathbf{c}_i)^T + \sum_{i=1}^C \eta_i \sum_{j=1}^N (1 - t_{ij})^n \\ & + \alpha \sum_{i=1}^C \left[\sum_{j=1}^N (au_{ij}^m + bt_{ij}^n) (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i) (\mathbf{x}_j - \mathbf{p}_{ij} \mathbf{E}_i)^T + \beta_i (M \text{trace}(\mathbf{E}_i \mathbf{E}_i^T) - \mathbf{1}_{1 \times M} \mathbf{E}_i \mathbf{E}_i^T \mathbf{1}_{M \times 1}) \right] \\ & - \sum_{j=1}^N \lambda_j \left(\sum_{i=1}^C u_{ij} - 1 \right) - \sum_{i=1}^C \sum_{j=1}^N \gamma_{ij} (\mathbf{1}_{1 \times M} \mathbf{p}_{ij}^T - 1) - \sum_{i=1}^C \sum_{j=1}^N \xi_{ij} \mathbf{p}_{ij}^T - \sum_{i=1}^C \delta_i (\det(\mathbf{A}_i) - \sigma_i), \quad (\text{E.6}) \end{aligned}$$

where $\mathbf{\Lambda} = [\lambda_1, \dots, \lambda_N]$ is a vector of Lagrange multipliers corresponding to the N constraints on the memberships u_{ij} in (3.4), $\mathbf{\Gamma} = [\gamma_{11}, \dots, \gamma_{CN}]$ and $\mathbf{\Xi} = [\xi_{11}, \dots, \xi_{CN}]$ are vectors of Lagrange multipliers corresponding to the $C \times N$ constraints on the proportions $\mathbf{p}_{ij}$ in (3.5), and $\mathbf{\Delta} = [\delta_1, \dots, \delta_C]$ is a vector of Lagrange multipliers corresponding to the C constraints on the norm matrices $\mathbf{A}_i$ in (3.14).

The proof concerning the proportions is similar to the one in Appendix A. The proof concerning the memberships is similar to the one in Appendix B. The proof concerning the endmember sets, the centers and the norm matrices is similar to the one in Appendix A and Appendix B, with the exception of using the weighted memberships $(au_{ij}^m + bt_{ij}^n)$ instead of the fuzzy memberships u_{ij}^m .

To obtain the optimal typicalities t_{ij} , we set the derivative of L in (E.6) with respect to t_{ij} to zero, i.e.,

$$\frac{\partial L}{\partial t_{ij}} = 0, \quad (\text{E.7})$$

and obtain:

$$nbt_{ij}^{n-1}(\mathbf{x}_j - \mathbf{c}_i)\mathbf{A}_i(\mathbf{x}_j - \mathbf{c}_i)^T + \alpha n t_{ij}^{n-1}(\mathbf{x}_j - \mathbf{p}_{ij}\mathbf{E}_i)(\mathbf{x}_j - \mathbf{p}_{ij}\mathbf{E}_i)^T - n\eta_i(1 - t_{ij})^{n-1} = 0. \quad (\text{E.8})$$

Rearranging the terms of (E.8), we get

$$\left[\frac{t_{ij}}{1 - t_{ij}} \right]^{n-1} = \frac{\eta_i}{b[(\mathbf{x}_j - \mathbf{c}_i)\mathbf{A}_i(\mathbf{x}_j - \mathbf{c}_i)^T + \alpha t_{ij}^{n-1}(\mathbf{x}_j - \mathbf{p}_{ij}\mathbf{E}_i)(\mathbf{x}_j - \mathbf{p}_{ij}\mathbf{E}_i)^T]}. \quad (\text{E.9})$$

Solving (E.9) for t_{ij} , we obtain

$$t_{ij} = \frac{1}{1 + \left[b \frac{\text{cost}_{ij}}{\eta_i} \right]^{\frac{1}{n-1}}}, \quad (\text{E.10})$$

where

$$\text{cost}_{ij} = (\mathbf{x}_j - \mathbf{c}_i)\mathbf{A}_i(\mathbf{x}_j - \mathbf{c}_i)^T + \alpha(\mathbf{x}_j - \mathbf{p}_{ij}\mathbf{E}_i)(\mathbf{x}_j - \mathbf{p}_{ij}\mathbf{E}_i)^T. \quad (\text{E.11})$$

We check the second-order sufficient condition by computing the second-order derivative of L in (E.6) with respect to t_{ij} :

$$\frac{\partial^2 L}{\partial t_{ij}^2} = n(n-1)bt_{ij}^{n-2}(\mathbf{x}_j - \mathbf{c}_i)\mathbf{A}_i(\mathbf{x}_j - \mathbf{c}_i)^T + \alpha n(n-1)t_{ij}^{n-2}(\mathbf{x}_j - \mathbf{p}_{ij}\mathbf{E}_i)(\mathbf{x}_j - \mathbf{p}_{ij}\mathbf{E}_i)^T + n(n-1)\eta_i(1 - t_{ij})^{n-2}. \quad (\text{E.12})$$

Since $n > 1$, $\frac{\partial^2 L}{\partial t_{ij}^2}$ is a positive scalar. Therefore, the solution memberships t_{ij} are indeed a local minimizer of the objective function J_R . $\square$

CURRICULUM VITAE

NAME: Hamdi Jenzri

ADDRESS: Computer Engineering & Computer Science Department
Speed School of Engineering
University of Louisville
Louisville, KY 40292

EDUCATION:

Ph.D., Computer Science & Engineering

August 2014

University of Louisville, Louisville, Kentucky

B.Eng., Telecommunications Engineering

June 2006

Higher School of Communications of Tunis, Tunis, Tunisia

CONFERENCE PUBLICATIONS:

1. **H. Jenzri** and H. Frigui, “*Robust Unmixing using Consensus Analysis*”, IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Lausanne, Switzerland, June 2014.
2. **H. Jenzri**, H. Frigui, and P. Gader, “*Context Dependent Hyperspectral Subpixel Target Detection*”, International Conference on Image Processing (ICIP), Paris, France, October 2014.
3. **H. Jenzri**, H. Frigui, and P. Gader, “*Robust Context Dependent Spectral Unmixing*”, International Conference on Pattern Recognition (ICPR), Stockholm, Sweden, August 2014.
4. **H. Jenzri**, H. Frigui, and P. Gader, “*Graph Constrained Multi-Model Unmixing Using LiDAR Information*”, IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Gainesville, Florida, June 2013.

5. **H. Jenzri**, H. Frigui, and P. Gader, “*Context Dependent Spectral Unmixing*”, IEEE International Workshop on Machine Learning for Signal Processing (MLSP), Santander, Spain, September 2012.
6. **H. Jenzri** and R. Abdelfattah, “*Contribution of the Inter-channel Polarimetric Coherence for Soil Classification*”, IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Cape Town, South Africa, July 2009.

HONORS AND AWARDS:

1. IEEE Outstanding CECS Student Award, April 2014
2. Golden Key International Honour Society Member, September 2010
3. Higher School of Communications of Tunis Travel Award, June 2008
4. Tunisian National Scholarship for Engineering Studies, September 2006