

# SEMI-SUPERVISED INTERACTIVE UNMIXING FOR HYPERSPECTRAL IMAGE ANALYSIS

---

A Thesis presented to  
the Faculty of the Graduate School  
at the University of Missouri

---

In Partial Fulfillment  
of the Requirements for the Degree  
Master of Science

---

by  
SHENG ZOU  
Dr. Alina Zare, Thesis Supervisor  
December 2016

The undersigned, appointed by the Dean of the Graduate School, have examined the thesis entitled:

SEMI-SUPERVISED INTERACTIVE UNMIXING  
FOR HYPERSPECTRAL IMAGE ANALYSIS

Sheng Zou,

a candidate for the degree of Master of Science and hereby certify that, in their opinion, it is worthy of acceptance.

---

Dr. Aina Zare

---

Dr. James Keller

---

Dr. Yi Shang

## ACKNOWLEDGMENTS

I would like to thank my adviser, Dr. Zare, for her precious assistance on my graduate study and research. I would like to thank committee members, Dr. Keller and Dr. Shang, for attending my defense and offering me valuable suggestions.

Thank you to all of my labmates for daily discussions and help, including Changzhe Jiao, Chao Chen, Hao Sun and Xiaoxiao Du. I benefited a lot from these discussions.

Finally, I would like to thank my parents for their everlasting love, support, encouragement from the other side of the earth.

# TABLE OF CONTENTS

<b>ACKNOWLEDGMENTS</b> . . . . .	<b>ii</b>
<b>LIST OF TABLES</b> . . . . .	<b>v</b>
<b>LIST OF FIGURES</b> . . . . .	<b>vi</b>
<b>ABSTRACT</b> . . . . .	<b>viii</b>
<b>CHAPTER</b> . . . . .	
<b>1 Introduction</b> . . . . .	<b>1</b>
1.1 Hyperspectral Unmixing . . . . .	2
1.2 Problem Statement . . . . .	4
<b>2 Literature Review</b> . . . . .	<b>6</b>
2.1 Semi-supervised Hyperspectral Unmixing . . . . .	7
2.2 Extended Functions of Multiple Instances (eFUMI) . . . . .	8
2.3 Spectral Variability . . . . .	13
2.3.1 Endmembers as Sets . . . . .	15
2.3.2 Endmembers as Statistical Distributions . . . . .	16
2.4 Partial Membership Latent Dirichlet Allocation (PM-LDA) . . . . .	21
2.4.1 Latent Dirichlet Allocation . . . . .	23
2.4.2 Partial Membership Latent Dirichlet Allocation . . . . .	24
2.5 Open Street Map (OSM) Guided Hyperspectral SLIC . . . . .	27
<b>3 Proposed Method</b> . . . . .	<b>29</b>

3.1	Instance Influence Estimation for Hyperspectral Target Signature Characterization using eFUMI . . . . .	29
3.2	Hyperspectral Unmixing with Endmember Variability using PM-LDA	32
3.2.1	PM-LDA Relationship with Normal Compositional Model . . .	32
3.2.2	Semi-supervised PM-LDA . . . . .	33
<b>4</b>	<b>Experiments . . . . .</b>	<b>36</b>
4.1	Data Set Description . . . . .	36
4.2	Instance Influence Estimation Experiments . . . . .	38
4.2.1	Single Point Experiment . . . . .	38
4.2.2	Mislabeled and Recovery Experiment . . . . .	42
4.2.3	Superpixel Influence Experiment . . . . .	46
4.3	Hyperspectral Unmixing via Semi-supervised PM-LDA Experiments .	49
4.3.1	Hyperspectral Unmixing on Pavia . . . . .	51
4.3.2	Hyperspectral Unmixing on Gulfport . . . . .	64
<b>5</b>	<b>Summary and Concluding Remarks . . . . .</b>	<b>69</b>
	<b>BIBLIOGRAPHY . . . . .</b>	<b>71</b>

## LIST OF TABLES

Table		Page
4.1	Overall proportion map entropy for three methods on Pavia. . . . .	55
4.2	Overall log-likelihood for three methods on Pavia. . . . .	55
4.3	Overall proportion map entropy and log-likelihood for semi-supervised PM-LDA with precise and imprecise labels on Pavia. . . . .	58
4.4	Overall proportion map entropy and log-likelihood for semi-supervised PM-LDA with or without labeling grey roof on Pavia. . . . .	61
4.5	Overall proportion map entropy and log-likelihood for semi-supervised PM-LDA using different initializations on Pavia. . . . .	64
4.6	Overall proportion map entropy for three methods on Gulfport. . . . .	66
4.7	Overall log-likelihood for three methods on Gulfport. . . . .	66
4.8	Overall proportion map entropy and log-likelihood for semi-supervised PM-LDA with precise and imprecise labels on Gulfport. . . . .	67

## LIST OF FIGURES

Figure	Page
1.1 Hyperspectral unmixing illustration on a subset of the RIT SHARE 2012 AVON AM Data. . . . .	2
2.1 Graphical model representation of LDA. . . . .	23
2.2 Graphical model representation of PM-LDA. . . . .	25
3.1 Hierarchical relationship on Pavia . . . . .	33
4.1 RGB image of MUUFL Gulfport and University of Pavia data sets. .	37
4.2 Labeled ground truth on Gulfport and Pavia . . . . .	39
4.3 Instance Influence on Gulfport . . . . .	41
4.4 Instance Influence on Pavia . . . . .	42
4.5 Estimated target endmembers . . . . .	45
4.6 Superpixels on Gulfport and Pavia . . . . .	46
4.7 Instance Influence on Gulfport superpixels . . . . .	48
4.8 Instance Influence on Pavia superpixels . . . . .	50
4.9 Superpixels and OSM regions on Pavia . . . . .	52
4.10 Estimated endmember means on Pavia with different methods . . . .	53

4.11	Estimated proportion maps on Pavia with different methods . . . . .	54
4.12	Estimated endmember means on Pavia with or without precise labels	57
4.13	Estimated proportion maps on Pavia with or without precise labels .	57
4.14	Estimated endmember means on Pavia with or without labeling grey roof . . . . .	60
4.15	Estimated proportion maps on Pavia with or without labeling grey roof	61
4.16	Estimated endmember means on Pavia with different initializations .	62
4.17	Estimated proportion maps on Pavia with different initializations . .	63
4.18	Superpixels and OSM regions on Gulfport . . . . .	65
4.19	Estimated endmember means on Gulfport with different methods . .	65
4.20	Estimated proportion maps on Gulfport with different methods . . .	66
4.21	Estimated endmember means on Gulfport with or without precise labels	67
4.22	Estimated proportion maps on Gulfport with or without precise labels	67



## ABSTRACT

In the past several decades, hyperspectral imaging has drawn a lot of attention in the field of remote sensing. Yet, due to low spatial resolutions of hyperspectral imagers, often the response from more than one surface material can be found in some hyperspectral pixels. These pixels are called mixed pixels. Mixed pixels bring challenges to traditional pixel-level applications, such as identification and detection of ground targets [1, 2]. To address these challenges, hyperspectral unmixing is often an important step during analysis of hyperspectral imagery. Hyperspectral unmixing is the task of decomposing each pixel into a set of pure material signatures (called *endmembers*) with the corresponding proportions of each material found in each pixel. In this thesis, novel hyperspectral unmixing approaches are proposed that leverage interactive labeling and semi-supervised approaches to improve unmixing results.

Hyperspectral unmixing is conducted using an unsupervised approach. Fully supervised hyperspectral unmixing approaches are generally infeasible as the number of pixels in a hyperspectral image is often in the tens- or hundreds-of-thousands and obtaining accurate pixel-level labels is prohibitively expensive or simply impossible. Thus, many hyperspectral unmixing algorithms in literature only use the hyperspectral image. However, there are often other ancillary information regarding the image available that could be used to assist the unmixing process and do not add an infeasible amount of labeling tasks. Motivated by this, novel semi-supervised hyperspectral unmixing techniques are proposed. The proposed semi-supervised approaches allow supervision over hyperspectral unmixing to improve the unmixing performance but do not require complete pixel-level accurate labels for a training data set. For the

semi-supervised hyperspectral unmixing, however, there are some challenges. The first one is to identify which pixels need to be (re-)labeled interactively during the unmixing process. We solve this problem by proposing an instance influence estimation approach. This approach estimates the influence of each pixel based on its impact on target endmember estimation and simplifies the job of the analyst by indicating the pixels with the highest influence during unmixing and endmember estimation. This allows an analyst to only (re-)label the data points with the highest level of influence and to refine the target endmember estimation interactively. The second one is to design a new semi-supervised hyperspectral unmixing method which is capable of introducing a semi-supervised learning phase to hyperspectral unmixing. We solve this challenge by proposing a semi-supervised Partial Membership Latent Dirichlet Allocation (sPM-LDA) approach. The supervision in this approach introduces document-level labeling, restricts and allows one to several candidate topics for each document, which leads to more sparse and precise estimation results. The third challenge is how to find available ancillary information regarding the image data set. We address this challenge by incorporating map data (e.g. Open Street Map) to provide the spatial information and imprecise labels.

Results on real hyperspectral data sets indicate that the proposed instance influence estimation methods can effectively identify the instances, of which labels are most influential for target endmember estimation. Furthermore, experiments on semi-supervised PM-LDA show that overall hyperspectral unmixing results are more accurate than the comparison unmixing that also address spectral variability.

# Chapter 1

## Introduction

Hyperspectral imaging originates from the principle that solar light interacting with a material surface results in the phenomenon where the radiation magnitude that is reflected, emitted and absorbed varies with wavelength [3]. A hyperspectral spectrometer collects the radiance reflected/emitted by each pixel area on the material surface over a wide range of wavelength bands [4]. This results in each hyperspectral pixel being a vector with the response over hundreds of continuous, single narrow wavelength bands. Each hyperspectral image, then, is a stack of hundreds of band-specific images. Mathematically, the single band image is a two-dimensional image covering spatial information represented by  $(x,y)$  for a particular wavelength,  $\lambda$ , and the hyperspectral image is a three-dimensional data cube represented by  $(x,y,\lambda)$  (i.e., two spatial and one spectral dimension).

## 1.1 Hyperspectral Unmixing

Each hyperspectral image corresponds to an image scene that generally consists of at least several different types of materials, such as, vegetation, asphalt, and soil. The pure spectral signature of each type of pure material is called an *endmember*.

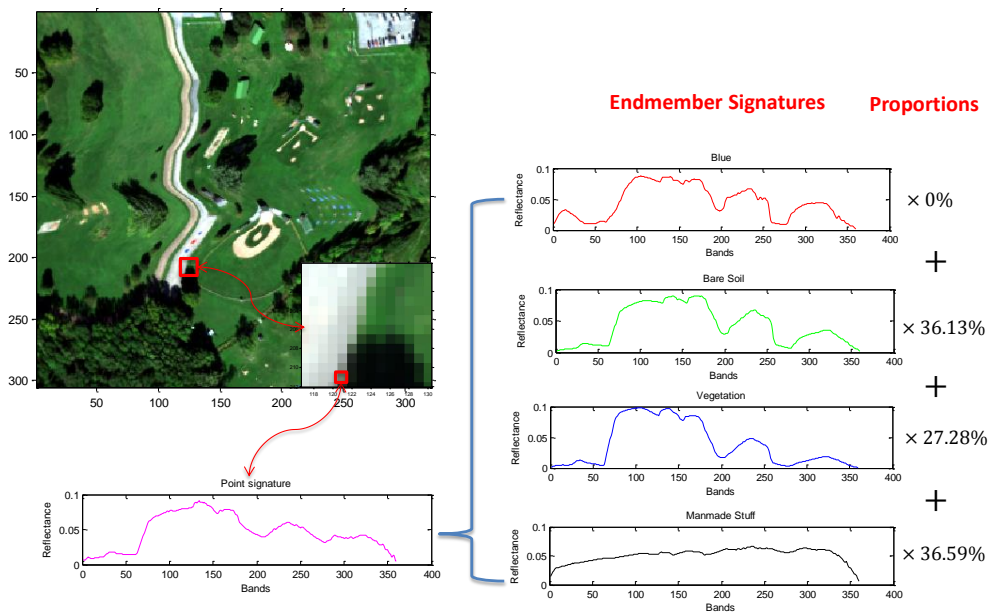


Figure 1.1: Hyperspectral unmixing illustration on a subset of the RIT SHARE 2012 AVON AM Data.

The spatial resolution of a hyperspectral sensor determines the smallest area that can be measured on the earth's surface by the sensor [4]. A hyperspectral pixel may cover more than one endmember because of the low spatial resolution. Due to the sub-pixel endmember mixture, each pixel signature is assumed to be the mixture of endmembers existing in the pixel with corresponding proportions. Therefore, hyper-

spectral unmixing is regarded as the process of pixel signature decomposition, that is, decomposing each mixed pixel signature resulting in the endmembers and associated proportions.

To solve the hyperspectral unmixing problem, many mixture models have been proposed to characterize the underlying mixed endmembers in hyperspectral image [5]. One simple and widely used mixed model is Linear Mixture Model (LMM). The LMM assumes that the multiple scattering effect among different endmembers can be ignored and the surface is linearly partitioned according to the proportions [6]. Therefore, the pixel signature can be well represented by a linear mixture of endmember signatures weighted by the associated proportions (and additive noise in some scenarios). Suppose there are  $M$  endmembers in a hyperspectral image. The  $k$ th endmember signature is denoted by  $e_k$  and the proportion vector of the  $k$ th endmember for this  $i$ th pixel is  $p_{ik}$ . The observed pixel signature  $\mathbf{x}_i$  is represented by

$$\mathbf{x}_i = \sum_{k=1}^M \mathbf{e}_k p_{ik} + \epsilon_i \quad (1.1)$$

where  $\epsilon_i$  is the error term accounting for noise.

Hyperspectral unmixing methods normally contain two steps: endmember extraction and proportion estimation. In general, most hyperspectral unmixing methods are summarized into two categories [6]: 1. geometrical based approach such as Pure Pixel (PP) based methods [7, 8], Minimum Volume (MV) based methods [9, 10], 2. statistical methods, such as Normal Compositional model based algorithms [11, 12, 13, 14], Beta Compositional model based algorithm [15], Normal Mixture Model based algorithm [16]. The methods in the first category are usually based on the linear mixture model which estimates the fixed endmembers by exploiting the data simplex. The

methods mentioned in second category could be applied to more general and real cases by assuming that the endmembers are not fixed. Instead, they are statistical distributions. For example, statistical methods are capable of representing the spectral variability within each endmember by assuming the underlying endmember distributions follow the Normal Composition Model (NCM), one of the most widely used models to model spectral variability.

## 1.2 Problem Statement

Hyperspectral unmixing approaches are generally defined as supervised/semi-supervised or unsupervised based on whether they rely on extra information or not, such as existing endmember signatures, spectral library containing the endmembers or another existing endmember extraction method. In this thesis, we introduce a novel supervised/semi-supervised scenario which is not found in current literature. We define a hyperspectral unmixing algorithm as supervised if every pixel label, indicating what endmembers each pixel contains, are available. It is semi-supervised if only a portion of pixels labels or uncertain pixel labels are available. It is unsupervised if no label information is available. Since a hyperspectral scene often consists of hundreds of thousands of pixels, it is generally infeasible to acquire all the pixel labels. Therefore, most of the current works in literature are unsupervised from this perspective since they only rely on the hyperspectral image itself. However, in fact, we often do have some extra information about the scene that can be used to help the unmixing. But there is no existing work that can be considered as a supervised (or even a semi-supervised) method to leverage the knowledge we have for the hyperspectral image.

So one goal of this thesis is to incorporate the additional information to improve the unmixing results in a way that does not add unrealistic burden on the user.

Interactive labeling is defined as a process in semi-supervised learning where labels for training data, coming from experts (human labeling) or other ways (e.g. tags from online resources or maps), can be adjusted or integrated in an intelligent fashion. This is a necessary and difficult problem in hyperspectral unmixing. For instance, the semi-supervised unmixing method *eFUMI* requires only inaccurate and uncertain labels that only indicate the approximate locations of target points. Thus, better label adjustments are good for achieving a better target endmember estimation result. However, no previous work has been done to guide the expert to adjust the labels (re-labeling) in an interactive way. Therefore, another goal of this thesis is to design a method to find the most influential points whose labels are needed to be adjusted with higher priority. Meanwhile, the method is expected to be automatic and capable of returning the desired influential points in a single run (of *eFUMI*) to ease the relabeling load of user.

In this thesis, two novel semi-supervised hyperspectral unmixing approaches that address the challenges introduced here are introduced.

# Chapter 2

## Literature Review

In this section, current methods for hyperspectral unmixing are reviewed, which are the conventional definitions about supervised, semi-supervised and unsupervised hyperspectral unmixing algorithms. The first problem solved in this thesis is the instance influence estimation, which is based on a semi-supervised hyperspectral unmixing method, the  $\epsilon$ FUMI method. Therefore, a brief review of the  $\epsilon$ FUMI method is provided. Then we proposed a novel semi-supervised hyperspectral unmixing approach. This approach aims to address one of the essential issues among hyperspectral imagery, which is spectral variability. A large number of related works have been proposed considering the endmembers either as sets or as statistical distributions to address spectral variability. These related works about spectral variability are reviewed. The proposed semi-supervised hyperspectral unmixing method stems from the work of PM-LDA, which is an extension of LDA. A brief review of both LDA and PM-LDA is provided. In order to map this proposed method to PM-LDA, the hyperspectral image needs to be segmented to superpixels first. So the used super-



pixel generation algorithm, hyperspectral SLIC, is briefly reviewed. Lastly, since the semi-supervised process becomes more intelligent and automatic with the assistance of Open Street Map (OSM). The OSM is also briefly reviewed.

## 2.1 Semi-supervised Hyperspectral Unmixing

Hyperspectral unmixing methods are commonly categorized as supervised, semi-supervised and unsupervised. Unsupervised hyperspectral unmixing represents the situation where the endmember signatures are unknown and no extra endmember extraction algorithm or spectral library is provided. Thus both endmember signatures and proportion values are estimated by a hyperspectral unmixing approach. Recently, the hyperspectral unmixing problem has been tackled in a supervised/semi-supervised fashion [17]. There is no strict difference in definition between supervised and semi-supervised hyperspectral unmixing. All current supervised and semi-supervised unmixing assume that the hyperspectral image signatures can be represented by the linear combinations of some pure endmember signatures that are known in advance. In the current literature, there are three major categories of supervised/semi-supervised techniques: 1. Endmember signatures are given (supervised); 2. Endmember signatures are unknown but the spectral library is given (semi-supervised); 3. Endmember signatures are unknown but endmember extraction is used from another approach (supervised/semi-supervised). The most simple case is when the endmember signatures are provided in advance as a priori so that the hyperspectral unmixing problem is totally a proportion estimation problem. Algorithms in this situation are categorized as the supervised hyperspectral unmixing [18]. The second case is when endmember

signatures are unknown and the spectral library is provided which contains the underlying endmembers. Information about endmembers is partially given in this scenario. Therefore, the algorithms are designed to find a subset of endmember signatures in the spectral library that can approximate the mixed pixels in the image well. The corresponding algorithms are called the semi-supervised hyperspectral unmixing [19]. In the last case, the endmember signatures are obtained directly from the hyperspectral data set without prior knowledge. However, these unmixing methods must rely on another popular endmember extraction algorithm such as the pixel purity index (PPI) [20], the N-FINDR [7] and the vertex component analysis (VCA) [8] methods. Different articles have different categorization for these external endmember extraction based methods. For example, [21] defines this case as supervised while [19, 22, 23] consider it as semi-supervised hyperspectral unmixing.

## 2.2 Extended Functions of Multiple Instances (eFUMI)

Multiple instance learning (MIL) is a variation of supervised learning [24]. Compared with conventional supervised learning, where each instance is labeled as positive or negative, only a set of bags with positive or negative bag-level labels are provided in multiple instance learning.  $\mathbf{X}_k = [x_1, \dots, x_n]$  is denoted as a  $k$ th bag, with a set of instances  $x_1, \dots, x_n$ . If one given bag label is negative,  $L_k = -$ , then we know that all instances in a bag are negative,  $\mathbf{X}_k = [x_1^-, \dots, x_n^-]$ . If it's positive,  $L_k = +$ , then we know that at least one instance in the bag is labeled as positive,  $\mathbf{X}_k = [x_1^+, \dots, x_i^+, x_{i+1}^-, \dots, x_n^-]$ . Now given only the positive and negative bag labels, the goal

of multiple instance learning is to learn a concept, the desired target feature vector.

Multiple instance learning method for supervised learning [25] is generalized by Function of Multiple Instance (FUMI). Each pixel is then considered as a function of a positive or negative bag, which is a linear combination of target and non-target endmembers. Both the target and non-target endmembers can be estimated by FUMI. The specific method applied to hyperspectral unmixing is the convex-FUMI (c-FUMI). Only a binary label indicating a data point does or doesn't contain some proportions of target endmember is necessary to be known for each pixel. The target detection is then performed after estimating the target endmember.

eFUMI was first proposed in [26, 27], based on the FUMI approach. The motivation for developing eFUMI is to allow further uncertainty with the data point labels. Even though cFUMI is capable of addressing the unspecificity for unknown proportion values of each data point, the instance-level binary labels are necessary for all data points. However, this may not be applicable in real hyperspectral imagery applications. For example, the typical hyperspectral data collection process does not always provide very accurate geo-locations. UTM coordinates of regions of interest may have the error range of about several pixels. For this situation, instead of accurate instance-level labels, the approximate location labels over several pixels are available. The bag-level labels are more realistic to be acquired than instance-level labels for target and non-target concept estimations. For this reason, the extended FUMI (eFUMI) algorithm was proposed to estimate concepts as well as proportions given uncertain locations for pixels.

eFUMI aims to estimate a target signature  $\mathbf{e}_T$ , non-target signatures  $\{\mathbf{e}_k\}_{k=1}^M$  and the proportion vector  $\mathbf{p}_i$  that characterizes how a mixed data point  $\mathbf{x}_i$  is represented by

endmembers. Suppose the entire data set are divided into  $K$  bags,  $\mathbf{B} = \{B_1, \dots, B_K\}$ , with their corresponding labels,  $L = \{L_1, \dots, L_K\}$ . The data point is represented as a function of concepts,  $\mathbf{x}_i = f(\mathbf{E}_i, \mathbf{p}_i)$ . Linear mixture model for hyperspectral unmixing is used so that this function could be expressed as a convex combination of endmember signatures and proportion vectors,  $\mathbf{x}_i = p_{iT}\mathbf{e}_T + \sum_{k=1}^M p_{ik}\mathbf{e}_k$ . Also, the function should meet the non-negative and sum-to-one constraints for proportions and normalization constraint for concepts, which are,  $p_{iT} + \sum_{k=1}^M p_{ik} = 1$ ,  $p_{iT} \geq 0$ ,  $p_{ik} \geq 0$ ,  $\|\mathbf{e}_k\|^2 = 1$ , and  $\|\mathbf{e}_T\|^2 = 1$ . For the bag-level labels, a bag  $B_j$  is labeled as the target bag  $L_j = 1$ , if at least one instance  $\mathbf{x}_i$  includes a positive target proportion  $p_{iT}$ . To be more specific, this situation can be described in (2.1), where  $\boldsymbol{\varepsilon}_i$  is the additive noise.

$$\text{if } L_j = 1, \exists \mathbf{x}_i \in B_j \text{ s.t. } \mathbf{x}_i = p_{iT}\mathbf{e}_T + \sum_{k=1}^M p_{ik}\mathbf{e}_k + \boldsymbol{\varepsilon}_i, p_{iT} > 0 \quad (2.1)$$

Similarly, a bag  $B_j$  is labeled as the non-target bag  $L_j = 0$ , if no instance in the bag contains any target proportion, which is expressed in (2.2).

$$\text{if } L_j = 0, \forall \mathbf{x}_i \in B_j, \mathbf{x}_i = \sum_{k=1}^M p_{ik}\mathbf{e}_k + \boldsymbol{\varepsilon}_i \quad (2.2)$$

Since the instance-level labels are hidden latent variables, they are addressed by an Expectation Maximization (EM) method. The complete log likelihood function is proportional to equation 2.3.

The goal is to minimize the objective function. The first term in the objective function 2.3 is the error term, the  $L_2$  norm between the original per-pixel signatures with reconstructed per-pixel signatures. This term is multiplied by a weight parameter  $w_i = \frac{\alpha N^-}{N^+}$  for pixels in positive bags and by  $w_i = 1$  for pixels in negative bags,

$$F = -\frac{1}{2}(1-u) \sum_{i=1}^N w_i \left\| \mathbf{x}_i - z_i p_{iT} \mathbf{e}_T - \sum_{k=1}^M p_{ik} \mathbf{e}_k \right\|_2^2 - \frac{u}{2} \sum_{k=1}^M \left\| \mathbf{e}_k - \boldsymbol{\mu}_0 \right\|_2^2 - \frac{u}{2} \left\| \mathbf{e}_T - \boldsymbol{\mu}_0 \right\|_2^2 - \sum_{k=1}^M \gamma_k \sum_{i=1}^N p_{ik} \quad (2.3)$$

$$\begin{aligned} E[F] = & \sum_{z_i \in \{0,1\}} \left[ -\frac{1}{2}(1-u) \sum_{i=1}^N w_i P(z_i | \mathbf{x}_i, \boldsymbol{\theta}^{(t-1)}) \left\| \mathbf{x}_i - z_i p_{iT} \mathbf{e}_T - \sum_{k=1}^M p_{ik} \mathbf{e}_k \right\|_2^2 - \frac{u}{2} \sum_{k=1}^M \left\| \mathbf{e}_k - \boldsymbol{\mu}_0 \right\|_2^2 \right. \\ & \left. - \frac{u}{2} \left\| \mathbf{e}_T - \boldsymbol{\mu}_0 \right\|_2^2 - \sum_{k=1}^M \gamma_k \sum_{i=1}^N p_{ik} \right] \end{aligned} \quad (2.4)$$

where  $N^+$  and  $N^-$  are the numbers of points in positive and negative bags, and  $\alpha$  is a user-selected hyperparameter. This weight parameter ensures that the error term is more fairly influenced by both positive and negative bags instead of by bags with more pixels only. The second and third terms are the endmember restriction terms, which restricts both the target and non-target endmembers to have a moderately tight fit around the data set by minimizing their distances to the global data mean,  $\boldsymbol{\mu}_0$ . The  $u$  is the regularization term set by the user to determine the influence between the error term and endmember restriction term. The last term in the objective function is the sparse promoting term,  $\gamma_k = \frac{\Gamma}{\sum_{i=1}^N p_{ik}^{(t-1)}}$ , which enforces the proportions associated with the unnecessary endmembers to be a very small number or zero. Within the framework of the EM algorithm, the optimization of the objective function is solved iteratively. Since an instance-level label is unknown for each training point, in the E-step, the conditional probability of an instance-level label given parameters and a bag-level label is computed in 2.4, where  $\boldsymbol{\theta}^t$  represents all the parameters estimated at iteration  $t$ , and  $p(z_i | \mathbf{x}_i, \boldsymbol{\theta}^{(t-1)})$  is the probability density function given the parameters estimated from the last iteration. The equations used for  $p(z_i | \mathbf{x}_i, \boldsymbol{\theta}^{(t-1)})$  are shown in 2.5.

$$\begin{aligned}
p(z_i|\mathbf{x}_i, \theta^{(t-1)}) = & \tag{2.5} \\
\left\{ \begin{array}{l}
p(z_i = 0|\mathbf{x}_i \in B_j^+, \theta^{(t-1)}) = e^{-\beta \|\mathbf{x}_i - \sum_{k=1}^M p_{ik} \mathbf{e}_k\|_2^2} \\
p(z_i = 1|\mathbf{x}_i \in B_j^+, \theta^{(t-1)}) = 1 - e^{-\beta \|\mathbf{x}_i - \sum_{k=1}^M p_{ik} \mathbf{e}_k\|_2^2} \\
p(z_i = 0|\mathbf{x}_i \in B_j^-, \theta^{(t-1)}) = 1 \\
p(z_i = 1|\mathbf{x}_i \in B_j^-, \theta^{(t-1)}) = 0
\end{array} \right.
\end{aligned}$$

where  $\beta$  is the scaling factor, and  $r_b = \left\| \mathbf{x}_i - \sum_{k=1}^M p_{ik} \mathbf{e}_k \right\|_2^2$  is the approximate reconstructed error between the pixel  $\mathbf{x}_i$  and its reconstructed signature by using only the non-target endmembers. The definition of  $p(z_i|\mathbf{x}_i, \theta^{(t-1)})$  in 2.5 shows that, for a non-target pixel in target bag, it should be fully represented by the non-target endmembers so that the reconstructed error  $r_b$  is small. Specifically,  $p(z_i = 0|\mathbf{x}_i \in B_j^+, \theta^{(t-1)}) = e^{-\beta r_b} \rightarrow 1$ . On the contrary, for a target pixel  $\mathbf{x}_i$ , if we represent it with only non-target endmembers, reconstructed error  $r_b$  may be large. So the probability density function is defined as  $p(z_i = 1|\mathbf{x}_i \in B_j^+, \theta^{(t-1)}) = 1 - e^{-\beta r_b} \rightarrow 1$ .  $z_i$  is unknown only for pixels in target bags because target bags can contain target and non-target pixels, which are a fixed value 0 for pixels in non-target bags because all pixels in non-target bags are non-target pixels.

The M-step is performed after E-step to maximize the expectation value 2.4 via optimization approaches to update each parameter. The whole process is outlined in Algorithm 1 <sup>1</sup>.

---

<sup>1</sup>The pseudo code part for this algorithm comes from the article [26].

---

**Algorithm 1** *e*FUMI EM algorithm [26]

---

**Input:**

- 1: Initialize  $\theta^0 = \{\mathbf{e}_T, \mathbf{E}, \mathbf{P}\}$ , normalize input data,  $t = 1$
  - 2: **repeat**
  - 3:        *E-step*: Compute (2.5) given  $\theta^{t-1}$
  - 4:        *M-Step*:
  - 5:            Update  $\mathbf{e}_T$  by minimizing (2.4) w.r.t.  $\mathbf{e}_T$
  - 6:            Update  $\mathbf{E}$  by minimizing (2.4) w.r.t.  $\mathbf{E}$
  - 7:            Update  $\mathbf{P}$  by minimizing (2.4) w.r.t.  $\mathbf{P}$
  - 8:         $t \leftarrow t + 1$
  - 9: **until** Convergence
  - 10: **return**  $\mathbf{e}_T, \mathbf{E}, \mathbf{P}$
- 

## 2.3 Spectral Variability

One prominent issue studied in hyperspectral image analysis is spectral variability. Spectral variability is categorized into two types: the variability inside the single end-member itself (intra-class variability) and the similarity among different endmembers (inter-class variability) [28]. The intra-class variability is discussed and addressed in this thesis. The term *spectral variability* simply denotes intra-class variability in all the following chapters.

The measured spectral signatures in the hyperspectral image can vary due to a lot of reasons such as variable illuminations and environmental, atmospheric and temporal conditions [29], as well as the intrinsic material spectral variability. Variation from illumination is commonly believed to be the major cause for spectral variability [30, 31]. The illumination change can result from different shapes or roughness of the same material on different locations of a hyperspectral image. The shape or roughness difference creates tilted surfaces resulting in different degrees of shadows on the source. For example, the canopies of different trees (each leaf has a different angle) or architectures of different buildings have different illumination levels. The

illumination can also be impacted by the light source, for instance, the sun. The variation of solar altitude and incident angle may lead to variation of illumination. For mineral materials, the illumination variation can result from distinct grain size and texture because grain size impacts the backscatter energy and further impacts the reflectance level [32].

Another predominant reason for spectral variability is different atmospheric conditions. To be more specific, a number of different types of gases (such as oxygen, water vapor, etc.) in the sky can absorb and scatter the radiation resulting in distortion of the spectral signatures of the hyperspectral image. These atmospheric gases can absorb or scatter radiation on certain wavelengths of materials. Therefore, the measured spectral signatures are changed during the downward and upward transmission process, where the solar light travels through the atmospheric gases to reach the materials and then reflects through the gases again and finally reaches the spectrometer. Sometimes the reflected lights don't totally travel through the gases but are reflected again, so multiple reflections occur during the process. Each reflection on the gases may lead to different levels of absorption and scattering.

Even though the effect of spectral variability exists, it has not been handled properly in most state-of-the-art hyperspectral unmixing algorithms, especially the linear mixture model based methods. It is logical that the higher the spectral variability, the higher the possibility that underlying signatures of the endmembers deviate from the fixed estimated endmembers [28].

A number of approaches have been proposed to address spectral variability in hyperspectral unmixing. There are two categories for representing these approaches: *endmembers as sets* and *endmembers as statistical distributions*. The former category



is based on the linear mixture model, and the latter category can be regarded as the stochastic mixing model. In the following subsections, methods that account for these two categories are reviewed, respectively.

### 2.3.1 Endmembers as Sets

Some algorithms address spectral variability by constructing a set of spectral signatures for each endmember instead of a single spectral signature under linear mixture model. The pixel signature can then be viewed as the convex combination of elements of endmember sets (one element from one set in most algorithms). From the perspective of the spectral library, the algorithms are further divided into two types.

For algorithms given the spectral library, MESMA [33] and its extensions MELSUM, BSMA, AutoMCU exhaustively search for one or more signatures for each endmember from the spectral library such that the estimated proportion values satisfy some predefined criteria. The signatures found, corresponding to each endmember, are grouped together to form the endmember sets. The endmember bundles unmixing approach [34] estimates the minimum and maximum proportion values for the  $k$ th material of pixel  $\mathbf{x}_i$  given  $\mathbf{E}_k$ , the endmember set containing all signature variants of the  $k$ th material. Band selection or weighting approaches [28] select the wavelengths that have the minimum spectral variability for proportion estimation. Support Vector Machine (SVM) based algorithms [35, 36] address the issue by training a set of SVM classifiers, representing different discretized proportion value combinations. The unmixing results are then found by the SVM giving the largest classification margin for a unknown pixel.

For algorithms without the spectral library, the endmember sets are directly ob-

tained from the pixels in the hyperspectral image. Automated endmember bundles [37] automatically selects a portion of pixels and applies an endmember extraction algorithm (e.g. VCA) to estimate endmembers. This process is repeated for several times and generates a number of endmembers. Then the endmember sets are obtained by applying a clustering algorithm (e.g. K-means) to these endmembers. Sparse unmixing [38] first estimates the endmember set of each endmember by minimizing a cost function with a sparse promoting term. After obtaining these dictionary elements, the unmixing is conducted by minimizing another cost function to estimate the proportion values. Local unmixing [39, 40] first extracts endmembers locally (from spatial neighborhoods) manually or by some endmember extraction algorithms. Then, the endmembers are grouped using spectral angle criteria to construct the endmember sets. After the endmember sets are determined, the proportion values are estimated by minimizing an objective function.

### 2.3.2 Endmembers as Statistical Distributions

Under the *endmembers as statistical distributions* approach, each endmember is considered as a statistical distribution rather than a set or a single value. Each sample from the endmember distribution can be regarded as a variant of the endmember signature. The variant is shown in 2.6.

$$\mathbf{e}_k \sim \mathcal{F}(\cdot|\theta_k) \tag{2.6}$$

where  $\mathcal{F}$  is the statistical distribution for the  $m$ th endmembers  $\mathbf{e}_k$  and  $\theta_k$  represents the unknown parameters for this distribution.

Therefore, for endmembers following the statistical distribution, each pixel is regarded as the convex combination of these distribution-based endmembers. To be more specific, the pixel signature  $\mathbf{x}_i$  can be written as

$$\mathbf{x}_i = \sum_{k=1}^M p_{ik} \mathbf{e}_k \quad (2.7)$$

where  $p_{ik}$  is the proportion value of one variant of the endmember distribution  $\mathbf{e}_k$  for pixel  $\mathbf{x}_i$ . Thus, compared with the standard linear mixture model, this category of models assumes that each pixel signature can be represented by a linear combination of variants from endmember distributions with associated proportion values.

A large number of endmember-distribution based, unmixing algorithms are based on a Bayesian framework. Once the distributions are indicated, both the distribution parameters and proportion values can be estimated simultaneously. The most commonly used distribution to represent endmembers is the normal distribution, that is,

$$\mathcal{F}(\mathbf{e}_k | \theta_k) = \mathcal{N}(\mathbf{e}_k | \mu_k, \Sigma_k) \quad (2.8)$$

where  $\mu_k$  is the mean parameter and  $\Sigma_k$  is the covariance parameter for endmember  $\mathbf{e}_k$ . The corresponding model assuming normal distributions for endmembers is named the *Normal Compositional Model* (NCM) [11]. We assume the variants from endmember distributions are mutually independent normal distribution variables. According to 2.6, 2.7 and 2.8, the pixel signature  $\mathbf{x}_i$  under the NCM is represented as

$$\mathbf{x}_i \sim \mathcal{N}\left(\cdot \mid \sum_{k=1}^M p_{ik} \mu_k, \sum_{k=1}^M p_{ik}^2 \Sigma_k\right) \quad (2.9)$$

Several approaches were proposed based on NCM for hyperspectral unmixing.

### Nested Stochastic EM

Parameter estimation was addressed initially in [11] with a method based on the nested stochastic expectation maximization algorithm (SEM). The proportion values  $p_{ik}$  are regarded as latent, hidden variables. The complete likelihood function is

$$p(\mathbf{x}_1, \dots, \mathbf{x}_N, p_{11}, \dots, p_{1M}, \dots, p_{N1}, \dots, p_{NM} | \{\mu_k, \Sigma_k\}) \quad (2.10)$$

$$= \prod_{i=1}^N \mathcal{N}(\mathbf{x}_i; \mu(\alpha_i), \Sigma(\alpha_i)) p(p_{i1}, \dots, p_{iM}) \quad (2.11)$$

where  $N$  is the number of pixels in the data set.

The SEM algorithm contains four main steps shown as below <sup>2</sup>.

(1) Initialization of the endmember distribution parameters

$$\Omega = \{(\mu(\alpha_i), \Sigma(\alpha_i)) | 0 \leq i \leq M\} \quad (2.12)$$

(2) Sampling of the hidden variables  $\{p_{ik} | 1 \leq k \leq M, 1 \leq i \leq N\}$  to generate a psuedo-complete sample through the selection of the proportion values maximizing the likelihood function, i.e., by solving

$$\hat{\alpha}_i = \arg(\max_{\alpha_i} (\mathcal{N}(\mathbf{x}_i; \mu(\alpha_i), \Sigma(\alpha_i)))) \quad (2.13)$$

(3) Maximization of the psuedo-complete likelihood function over  $\Omega$ , the update

equations for distribution parameters  $\{\mu_k, \Sigma_k\}$  after  $l$  iterations are

$$\mu_k^{l+1} = \frac{1}{N} \sum_{i=1}^N E(\varepsilon_k | \mathbf{x}_i, \{p_{ik}\}, \Omega^l) \quad (2.14)$$

$$\Sigma_k^{l+1} = \frac{1}{N} \sum_{i=1}^N cov(\varepsilon_k | \mathbf{x}_i, \Omega^l) + [E(\varepsilon_k | \mathbf{x}_i, \Omega^l) - \mu_k^{l+1}][E(\varepsilon_k | \mathbf{x}_i, \Omega^l) - \mu_k^{l+1}]^t \quad (2.15)$$

(4) Iteration of step (2) and (3) until a convergence criterion is satisfied.

### MCMC given endmember means

[12] developed a Markov chain Monte Carlo (MCMC) method for NCM-based hyperspectral unmixing if the endmember means are known. The endmember extraction algorithm such as VCA or N-FINDR, is applied first to estimate the endmember means. Then, the endmember covariances and proportion values are estimated using MCMC. Several parameter priors are imposed. The Dirichlet prior is assigned to the proportions to enforce the non-negative and sum-to-one constraints, that is,  $\mathbf{p}_n \sim \mathcal{D}(\cdot | \mathbf{1})$ . The covariance matrix of each endmember can be written as  $\sigma^2 \mathbf{I}_L$ , where  $\mathbf{I}_L$  is the  $L \times L$  identity matrix and  $\sigma^2$  is the endmember variance in any spectral band. A conjugate inverse Gamma distribution is imposed on the variance  $\sigma^2$  as the prior, given by  $\sigma^2 | \delta \sim \mathcal{IG}(\nu, \delta)$ , where  $\nu$  and  $\delta$  are two user-defined hyperparameters (shape and scale parameters). A non-informative Jeffreys' prior is assigned to the  $\delta$ , which is  $f(\delta) \propto \frac{1}{\delta} \mathbf{1}_{R^+}(\delta)$ .

The authors further derived the distributions for each parameter. For the proportion vector, the conditional distribution is

---

<sup>2</sup>The main steps part for this algorithm comes from the article [11].

$$f(\mathbf{p}|\mathbf{x}, \sigma^2) \propto \frac{1}{[\sigma^2 c(\mathbf{p})]^{L/2}} \exp\left(-\frac{\|\mathbf{x} - \mu(\mathbf{p})\|^2}{2\sigma^2 c(\mathbf{p})}\right) \mathbf{1}_S(\mathbf{p}) \quad (2.16)$$

where  $\mathbf{1}_S(\cdot)$  is the indicator function defined on the simplex  $S$ .

For the variance, the conditional distribution is

$$\sigma^2|\mathbf{x}, \mathbf{p}, \delta \sim \mathcal{IG}\left(\frac{L}{2} + 1, \frac{\|\mathbf{x} - \mu(\mathbf{p})\|^2}{2c(\mathbf{p})} + \delta\right) \quad (2.17)$$

For  $\delta$ , the conditional distribution is

$$\delta|\sigma^2 \sim \mathcal{G}\left(1, \frac{1}{\sigma^2}\right) \quad (2.18)$$

where  $\mathcal{G}(a, b)$  is the Gamma distribution with shape parameter  $a$  and scale parameter  $b$ .

The pseudo code for using hybrid Gibbs sampler for hyperspectral unmixing is summarized in Algorithm 2 <sup>3</sup>.

---

**Algorithm 2** Hybrid Gibbs sampler for hyperspectral unmixing using the NMC [12]

---

- 1: **Initialization:** Sample  $\delta^0$  from the probability density function (pdf),  $f(\delta) \propto \frac{1}{\delta} \mathbf{1}_{R^+}(\delta)$ . Sample  $\sigma^{2^0}$  from the pdf,  $\sigma^2|\delta \sim \mathcal{IG}(\nu, \delta)$ .
  - 2: **repeat**
  - 3:       Sample  $\mathbf{p}^t$  from the pdf in 2.16 using Metropolis-within-Gibbs step.
  - 4:       Sample  $\sigma^{2^t}$  from the pdf in 2.17
  - 5:       Sample  $\delta^t$  from the pdf in 2.18
  - 6:        $t \leftarrow t + 1$
  - 7: **until** Convergence
  - 8: **return**  $\mathbf{p}, \sigma^2, \delta$
- 

### MCMC given endmember covariances

<sup>3</sup>The pseudo code part for this algorithm comes from the article [12].

In terms of hyperspectral data which are usually nonconvex, [13] proposed to use several convex regions instead of a single convex region to represent the whole data set. The algorithm can automatically determine a few endmember distribution sets. Each set is viewed as a random simplex, where each vertex is modeled as an endmember distribution under the normal compositional model. The pixels are then divided into different sets according to the convex regions using a Dirichlet process prior. The Metropolis-within-Gibbs sampler is applied to divide the data set into convex regions with learned number of regions and estimating the endmember distributions and proportion values for each convex region.

The endmember distribution covariances are assumed to be known in advance and data means are assumed to be drawn from the normal distributions  $\mu_k \sim \mathcal{N}(\cdot | \mathbf{m}, \mathbf{C})$ . The proportion vector for pixel  $\mathbf{x}_i$  is modeled as  $\mathbf{p}_i \sim \mathcal{D}(\cdot | \mathbf{1})$ . The  $\mathbf{m}$  and  $\mathbf{C}$  are hyperparameters defining the prior distribution on the endmember means. The sampling method is shown in Algorithm 3 <sup>4</sup>.

## 2.4 Partial Membership Latent Dirichlet Allocation (PM-LDA)

The proposed method is based on the framework of Partial Membership Latent Dirichlet Allocation (PM-LDA). PM-LDA is based on previous work of Latent Dirichlet Allocation (LDA). This section is partitioned into two parts, the former subsection is a review of LDA and the latter subsection introduces the basic principle of PM-LDA.

---

<sup>4</sup>The pseudo code part for this algorithm comes from the article [13].

---

**Algorithm 3** Sampling Piecewise Convex Unmixing and Endmember Extraction (S-PCUE) [13]

---

```

1: Set parameter values and initialize partition.
2: for  $r \leftarrow 1$  to  $R_{initial}$  convex sets do
3:     Initialize  $\mathbf{E}_r$ ,  $\mathbf{P}_r$ , and  $\mathbf{C}_r$ .
4: end for
5: for  $k \leftarrow 1$  to number of total iterations do
6:     for  $r \leftarrow 1$  to number of convex sets do
7:         for  $j \leftarrow 1$  to number of data points do
8:             Sample proportions  $\mathbf{p}_j$  for  $\mathbf{x}_j$  for each set of endmembers
9:         end for
10:        for  $k \leftarrow 1$  to number of endmembers in convex set  $r$  do
11:            Sample  $e_{k,r}$  in convex set  $r$ 
12:        end for
13:        Sample  $\mu_r$ 
14:        Sample  $\mathbf{C}_r$ 
15:    end for
16:    for  $k \leftarrow 1$  to  $K$  do
17:        Sample new  $\mathbf{E}_k^*$  and  $\mathbf{P}_k^*$  matrices.
18:    end for
19:    for  $j \leftarrow 1$  to number of data points do
20:        Remove  $\mathbf{x}_j$  from its current convex set.
21:        Compute DP partition probabilities for  $\mathbf{x}_j$ .
22:        Sample a convex set for  $\mathbf{x}_j$  based on the DP partition probabilities.
23:        if A new convex set is sampled then
24:            Add the new endmember distribution set to  $\mathbf{E}$  and assign  $\mathbf{x}_j$  to this set
25:        else
26:            Update the label of  $\mathbf{x}_j$  to the sampled endmember distribution set
27:        end if
28:    end for
29: end for

```

---



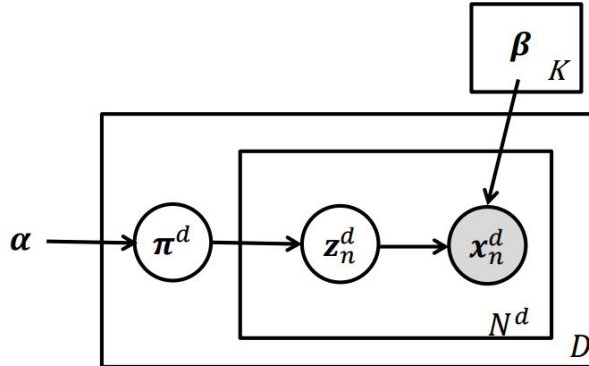


Figure 2.1: Graphical model representation of LDA.

### 2.4.1 Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) [41] is a generative statistical model for sets of observations such as text corpora, which learns the latent topics and assigns documents with latent topics based on a hierarchical Bayesian model. The LDA assumes a three-layer hierarchical model, which is corpus-document-word. Each corpus contains a bag of documents and each document contains a bag of words. Each word is assumed to belong to one and only one topic. Thus, each document generally has several topics with a certain combination of topic proportions. For parameter inference of LDA, given the corpus and prior parameters for topic proportion and topic distribution, the essential problem is to determine the underlying topic proportion for each document and topic assignment for each word.

LDA assumes the following generative process for each document  $\mathbf{w}_d$  in a corpus  $\mathbf{D}$ :

- (1) For each document,  $\mathbf{w}_d$ , choose topic proportions  $\pi_d \sim Dir(\alpha)$ .
- (2) For each word position,
  - (a) Choose topic assignment  $z_{d,n} \sim Mult(\pi_d)$ .

(b) Choose word  $w_{d,n} \sim Mult(\beta_{z_{d,n}})$ .

Step (1) is at document-level; LDA assumes that the topic proportions are drawn from Dirichlet distribution with hyperparameter  $\alpha$ . Step (2) is at word-level; The topic  $z_{d,n}$  assigned to each word is assumed to be drawn from a Multinomial distribution given the parameter topic proportion. The word assigned for each word location is assumed to be drawn from a Multinomial distribution given the parameter  $\beta_{z_{d,n}}$ .

The graphical model representation of LDA is shown in Figure 2.1 <sup>5</sup>.

## 2.4.2 Partial Membership Latent Dirichlet Allocation

Partial Membership Latent Dirichlet Allocation (PM-LDA) [42, 43] is an extension of LDA allowing each word to have partial membership over several topics. The use of partial memberships allows for topic modeling in which crisp topic assignments are insufficient since words may straddle multiple topics simultaneously.

PM-LDA assumes the following generative process for each document  $\mathbf{w}_d$  in a corpus  $\mathbf{D}$ , as is shown in Figure 2.2 <sup>6</sup>:

(1) For each document,  $\mathbf{X}_d$ ,

(a) choose topic proportions  $\pi_d \sim Dir(\alpha)$

(b) Choose scaling factor  $s^d$  from an exponential distribution  $s^d \sim exp(\lambda) = \lambda e^{-\lambda s}$

(2) For each word position,

(a) Choose the topic membership  $z_n^d \sim Dir(\pi^d s^d)$

(b) Choose the word  $x_n^d \sim Expon(\sum_k z_{nk}^d \eta_k)$

---

<sup>5</sup>The Figure comes from the article [42].

<sup>6</sup>The Figure comes from the article [42].

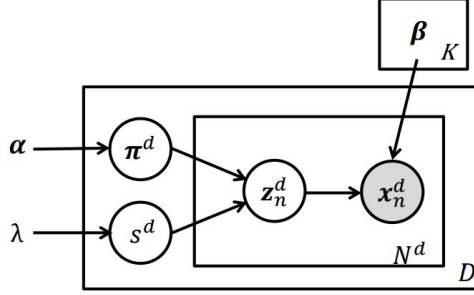


Figure 2.2: Graphical model representation of PM-LDA.

In the PM-LDA model, the random variable associated with a data point,  $\mathbf{x}$ , is assumed to be distributed according to multiple topics with a continuous partial membership in each topic,  $\mathbf{z}$ . Specifically, the PM-LDA model is

$$p(\boldsymbol{\pi}^d, s^d, \mathbf{z}_n^d, \mathbf{x}_n^d | \boldsymbol{\alpha}, \lambda, \boldsymbol{\beta}) = p(\boldsymbol{\pi}^d | \boldsymbol{\alpha}) p(s^d | \lambda) p(\mathbf{z}_n^d | \boldsymbol{\pi}^d, s^d) \prod_{k=1}^K p_k(\mathbf{x}_n^d | \beta_k)^{z_{nk}^d} \quad (2.19)$$

where  $\mathbf{x}_n^d$  is the  $n$ th word in document  $d$ ,  $\mathbf{z}_n^d \sim \text{Dir}(\boldsymbol{\pi}^d \mathbf{s}^d)$  is the partial membership vector of  $\mathbf{x}_n^d$ ,  $\boldsymbol{\pi}^d \sim \text{Dir}(\boldsymbol{\alpha})$  and  $s^d \sim \exp(\lambda)$  are the topic proportion vector and the level of topic mixing in document  $d$ , respectively. Given hyperparameters  $\boldsymbol{\Psi} = \{\alpha, \lambda\}$  and the data set (which has been partitioned into documents),  $\mathbf{X} = \{\mathbf{X}^1, \mathbf{X}^2, \dots, \mathbf{X}^D\}$ , the goal of parameter estimation given the PM-LDA model is to estimate the topic proportion of each document,  $\boldsymbol{\pi}^d$ , the topic mixing level in each document,  $s^d$ , the partial memberships of each word in each topic,  $\mathbf{z}_n^d$ , and the parameters defining the probability distribution of each topic,  $\beta_k$ . Algorithm 4<sup>7</sup> summarizes a Metropolis-within-Gibbs sampler to perform parameter estimation.  $\boldsymbol{\Pi}, \mathbf{S}, \mathbf{M}$  denote the set for topic proportion  $\boldsymbol{\pi}, \mathbf{s}, \mathbf{Z}$ , respectively.

---

**Algorithm 4** Metropolis-within-Gibbs Sampling Method for Parameter Estimation [14]

---

**Input:** A corpus  $\mathbf{D}$ , the number of topics  $K$ , hyperparameters  $\Psi = \{\alpha, \lambda\}$ , and the number of iterations  $T$

**Output:** Collection of all samples:  $\Pi^{(t)}, \mathbf{S}^{(t)}, \mathbf{M}^{(t)}, \beta^{(t)}$

- 1: **for**  $t = 1 : T$  **do**
  - 2:   **for**  $d = 1 : D$  **do**
  - 3:     Sample  $\pi^d$ : Draw candidate:  $\pi^\dagger \sim \text{Dir}(\alpha)$   
    Accept candidate with probability:  

$$a_\pi = \min \left\{ 1, \frac{p(\pi^\dagger, s^{(t-1)}, \mathbf{Z}^{(t-1)}, \mathbf{X} | \Psi) p(\pi^{(t-1)} | \alpha)}{p(\pi^{(t-1)}, s^{(t-1)}, \mathbf{Z}^{(t-1)}, \mathbf{X} | \Psi) p(\pi^\dagger | \alpha)} \right\}$$
  - 4:     Sample  $s^d$ : Draw candidate:  $s^\dagger \sim \exp(\lambda)$   
    Accept candidate with probability:  

$$a_s = \min \left\{ 1, \frac{p(\pi^{(t)}, s^\dagger, \mathbf{Z}^{(t-1)}, \mathbf{X} | \Psi) p(s^{(t-1)} | \lambda)}{p(\pi^{(t)}, s^{(t-1)}, \mathbf{Z}^{(t-1)}, \mathbf{X} | \Psi) p(s^\dagger | \lambda)} \right\}$$
  - 5:     **for**  $n = 1 : N^d$  **do**
  - 6:       Sample  $\mathbf{z}_n^d$ : Draw candidate:  $\mathbf{z}_n^\dagger \sim \text{Dir}(\mathbf{1}_K)$   
       Accept candidate with probability:  

$$a_{\mathbf{z}} = \min \left\{ 1, \frac{p(\pi^{(t)}, s^{(t)}, \mathbf{z}_n^\dagger, \mathbf{x}_n | \Psi)}{p(\pi^{(t)}, s^{(t)}, \mathbf{z}_n^{(t-1)}, \mathbf{x}_n | \Psi)} \right\}$$
  - 7:       **end for**
  - 8:     **end for**
  - 9:     **for**  $k = 1 : K$  **do**
  - 10:       Sample  $\mu_k$ : Draw proposal:  $\mu_k^\dagger \sim \mathcal{N}(\cdot | \mu_{\mathbf{D}}, \Sigma_{\mathbf{D}})$   
        $\mu_{\mathbf{D}}$  and  $\Sigma_{\mathbf{D}}$  are mean and covariance of the data  
       Accept candidate with probability:  

$$a_k = \min \left\{ 1, \frac{p(\Pi^{(t)}, \mathbf{S}^{(t)}, \mathbf{M}^{(t)}, \mathbf{D} | \mu_k^\dagger) \mathcal{N}(\mu_k^{(t-1)} | \mu_{\mathbf{D}}, \Sigma_{\mathbf{D}})}{p(\Pi^{(t)}, \mathbf{S}^{(t)}, \mathbf{M}^{(t)}, \mathbf{D} | \mu_k^{(t-1)}) \mathcal{N}(\mu_k^\dagger | \mu_{\mathbf{D}}, \Sigma_{\mathbf{D}})} \right\}$$
  - 11:       **end for**
  - 12:       Sample covariance matrices  $\Sigma = \sigma^2 \mathbf{I}$ :  
       Draw candidate from:  $\sigma^2 \sim \text{Unif}(0, u)$   
       with  $u = \frac{1}{2} \{ \max_{\mathbf{x}_n} d^2(\mathbf{x}_n - \mu_{\mathbf{D}}) - \min_{\mathbf{x}_n} d^2(\mathbf{x}_n - \mu_{\mathbf{D}}) \}$   
       Accept candidate with probability:  

$$a_\Sigma = \min \left\{ 1, \frac{p(\Pi^{(t)}, \mathbf{S}^{(t)}, \mathbf{M}^{(t)}, \mathbf{D} | \Sigma^\dagger)}{p(\Pi^{(t)}, \mathbf{S}^{(t)}, \mathbf{M}^{(t)}, \mathbf{D} | \Sigma^{(t-1)})} \right\}.$$
  - 13:     **end for**
-

## 2.5 Open Street Map (OSM) Guided Hyperspectral SLIC

In order to map the proposed method to PM-LDA, the image is segmented to superpixels first. A superpixel is a spatial-continuous region. The used superpixel generation algorithm is hyperspectral SLIC [44], an extension of SLIC algorithm [45].

In original SLIC algorithm, the test images used in [45] are color images in the CIELAB color space. *Lab* features are selected where *L* represents lightness, *a* and *b* represent color-opponent dimensions [46]. The SLIC algorithm can be viewed as an adaption of *k*-means to generate superpixels. The search space of SLIC is restricted to an area that is proportional to the size of the superpixel which dramatically reduces the complexity. To be more specific, the size of the desired superpixel is assumed to be  $S \times S$ , then the search space looking for the similar surrounding pixels should have a size of  $2S \times 2S$ .

[44] suggests a novel approach based on SLIC by using the hyperspectral data signatures as the new feature vectors instead of the *Lab* features in original SLIC algorithm. The *Lab* features are replaced by the spectral information. A hyperspectral image is denoted as  $n$  pixel vectors  $\mathbf{X} = \mathbf{x}_i \in \mathbf{R}^B, i = 1, 2, \dots, n$ . Here the number of spectral bands is  $B$ . The spectral distance of two pixels is defined as:

$$d_{spectral} = \sum_{\lambda=1}^B \|\mathbf{x}_i(\lambda) - \mathbf{x}_j(\lambda)\|_2^2. \quad (2.20)$$

The coordinates  $(a_i, b_i)$  and  $(a_j, b_j)$  are the spatial locations for  $\mathbf{x}_i$  and  $\mathbf{x}_j$ . The spatial distance of these two pixels is defined as:

---

<sup>7</sup>The pseudo code for this part comes from the article [14].

$$d_{spatial} = \sqrt{(a_i - a_j)^2 + (b_i - b_j)^2}. \quad (2.21)$$

The combination function of these two distances is defined as:

$$d_{i,j} = d_{spectral} + \frac{m}{S} d_{spatial}, \quad (2.22)$$

where  $S$  is the grid intervals and  $m$  is the scaling factor which is used to control the importance of spatial distance.

The hyperspectral SLIC algorithm is shown in Algorithm 5 <sup>8</sup>.

OSM data is exported online for a particular region by setting the latitude and longitude ranges. The exported OSM data contain the *node* information associated with the geo-coordinates and *way* information showing how *nodes* are connected. Then affine transformation is performed to map the geographical locations to the coordinates in the hyperspectral image. With the connection information, we can connect specific nodes to form a polygon to represent the object. The superpixels, from hyperspectral SLIC algorithm, are grouped as long as they have overlaps with the polygons from OSM data, and given the same tag of the polygon they overlapped.

---

**Algorithm 5** Hyperspectral SLIC superpixel segmentation [45]

---

- 1: Initialize cluster centers  $C_k$  by sampling pixels at regular grid steps  $S$ .
  - 2: Perturb cluster centers in an  $n \times n$  neighborhood, to the lowest gradient position.
  - 3: repeat
  - 4: **for** each cluster enter  $C_K$  do **do**
  - 5:   Assign the pixels with the shortest distance from a  $2S \times 2S$  square neighborhood around the cluster center according to the combination distance function
  - 6: **end for**
  - 7: compute new cluster centers
  - 8: until stopping criterion reaches.
- 

<sup>8</sup>The pseudo code for this part comes from the article [44].

# Chapter 3

## Proposed Method

In this section, the proposed methods for instance influence estimation of eFUMI and semi-supervised PM-LDA based hyperspectral unmixing are discussed.

### 3.1 Instance Influence Estimation for Hyperspectral Target Signature Characterization using eFUMI

Given the eFUMI algorithm outlined in the previous section, our goal in this paper is to estimate the *influence* of each pixel or an image region on the result of the eFUMI algorithm. Specifically, our goal is to estimate which pixels would cause the largest change in the target spectral signature estimated by eFUMI if their labels were switched. By identifying which pixels would cause the largest change in the eFUMI algorithm, then one could guide an analyst to focus on (re-)labeling only those pixels that would have a significant influence on the eFUMI result to reduce the analyst's

labeling load.

Given the goal of determining which pixels cause the largest change in the estimate target spectral signature, we define influence of a pixel as:

$$I_i = \|\mathbf{e}_{t_i} - \mathbf{e}_t\|^2 \quad (3.1)$$

where  $\mathbf{e}_t$  is the estimated target endmember without changing any label values (i.e., the target estimation result using an initial set of bag-level labels) and  $\mathbf{e}_{t_i}$  is the estimated target endmember after changing the label of point  $\mathbf{x}_i$  (i.e., switching the label to be “negative” to “positive” or vice versa. One could certainly determine influence of each pixel by laboriously switching the label of each pixel in series, re-running *e*FUMI for the modified label set, and computing the resulting change in the estimated target signature. However, this would be extremely time consuming and, thus, infeasible for the development of an interactive method for hyperspectral unmixing analysis.

Therefore, this thesis proposes two fast methods to approximate the relative influence of a single point or image region (e.g., a superpixel). Both of the proposed methods extremely reduce the time needed to estimate relative influence as compared with original relabeling-rerunning process for each point/superpixel. Instead, only running *e*FUMI once is necessary. After running *e*FUMI with initial user-selected labels, the first proposed method estimates relative influence by applying the fully-constrained least-squares [18] unmixing algorithm to all data points using the endmembers estimated from the initial run. Then, we propose to use the target proportion value estimated for each point as a surrogate for the influence measure. Namely, for each



point  $\mathbf{x}_i$ , we find:

$$\mathbf{p}_i = \arg \min_{\mathbf{p}_i} \|\mathbf{x}_i - \mathbf{E}\mathbf{p}_i\|_2^2 \quad (3.2)$$

$$s.t. \quad \begin{cases} p_{ij} \geq 0 & \forall i, j \\ \sum_{j=1}^{M+1} p_{ij} & \forall i, j \end{cases} \quad (3.3)$$

where  $\mathbf{E}$  is a matrix containing the estimated endmembers,  $\mathbf{p}_i$  is the proportion vector for point  $\mathbf{x}_i$ ,  $p_{ij}$  is the proportion value of endmember  $j$ , and  $M$  is the number of non-target endmembers. The surrogate influence measure is set to  $p_{it}$ , the proportion value estimated for the target endmember for the  $i^{th}$  data point. This method identifies the data points with large target portion. Thus, changing the label of these data points would have a large influence of the resulting estimated target endmember.

For the second proposed method, given the estimated endmembers and proportions for each point, the residual error is computed:

$$r_i = \|\mathbf{x}_i - \mathbf{E}\mathbf{p}_i\|_2^2 \quad (3.4)$$

where  $\mathbf{E}$  is the estimated endmembers,  $\mathbf{p}_i$  is the estimated proportion for point  $\mathbf{x}_i$ . This method identifies the data points that are not well represented by the estimated endmembers and, thus, lie outside the convex hull defined by the endmembers.

## 3.2 Hyperspectral Unmixing with Endmember Variability using PM-LDA

PM-LDA is capable of addressing the issue of spectral variability in hyperspectral unmixing because it is a NCM. In the following subsection, the relationship between PM-LDA and NCM is discussed to show that PM-LDA is actually a NCM method by properly associating each component in the graphical model of PM-LDA to each component in hyperspectral unmixing. Then, semi-supervised learning is discussed to be incorporated with PM-LDA.

### 3.2.1 PM-LDA Relationship with Normal Compositional Model

PM-LDA has been shown effective for NCM based hyperspectral unmixing and end-member estimation. The hyperspectral scene is first segmented into spatially-contiguous superpixels. Each superpixel is assumed to be a *document*. Given the superpixel segmentation and the assumption that the topic distributions are Gaussian (to assume the NCM), then the parameters of the PM-LDA model can be directly related to parameters of interest in the NCM unmixing model. Namely, the  $K$  topic distributions governed by parameters  $\beta_k = \{\mu_k, \Sigma_k\}$  correspond to the  $K$  Gaussian endmember distributions. The partial membership vector for data point  $n$  in document  $d$ ,  $\mathbf{z}_n^d$ , is the proportion vector associated with the  $n^{th}$  data point in the  $d^{th}$  superpixel. The topic proportion vectors for a document,  $\boldsymbol{\pi}^d$ , correspond to the average proportion vector for a superpixel with the mixing level  $s^d$  corresponding to how much each proportion vector in the document is likely to vary from the average proportion vector. Thus, an entire hyperspectral scene is modeled as a corpus in PM-LDA.

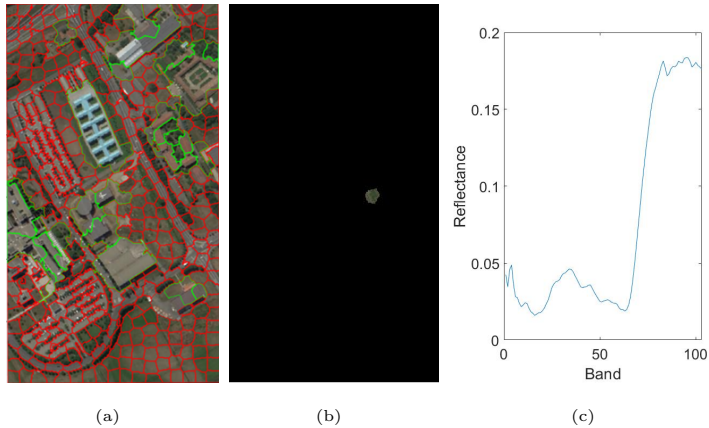


Figure 3.1: Hierarchical relationship on Pavia: (a) Hyperspectral image (corpus); (b) One superpixel (document); (c) one pixel (word).

An advantage of the use of a superpixel segmentation to define documents during unmixing allows us to leverage the expected similarity of the materials found in neighboring pixels. In other words, spectrally homogeneous neighborhoods are likely to be grouped within a superpixel. Using PM-LDA, all of the pixels in a superpixel are paired with proportion vectors drawn from the same Dirichlet distribution with a shared average proportion vector ( $\boldsymbol{\pi}^d$ ) and the variance around that mean is governed by  $s^d$ . Larger  $s^d$  values correspond to more spatially and spectrally homogeneous superpixels.

### 3.2.2 Semi-supervised PM-LDA

The proposed semi-supervised partial membership Latent Dirichlet Allocation model comprises three separate steps: 1) OSM-guided hyperspectral SLIC; 2) endmember label assignment for each superpixel; 3) incorporation of endmember labels into PM-LDA for inference and parameter estimation.

### Endmember label assignment for each superpixel

After generating superpixels using OSM-guided hyperspectral SLIC in the first phase, we need to acquire the endmember labels for semi-supervised process in the second stage. Given the superpixels for a hyperspectral image, the supervision is imposed by assigning an endmember label vector,  $\tau_j$ , for each superpixel  $j$ . To be specific, superpixel  $j$  may contain endmember  $i$  if  $\tau_{ij} = 1$ ; while superpixel  $j$  doesn't contain endmember  $i$  if  $\tau_{ij} = 0$ . At the beginning, the expert determines the number of endmembers in the hyperspectral image. For example, in University of Pavia dataset, the number of endmembers is set as 6 and they are denoted as  $e_1, e_2, e_3, e_4, e_5, e_6$ . Then, the labeling process is divided into two parts.

Based on the degree of dispersion, the expert decides which endmembers need to be labeled (supervised) and which endmembers do not (unsupervised). Endmembers that are compact in spatial extent and found existing in a few superpixels may be easily labeled. This selection not only reduces the labeling time for the expert, but also increases the efficiency of supervision because the more superpixels containing a particular endmember, the less informative it is to apply supervision on that endmember. For example, the endmember  $e_2$  is a building material (which can be considered as red roof after endmember estimation), existing in several superpixels. It is easy for the expert to label the superpixels containing  $e_2$  as  $\tau_{2j} = 1$  and all other superpixels as  $\tau_{2j} = 0$ .

On the contrary, the endmembers may be better treated as unsupervised if the pixels containing them are widely distributed over a large number of superpixels. For instance, if there are a large number of superpixels containing endmembers  $e_3, e_4, e_5, e_6$  (soil, asphalt, shadow, vegetation), it would be time consuming and less efficient

to label them. Thus, we suggest that these endmembers are labeled as 1 for all superpixels. However, one can still label these endmembers to further improve results. Consider the extreme case, if the endmember label matrix  $\tau_{ij} = \mathbf{1}_{\mathbf{KC}}$ , the supervised PM-LDA degrades to standard unsupervised PM-LDA.

### **Incorporation of endmember label into PM-LDA for inference and parameter estimation**

Under standard PM-LDA, topic proportion for each superpixel is drawn from a Dirichlet distribution with hyperparameter  $\alpha$ . To impose supervision, it is proposed to draw a topic proportion from the product of the original distribution and an endmember label vector, followed by normalization for each pixel. By applying the endmember label, the candidate topic proportion vector is restricted to have nonzero entries on the supervised endmembers. The change of target proportion impacts the candidate topic membership (proportions) for each pixel because topic membership is drawn from the Dirichlet distribution of which the parameter is the product of the scaling factor and topic proportion. Therefore, the topic membership is also confined within supervised endmembers for the superpixel it belongs to. To be more specific, the topic proportion sampling step (line 3, Algorithm 4) is changed to  $\boldsymbol{\pi}^\dagger \sim \tau^d \text{Dir}(\boldsymbol{\alpha})$ , where  $\tau^d$  denotes the endmember label for the  $d^{\text{th}}$  superpixel. We suggest randomly sampling a portion of data points for the endmember distribution parameter estimation steps (line 9-12, Algorithm 4) to significantly reduce the running time with sacrificing a little estimation accuracy by choosing a reasonable sampling rate  $\epsilon$ .

# Chapter 4

## Experiments

The proposed methods are further evaluated using two real hyperspectral images. Both methods demonstrate their effectiveness either on influence estimation or hyperspectral unmixing with spectral variability.

### 4.1 Data Set Description

The proposed instance influence estimation methods and hyperspectral unmixing using semi-supervised PM-LDA were both tested on two hyperspectral data sets: *University of Pavia* and *MUUFL Gulfport*. Here are the description of the two data sets.

#### **MUUFL Gulfport Data Set**

The first hyperspectral data set used in experiments in this paper was acquired by the CASI-1500 hyperspectral imager over the campus of the University of South-

ern Mississippi-Gulfpark in Long Beach, Mississippi in November 2010. The scene contains  $325 \times 337$  pixels and consists of 72 bands covering a wavelength range of 375 to 1050 nm with a 10 nm bandwidth. The spatial resolution is 1 m. The RGB image of the Gulfport hyperspectral image is shown in Figure 4.1 (a).

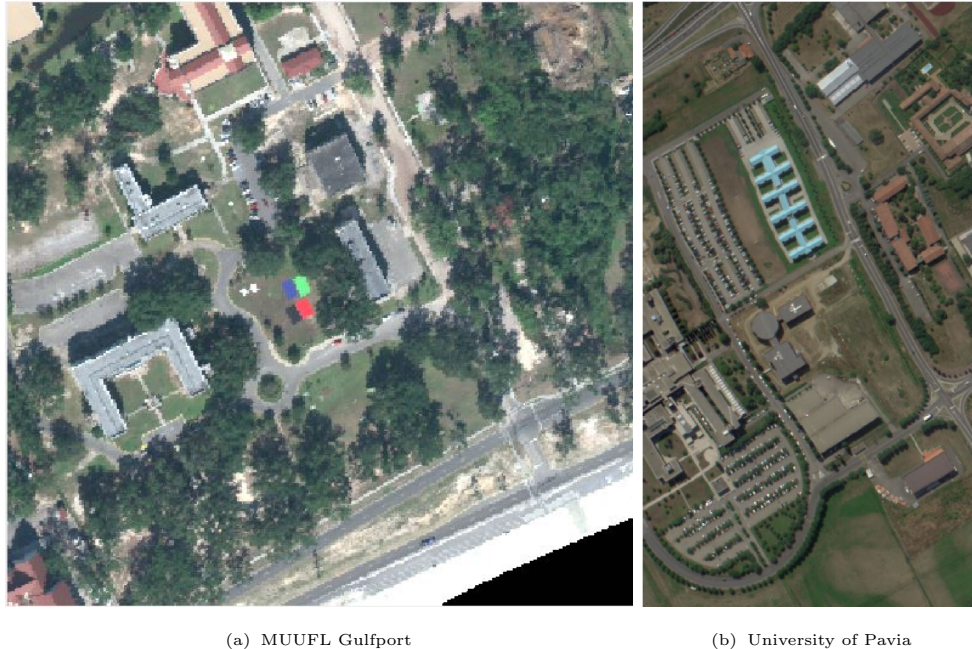


Figure 4.1: RGB image of MUUFL Gulfport and University of Pavia data sets.

### **Pavia University Data Set**

The second data set was collected by the Reflective Optics System Imaging Spectrometer (ROSIS) around the Engineering School at the University of Pavia, Italy in July 2002. The flight was operated by the Deutschen Zentrum for Luftund Raumfahrt (DLR, the German Aerospace Agency) in the framework of the HySens project, managed and sponsored by the European Union [47]. The ROSIS generates 115 bands covering 430 to 860 nm with 4nm bandwidth. The scene contains  $340 \times 610$  pixels

and consists of 103 bands with the 12 noisiest bands removed. The spatial resolution is 1.3 m. We obtained this data set from the website of Computational Intelligence Group at the Basque University (UPV/EHU) [48]. The RGB image of the University of Pavia hyperspectral image is shown in Figure 4.1 (b).

## 4.2 Instance Influence Estimation Experiments

Experimental evaluations have been carried out to investigate the performance of the proposed methods.

For Gulfport data set, four different types of colored fabric cloths were placed throughout the scene as targets. The target colors (used to identify the different target types) are Brown, Dark Green, Faux Vineyard Green, and Pea Green. Each type of target varied in size from  $0.5 \times 0.5 \text{ m}^2$ ,  $1 \times 1 \text{ m}^2$ , to  $3 \times 3 \text{ m}^2$ . A 5m-by-5m halo around the center of each target was considered as the target “bag” since the accuracy of the GPS device used to collect the ground truth was 5 m. [49]

For Pavia data set, the target of interest for this data set was selected as a region of sidewalk (found around the blue painted metal sheets). The label of target bag and non-target bags were chosen to ensure that the target bag contains the target material and non-target bags contain a variety of non-target materials such as painted metal sheets, vegetation, red roof, bare soil and shadow.

### 4.2.1 Single Point Experiment

The first experiment investigated the correlation between the influence of a single data point and the proposed metrics. The eFUMI was first run given an initial set of





Figure 4.2: Labeled ground truth on (a) Gulfport. 5-by-5 red regions are the target bags (Brown) and the rest are non-target bags; (b) Pavia University. Red region is the target bag (Sidewalk) and green regions are the non-target bags.

labels (as shown in Figure 4.2). When running  $eFUMI$ , endmembers were initialized using the VCA algorithm and all proportion values were initialized as  $\frac{1}{M+1}$  where  $M$  is the number of background endmembers and  $M + 1$  is the total number of endmembers (given only one target endmember). After this initial run, both of the proposed methods to approximate relative influence were computed. Thus, the target proportion value and the residual error were computed for every data point. Then, the labels of 1000 non-target data points were flipped in series. For each data point,  $eFUMI$  was run again given the new label set and the resulting influence value was computed. After the influence value was computed for a data point, the label of that point was restored and the next label was flipped to repeat the same process such that the influence values for all 1000 points were calculated. This process is outlined in Alg. 6.

Figure 4.3 (a) and (b) display the scatter plots of log of the influence value versus the target proportion and residual error metrics, respectively, for 1000 randomly

---

**Algorithm 6** Single Point Experiment

---

- 1: Initialize  $\mathbf{E}_{init}$  via VCA,  $\mathbf{P}_{init}$  as  $\frac{1}{M+1}$
  - 2:  $\mathbf{E}_0, \mathbf{P}_0 \leftarrow e\text{FUMI}(L, \mathbf{X}, \mathbf{E}_{init}, \mathbf{P}_{init})$  (where  $L$  is the initial label set and  $\mathbf{X}$  is the data)
  - 3: Compute residual error for each data point:  $\mathbf{r} \leftarrow \|\mathbf{X} - \mathbf{E}_0\mathbf{P}_0\|_2^2$
  - 4: Obtain target proportion for each data point,  $\mathbf{p}_t$  via unmixing using (3.3)
  - 5: **for**  $i=1:\text{NumDataPoints}$  **do**
  - 6:       Flip the label  $\mathbf{l}_i$  of point  $\mathbf{x}_i$  to obtain updated label set,  $L^i$ :  $\mathbf{l}_i \leftarrow \mathbf{l}_i - 1$
  - 7:        $\mathbf{E}_i, \mathbf{P}_i \leftarrow e\text{FUMI}(L^i, \mathbf{X}, \mathbf{E}_0, \mathbf{P}_0)$
  - 8:       Compute Influence:  $\mathbf{I}_i$  using (3.1)
  - 9:       Restore the label  $\mathbf{l}_i$ :  $\mathbf{l}_i \leftarrow \mathbf{l}_i - 1$
  - 10: **end for**
  - 11: **return**
- 

selected points. The Spearman’s rank correlation coefficients for both methods are  $\rho_{p_t} = 0.3875$  and  $\rho_{r_e} = 0.1005$ . Figure 4.3 (c) and (d) display the scatter plots of log of the influence value versus the target proportion and residual error metrics, respectively, for the 1000 data points with the largest target proportion value. The Spearman’s rank correlation coefficients for both methods are  $\rho_{p_t} = 0.6434$  and  $\rho_{r_e} = 0.5674$ . As can be seen in the scatter plots, points with large influence values tend to also have large target proportion and large residual error values.

The proposed methods are also shown to be effective on the Pavia University dataset. As is shown in Figure 4.2 (b), a sub-image of Pavia University was chosen as the test dataset. The red rectangular region was marked as the target bag and green rectangular regions were identified as the non-target bags. In this example, the target material was “sidewalk”. The labels of all non-target data points were flipped sequentially, following the procedure described above. As shown in Figure 4.4 (a), there is a strong relationship between high influence value and high target proportion on this data set. Figure 4.4 (b) displays the relationship between influence value and

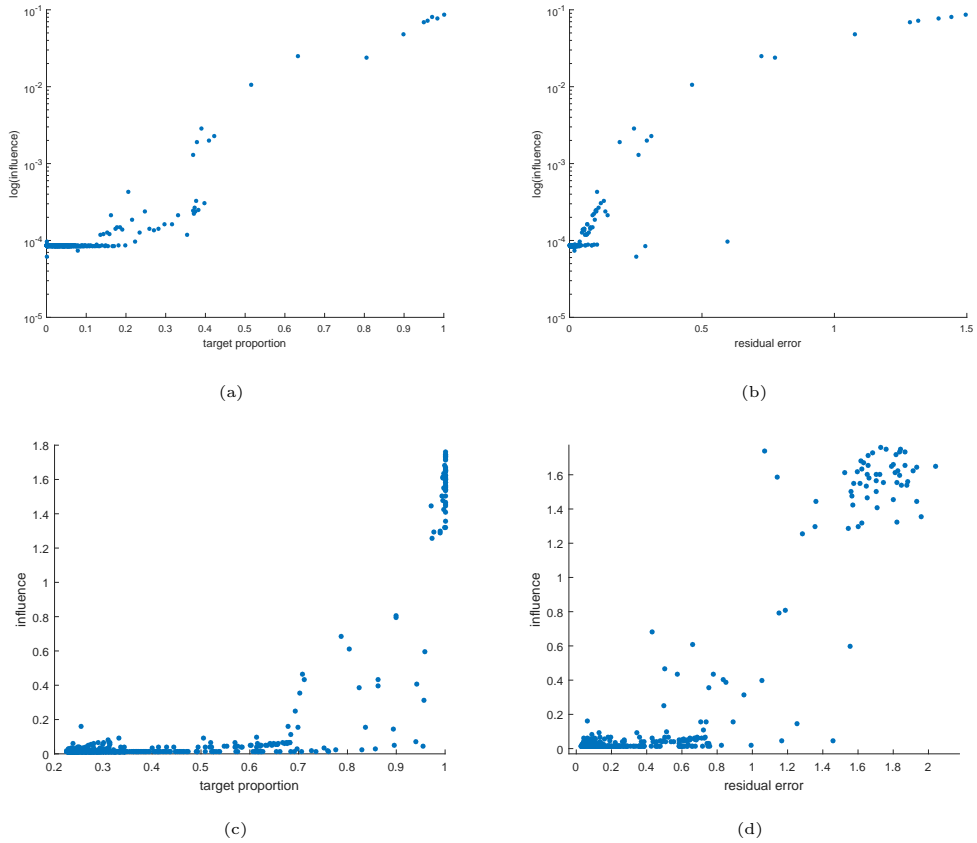


Figure 4.3: Influence caused by selecting 1000 points randomly(a,b) or with high target proportion(c,d) from Gulfport and changing labels, compared with their corresponding (a,c) target proportion, (b,d) residual error.

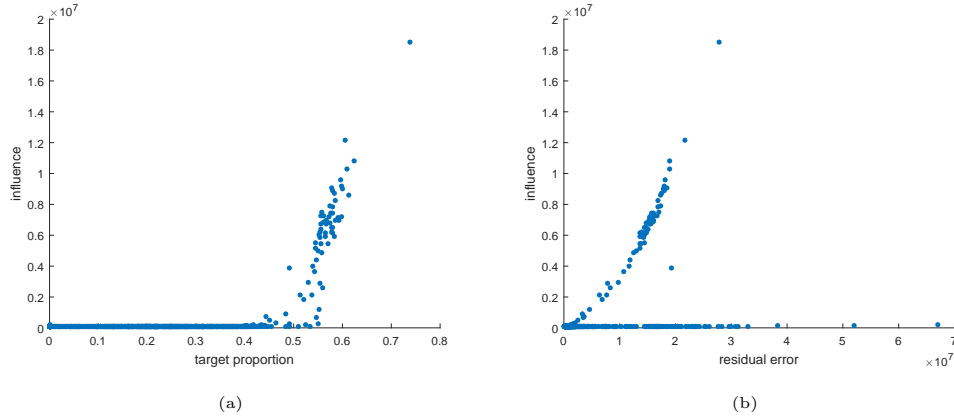


Figure 4.4: Influence caused by selecting all data points in the non-target bags with the large target proportion value from the Pavia data set and changing labels, compared with their corresponding (a) target proportion, (b) residual error.

residual error. The data points that have a target proportion above 0.4 in Figure 4.4 (a) were identified as pixels from red roof regions in the non-target bags. Further investigation shows that these red roof materials have a similar spectral signature to the sidewalk in the area. In this way, act as a “confuser” during target characterization and, thus, they are influential data points in the *e*FUMI algorithm. In this example, target portion was seen to be more effective than residual error in predicting high influence values (as can be seen with the divergence of influence value in the scatter plot among points with large residual error).

### 4.2.2 Mislabeling and Recovery Experiment

In this experiment, we investigate the improvement in target endmember estimation by correcting the labels of groups of mislabeled data points. Specifically, we compare the improvement of selecting points to relabel using the proposed methods versus random selection.

The experiment was structured as follows. First, *e*FUMI was run with a fixed set of correct initial labels. The resulting endmembers and proportion values were labeled as  $E_{true}$  and  $P_{true}$ . Then, 0.5% of the labels (500 non-target points) were randomly selected and changed to incorrect label values. *e*FUMI was run using these incorrect labels and the results were named  $E_{err}$  and  $P_{err}$ . Finally, 20% data points were selected using one of three different strategies: (1) selected 20% of the data points randomly, (2) selected 20% of the data points with largest target proportion, and (3) selected 20% of the data points with the largest residual error. Among the 20% of selected pixels, their labels were corrected if incorrect. These updated labels were then used to run *e*FUMI. These final set of estimated endmembers were named  $E_{rand}$ ,  $E_{pt}$ , and  $E_{re}$ , respectively, depending on the sample selection strategy. This process is outlined in Alg. 7.

Degrees of improvement (DoI) for the target endmember spectra for the three methods (random, target proportion and residual error) were calculated using the following:

$$DoI_k = \frac{\|e_{t_{true}} - e_{t_{err}}\|^2 - \|e_{t_{true}} - e_{t_k}\|^2}{\|e_{t_{true}} - e_{t_{err}}\|^2}, k = rand, P_t, re, \quad (4.1)$$

where  $e_{t_m}$  is the target endmember for  $e_m = true, err, rand, P_t, re$ . DoI for the three methods were computed as  $DoI_{rand} = 18.7\%$ ,  $DoI_{P_t} = 96.11\%$ ,  $DoI_{re} = 99.42\%$ , respectively, considering the brown target in the MUUFL Gulfport data set. Figure 4.5 and the resulting DoIs show that the label correction obtained using the proposed methods could better recover the estimated target signature as opposed to random selection.

The mislabeling and recovery experiment was repeated on the Pavia Dataset. 5%

---

**Algorithm 7** Mislabeling and Recovery Experiment

---

- 1: Initialize  $\mathbf{E}_{init}$  via VCA,  $\mathbf{P}_{init}$  as  $\frac{1}{M+1}$
- 2:  $\mathbf{E}_{true}, \mathbf{P}_{true} \leftarrow e\text{FUMI}(L, \mathbf{X}, \mathbf{E}_{init}, \mathbf{P}_{init})$  (where  $L$  is the initial label set and  $\mathbf{X}$  is the data)
- 3: Compute residual error for each data point:  $\mathbf{r} \leftarrow \|\mathbf{X} - \mathbf{E}_0 \mathbf{P}_0\|_2^2$
- 4: Obtain target proportion for each data point,  $\mathbf{p}_t$  via unmixing using (3.3)
- 5: Randomly select and flip  $\alpha\%$  labels obtaining  $L^{err}$
- 6:  $\mathbf{E}_{err}, \mathbf{P}_{err} \leftarrow e\text{FUMI}(L^{err}, \mathbf{X}, \mathbf{E}_{init}, \mathbf{P}_{init})$  (where  $L$  is the updated label set and  $\mathbf{X}$  is the data)
- 7: Initialize  $\mathbf{E}_0$  via VCA,  $\mathbf{P}_0$  as  $\frac{1}{M+1}$
- 8: **for**  $i=1$  to 3 **do**
- 9:     **if**  $i=1$  **then**
- 10:         Randomly select  $\beta\%$  labels and correct if wrong to obtain  $L^{rand}$
- 11:          $\mathbf{E}_{rand}, \mathbf{P}_{rand} \leftarrow e\text{FUMI}(L^{rand}, \mathbf{X}, \mathbf{E}_0, \mathbf{P}_0)$
- 12:         Compute  $DoI_{rand}$  using (4.1)
- 13:     **end if**
- 14:     **if**  $i=2$  **then**
- 15:         Select points with highest  $\beta\%$  target proportions and correct their labels if wrong to obtain  $L^{pt}$
- 16:          $\mathbf{E}_{pt}, \mathbf{P}_{pt} \leftarrow e\text{FUMI}(L^{pt}, \mathbf{X}, \mathbf{E}_0, \mathbf{P}_0)$
- 17:         Compute  $DoI_{pt}$  using (4.1)
- 18:     **end if**
- 19:     **if**  $i=3$  **then**
- 20:         Select points with highest  $\beta\%$  residual errors and correct their labels if wrong to obtain  $L^{re}$
- 21:          $\mathbf{E}_{re}, \mathbf{P}_{re} \leftarrow e\text{FUMI}(L^{re}, \mathbf{X}, \mathbf{E}_0, \mathbf{P}_0)$
- 22:         Compute  $DoI_{re}$  using (4.1)
- 23:     **end if**
- 24: **end for**
- 25: **return**

---

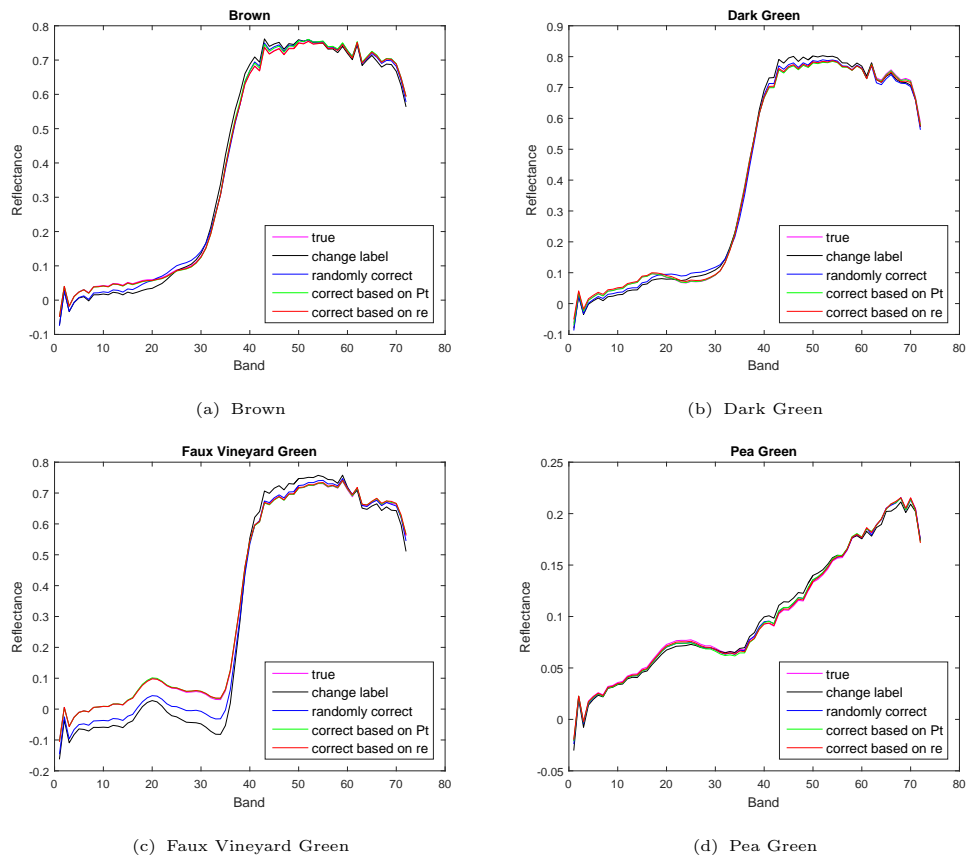


Figure 4.5: Target endmembers on Gulfport estimated by label of (1) ground truth; (2) 0.5% flipped; (3) correct 20% randomly; (4) correct 20% based on influential points of method 1; (5) correct 20% based on influential points of method 2.

of the labels (58 non-target points) was chosen as the mislabeling ratio and 10% was chosen as the correction ratio. The initial labels for *eFUMI* was the same one shown in Figure 4.2 (b). The DoI for three methods were found to be  $DoI_{rand} = 25.26\%$ ,  $DoI_{P_t} = 73.4\%$ , and  $DoI_{re} = 28.06\%$ . Similar to the single point experiment, it was found that the target proportion metric was more effective on this data set.

### 4.2.3 Superpixel Influence Experiment

A superpixel is defined as a small, spatially continuous segment in an image. In this experiment, we investigate the influence of modifying labels of superpixels (as opposed to individual pixels) in a hyperspectral data set. The influence for superpixel was investigated since each superpixel generally contains data points with similar spectral signatures (and, thus, the influence of the set of data points is likely to be similar) and because it is much easier and more intuitive to relabel superpixel regions instead of individual pixels.

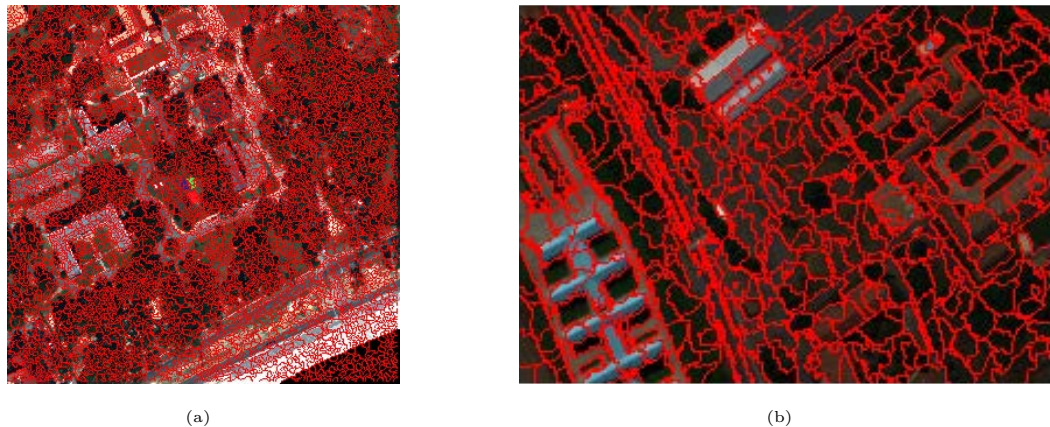


Figure 4.6: Influence of Superpixel for (a) Gulfport. 7419 segments were generated via Normalize Cut; (b) Pavia University. 744 segments were generated via Normalize Cut.



To understand the correlation between influence of superpixel and proposed approaches, the Gulfport dataset was over-segmented into about seven thousand superpixels having similar sizes using a Normalized Cut method [50]. Then, the rest of experiment was very similar with single point influence experiment with one difference which is instead of computing the target proportion of all points in each superpixel, we relied on the largest target proportion in each superpixel as the surrogate influence metric. This process is outlined in Alg. 8. Figure 4.6 (a) shows the superpixels of Gulfport dataset. Figure 4.7 illustrates the plot of log of the influence value versus the log of largest target proportion, residual error, sum of target proportion, and sum of residual error in each superpixel, respectively. As is shown in Figure 4.7, the log of the influence is most correlated with the largest target proportion in each superpixel among all proposed estimation methods.

---

**Algorithm 8** Superpixel Experiment

---

- 1: Segment input image into  $N$  superpixels
  - 2: Initialize  $\mathbf{E}_{init}$  via VCA,  $\mathbf{P}_{init}$  as  $\frac{1}{M+1}$
  - 3:  $\mathbf{E}_0, \mathbf{P}_0 \leftarrow e\text{FUMI}(L, \mathbf{X}, \mathbf{E}_{init}, \mathbf{P}_{init})$
  - 4: Compute residual error for each data point:  $\mathbf{r} \leftarrow \|\mathbf{X} - \mathbf{E}_0\mathbf{P}_0\|_2^2$ , compute  $\max(\mathbf{r})$  in each superpixel, compute  $\text{sum}(\mathbf{r})$  in each superpixel
  - 5: Obtain target proportion for each data point,  $\mathbf{p}_t$  via unmixing using (3.3), compute  $\max(\mathbf{p}_t)$  in each superpixel, compute  $\text{sum}(\mathbf{p}_t)$  in each superpixel
  - 6: **for**  $i=1:N$  **do**
  - 7: Flip all of the labels for each data point in superpixel  $\mathbf{s}_i$  to obtain  $L^i$
  - 8:  $\mathbf{E}_i, \mathbf{P}_i \leftarrow e\text{FUMI}(L^i, \mathbf{X}, \mathbf{E}_0, \mathbf{P}_0)$
  - 9: Compute Influence:  $\mathbf{I}_i$  using (3.1)
  - 10: Restore the labels to  $L$
  - 11: **end for**
  - 12: **return**
- 

To further investigate superpixel influence, the same experiment was repeated on the Pavia dataset. Nearly seven hundred segments was generated and shown in 4.6

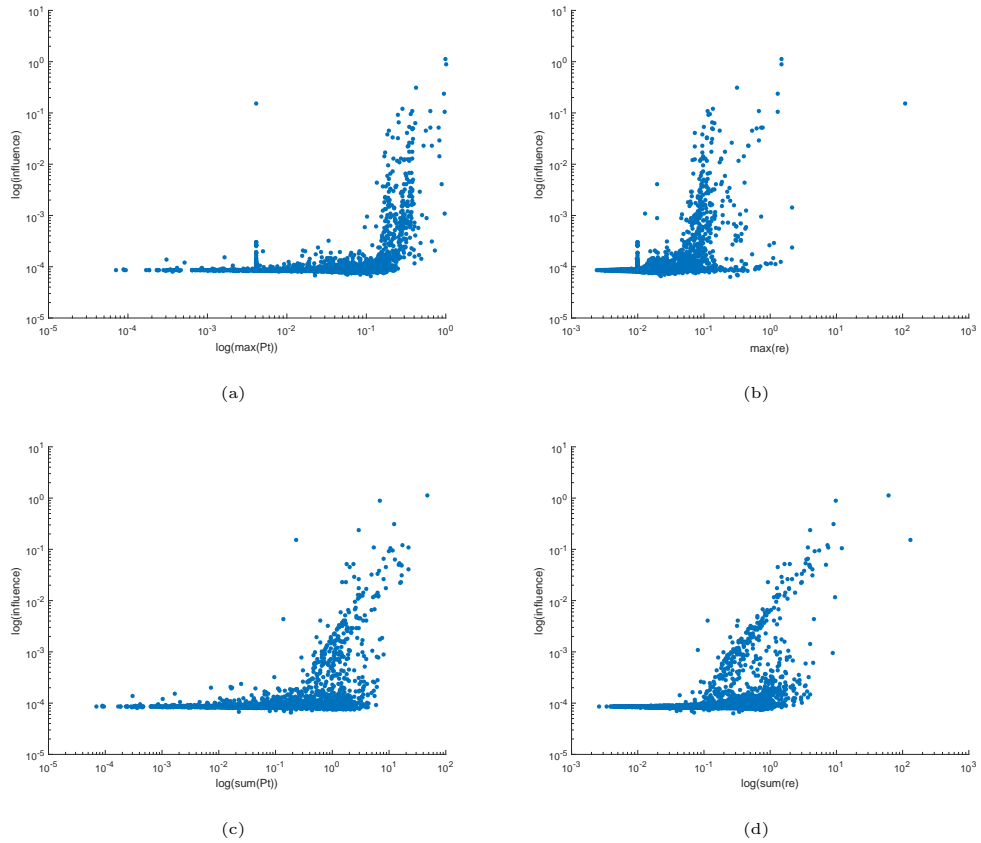


Figure 4.7: Influence of Gulfport superpixels caused by flipping the labels of each superpixel compared with (a) maximum target proportion in each superpixel, (b) maximum residual error in each superpixel, (c) sum of target proportions in each superpixel, and (d) sum of residual errors in each superpixel.

(b). The average size of Pavia segments is much larger than those in Gulfport. Figure 4.8 shows the influence value versus all of the proposed metrics for the Pavia data set. The performance of the sum of target proportions and the sum of the residual errors were found to be more effective than the maximum values. We believe this to be the case because the sizes of the superpixels on Pavia are much larger than Gulfport so that considering only one maximum point in the superpixel is not comprehensive enough as compared to metrics that take into account all target proportions or residual errors in each superpixel.

### 4.3 Hyperspectral Unmixing via Semi-supervised PM-LDA Experiments

Two real hyperspectral data sets, *University of Pavia* and *MUUFL Gulfport*, are used for hyperspectral unmixing using the proposed method. There are three experiments designed to evaluate the proposed method from different aspects. The first experiment compared supervised PM-LDA with two NCM based algorithms, PM-LDA [14] and NCM-Bayes [12], to demonstrate the improved performance on unmixing results. The second experiment adjusted the semi-supervised labels (endmember labels) from precise labels to imprecise labels, to test the robustness of the algorithm when addressing imprecise labels/tags. The third experiment showed how to combine open street map information (tags and polygons) with human labeling and how to select and handle this information to achieve different customized unmixing results, which is an advantage of incorporating semi-supervised learning.

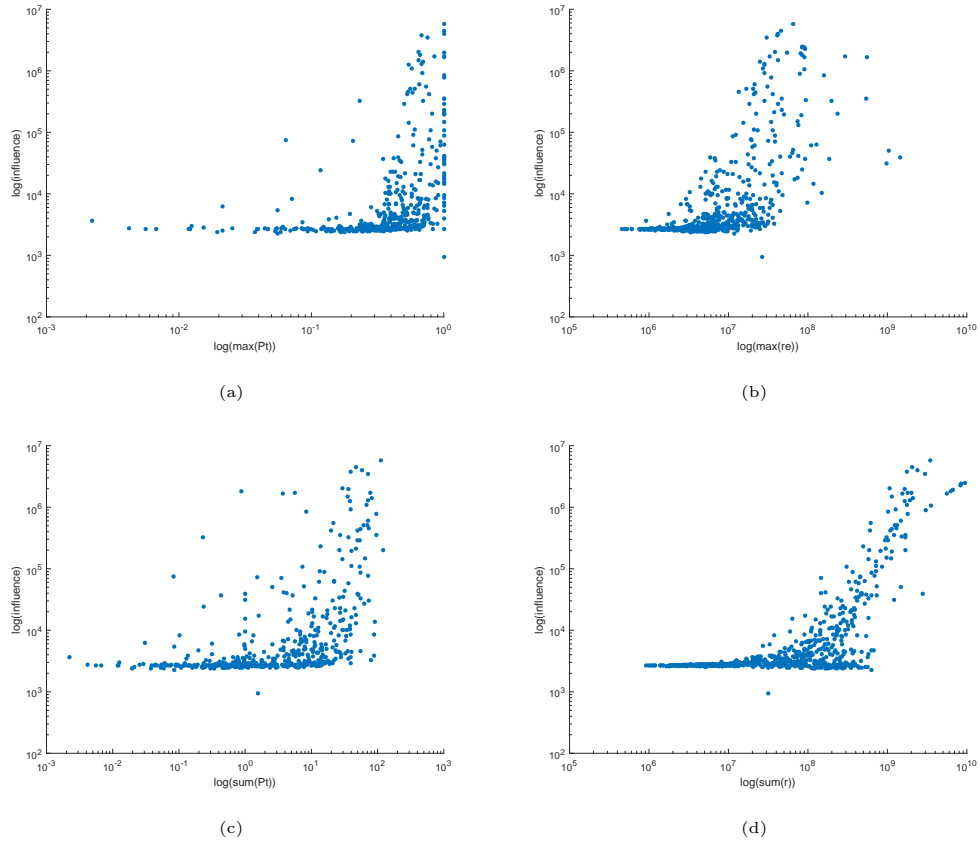


Figure 4.8: Influence of Pavia superpixels caused by flipping the labels of each superpixel compared with (a) maximum target proportion in each superpixel, (b) maximum residual error in each superpixel, (c) sum of target proportions in each superpixel, (d) sum of residual errors in each superpixel.

### 4.3.1 Hyperspectral Unmixing on Pavia

#### Comparison between semi-supervised PM-LDA with other methods

The test image is segmented into superpixels using the SLIC algorithm. The parameters for SLIC are the desired number of superpixels,  $k = 500$ , and the weighting parameter between spectral and spatial difference,  $m = 20$ . Fig 4.9(a) shows the superpixels generated by SLIC. Then, building polygons selected and pulled from Open Street Map from the region of interest are shown in Fig 4.9(b). The final superpixels are shown in Fig 4.9(c) by merging all superpixels overlapping any building polygons.

The number of endmembers is set as 6 for both the proposed method and comparison methods, which is commonly believed to be a reasonable number of endmembers in Pavia data set. As a pre-processing step, all the pixel signatures in the data set are normalized to have unit length as the input for all methods. The initialization step for semi-supervised PM-LDA and PM-LDA uses the Vertex Component Analysis (VCA), which is the endmember extraction method applied in NCM-Bayes. The parameter setting for all methods are manually tuned as follows to yield the best performance. (Note that the tuned parameters are necessarily fixed to be these values as below. Small skews for the parameters have little impact expect for  $K$  since the change of  $K$  may result in a mixture or decomposition of two materials): For semi-supervised PM-LDA,  $K = 6$ ,  $\lambda = 1$ ,  $\alpha = 0.3$ ,  $\epsilon = 5\%$  and  $T = 200$ . Blue roof and red roof are selected to be supervised in this study. To be more specific, the superpixels overlapping with the 3 blue roof or red roof polygons in Fig 4.9(b) are given precise labels. For instance, superpixels in blue roof regions are given labels  $\tau = [101111]$  while superpixels in red roof regions are given labels  $\tau = [011111]$ , where '1' in the

first entry in  $\tau$  denotes the possible existence of blue roof and '1' in the second entry denotes the possible existence of red roof, while '1' in the other four entries denotes the possible existence of the other four endmembers. All other pixels in the scene are given  $\tau = [001111]$ . The reason for only selecting partial information about polygons (no grey building or road polygons used) for semi-supervised learning is discussed in the third experiment. For NCM-Bayes, the Markov chain length was set to 250, the length of the burn-in period was set to 1000,  $\delta = 0.001$ , and the initial endmember variance was set to 0.001. For PM-LDA,  $K = 6$ ,  $\lambda = 1$ ,  $\alpha = 0.3$  and  $T = 200$ .

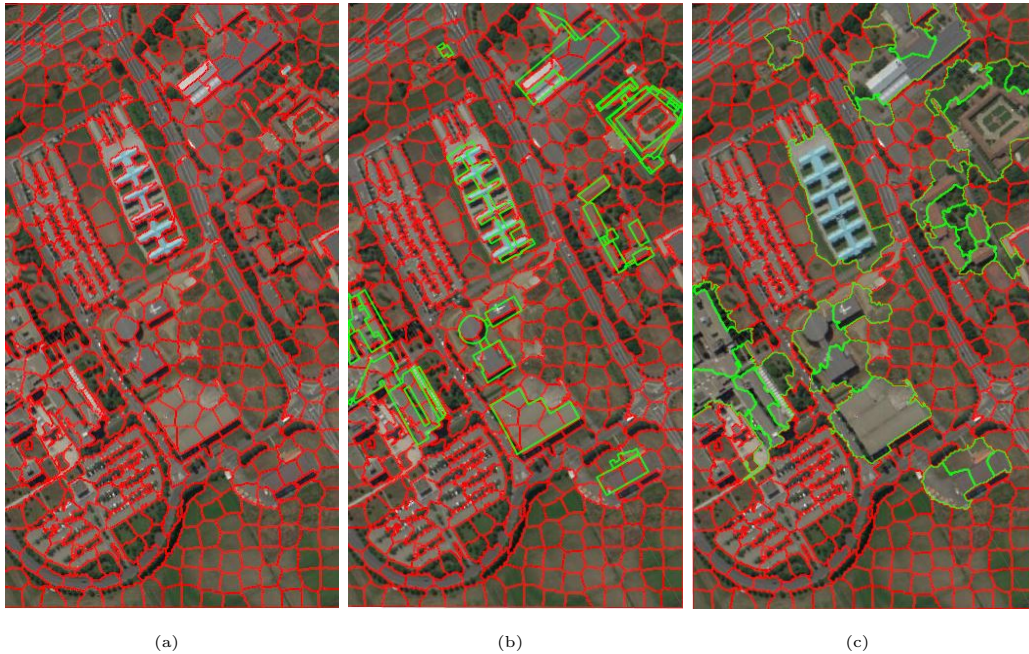


Figure 4.9: Superpixels on Pavia: (a) superpixels by SLIC; (b) red: superpixels by SLIC, green: open street map; (c) merge of superpixels and open street map.

For experimental results, all six endmember distribution parameters including endmember means and covariance matrices are learned. These estimated endmember means are shown in Fig 4.10. From these estimated endmember signatures, we

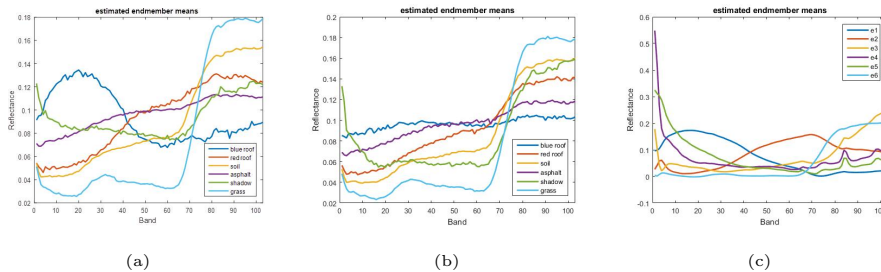


Figure 4.10: Endmember means on Pavia estimated by (a) semi-supervised PM-LDA, (b) PM-LDA, (c) VCA.

see that only signatures learned by supervised PM-LDA are close to the underlying ground truth of the Pavia data set. The major issue for the estimated endmember means using PM-LDA in Figure 4.10 (b) is that the signature of blue roof is incorrect. The estimated endmember means using VCA are worse in Figure 4.10 (c), for example, the signature of grass is not estimated. The estimated proportion maps in Fig 4.11 further illustrate the performance of the different methods. All proportion maps are permuted and aligned vertically according to the same major materials on each map so that each vertical map could be the same and compared. The 6 desired endmembers are blue roof, red roof, soil, asphalt, shadow and grass. Proportion maps estimated by supervised PM-LDA in the first row of Fig 4.11 are found to be smooth and the estimated proportion values are high for corresponding pixels dominated by single desired material and low for other materials. In contrast, PM-LDA is not able to separate blue roof from some other buildings and roads consisting of different materials. For NCM-Bayes, at most, three endmember proportions are estimated with a large error.

For quantitative evaluation, two evaluation metrics are suggested, entropy of proportion maps and data likelihood given estimation results. The proportion entropy

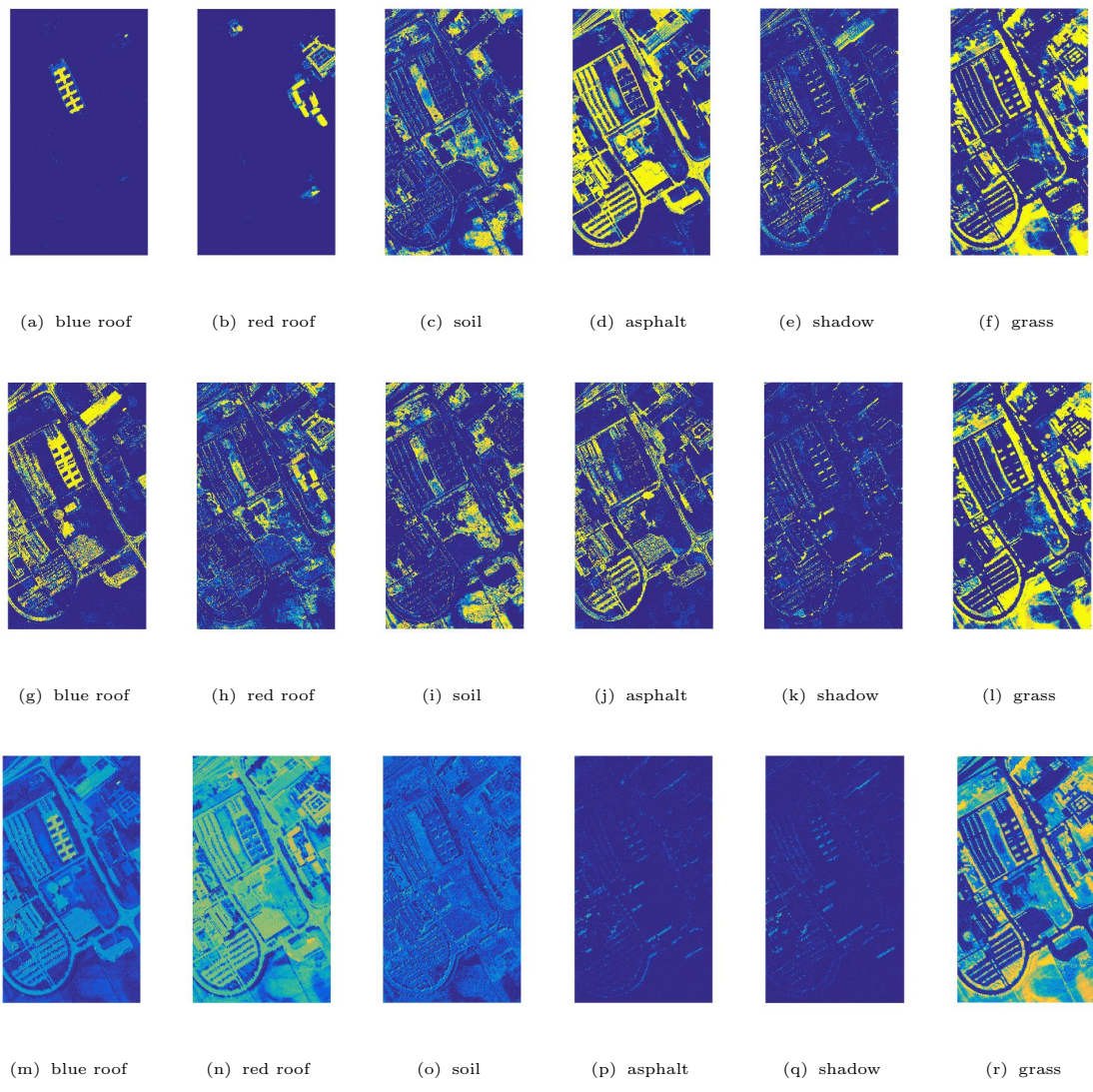


Figure 4.11: Estimated proportion maps on Pavia using (a)-(f) semi-supervised PM-LDA, (g)-(i) PM-LDA, (m)-(r) NCM-Bayes.



is defined in (4.2).

$$H(\mathbf{P}) = - \sum_{n=1}^N \sum_{k=1}^K p_{nk} \ln p_{nk} \quad (4.2)$$

where  $p_{nk}$  is the proportion value for the  $n^{th}$  pixel and  $k^{th}$  endmember,  $N$  is the number of pixels and  $K$  is the number of endmembers. The motivation for applying proportion entropy is that in real hyperspectral images, there are usually a small number of the available endmembers present in each pixel. Therefore, accurate proportion values for each pixel yield low proportion entropy indicating that only a few endmembers are present in each pixel. The second evaluation metric is the NCM log-likelihood over all pixels in the data set. The NCM-likelihood provides a measure of the overall fit between the hyperspectral data points and endmember distributions under the NCM model. The metric is indicated in (4.3).

$$f(\mathbf{X}|\mathbf{E}, \mathbf{P}, \mathbf{\Sigma}) = \sum_{n=1}^N \ln \mathcal{N} \left( \mathbf{x}_n \left| \sum_{k=1}^K p_{nk} \mathbf{e}_k, \sum_{k=1}^K p_{nk}^2 \mathbf{\Sigma}_k \right. \right) \quad (4.3)$$

The quantitative evaluation using two metrics are shown in Table 4.1 and Table 4.2, respectively.

Dataset	NCM-Bayes	PM-LDA	sPM-LDA
Pavia	2.23e5	8.81e4	<b>8.39e4</b>

Table 4.1: Overall proportion map entropy for three methods on Pavia.

Dataset	NCM-Bayes	PM-LDA	sPM-LDA
Pavia	8.97e6	<b>6.68e7</b>	6.53e7

Table 4.2: Overall log-likelihood for three methods on Pavia.

From the quantitative evaluation results, the lowest overall proportion map entropy is achieved by semi-supervised PM-LDA. It indicates the feature of semi-supervised PM-LDA. Semi-supervised learning restricts the candidate endmembers in the supervised regions to make the proportion vector sparse, which eventually leads to a low proportion map entropy. However, the highest log-likelihood is achieved by PM-LDA, instead of semi-supervised PM-LDA. The underlying reason is that sometimes a pixel signature may be better approximated with more endmembers, which is the situation for PM-LDA. In other word, there is no restriction on the candidate endmembers for each pixel and all pixels in PM-LDA are represented by all endmembers. However, the usage of too many endmembers to represent a pixel signature in standard PM-LDA may not reflect the real endmembers in the pixel and the estimated proportion vector may not be close to the underlying proportion vector. In other words, the pixel signature is over-fit if represented by all endmembers.

### **Robustness analysis using imprecise labels**

In the first experiment, the precise labels,  $\tau = [101111]$ , are given to indicate no red roof endmember exists for the superpixels in blue roof region by setting the second entry of  $\tau$  to '0'. The way to provide precise labels for superpixels in red roof regions is similar. In this experiment, the imprecise labels are given to test the algorithm performance and limitation. For imprecise labels, the superpixels in blue roof and red roof regions are all given the same label,  $\tau = [111111]$ , which relaxes the restriction for candidate endmembers in blue and red roof regions. All other pixels in non-blue-roof, non-red-roof regions are still labeled as  $\tau = [001111]$ . All other parameters are kept the same as in the first experiment.

For the experimental results, the overall estimated endmember means and pro-

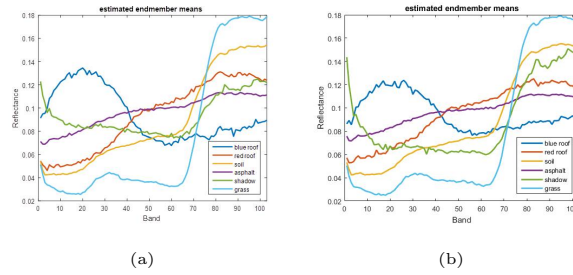


Figure 4.12: Endmember means estimated on Pavia by (a) semi-supervised PM-LDA with precise labels, (b) semi-supervised PM-LDA with imprecise labels.

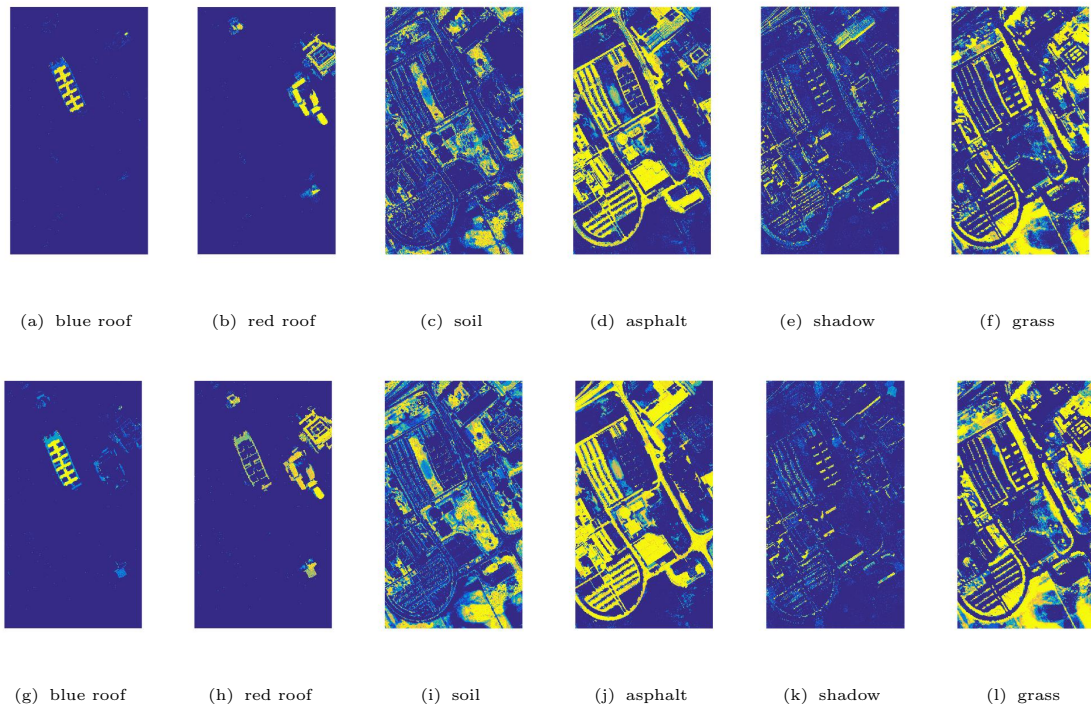


Figure 4.13: Estimated proportion maps on Pavia using (a)-(f) semi-supervised PM-LDA with precise labels, (g)-(i) semi-supervised PM-LDA with imprecise labels.

portion maps by semi-supervised PM-LDA with imprecise labels are slightly worse compared with those of semi-supervised PM-LDA with precise labels. For instance, the estimated signatures are worse than those with precise labels, especially the magnitude of blue roof signature shape in the experiment with imprecise labels is lower than the underlying magnitude, as shown in Fig 4.12. For proportion maps, the maps of endmembers which are unsupervised have little change. For the proportion maps of blue roof and red roof, even with imprecise labels, the two different endmembers are still separated. However, the pixels of sidewalk around blue roof which are believed to be asphalt are wrongly represented by blue roof and red roof. This error comes from imprecise labels because without the endmember constraints between red and blue roof endmembers, the endmember signatures of blue roof and red roof are not well estimated. The sidewalk is better represented by the linear combination of these two endmembers than the asphalt signature. Therefore, from the experiment on Pavia, if the semi-supervised learning are provided with imprecise labels, the desired endmembers may still be learned and separate from each other but not as good as the results with precise labels. The quantitative results, both the entropy and likelihood in Table 4.3, show that precise labels yield better unmixing performance than imprecise labels.

Labels	Precise	Imprecise
Entropy	<b>8.39e4</b>	8.81e4
Likelihood	<b>6.53e7</b>	6.46e7

Table 4.3: Overall proportion map entropy and log-likelihood for semi-supervised PM-LDA with precise and imprecise labels on Pavia.

### Open Street Map polygon selection and interactive human selection

Theoretically, the semi-supervised superpixels can be selected entirely by an expert

and labeled. In the thesis, open street map polygons are utilized by incorporating them into semi-supervised learning to assist the human selection and labeling of the superpixels. For example, in the second experiment, the semi-supervised superpixels are directly selected by finding the superpixels overlapping with red roof and blue roof building polygons. Since only the tag 'building' is provided for all possible building structures in open street map, human interaction is necessary to provide more precise labels by indicating which polygon is blue roof building and which polygon is red roof building, and to determine which polygon in the building area should not be selected as the semi-supervised labels, for example, grey building. The grey building should not be selected because its signature is very similar to asphalt in the 'road' polygons. Thus, it's unreasonable to assign different labels for both grey building and road but it's reasonable to select them as unsupervised superpixels. (Experimental results regarding grey roof as an individual endmember are discussed in this part.) This shows one drawback when using open street map. The open street map polygons and their associated tags are most likely to be purposed-based, such as building or road, instead of material-based in hyperspectral unmixing. Therefore, when exploiting the advantage of using open street map polygons to select superpixels, human interactions are necessary to ensure the semi-supervised superpixels and their associated labels are reasonable. Sometimes, human interactive selection is a must if a region of interest, that is to be semi-supervised, is not provided with polygons in open street map. An example of this is the beach sand area in Gulfport data set which is discussed in the next subsection.

The experimental setup is based on the previous experiment which uses the semi-supervised PM-LDA with imprecise labels. However, it adds the grey building to the

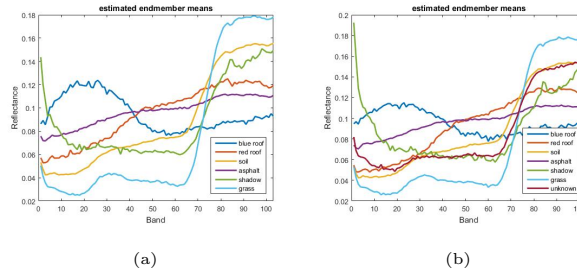


Figure 4.14: Endmember means on Pavia estimated by (a) semi-supervised PM-LDA without grey roof, (b) semi-supervised PM-LDA with grey roof.

semi-supervised learning process. The number of endmembers,  $K$ , is set as 7. The labels for superpixels in blue roof, red roof, or grey roof regions are all  $\tau = [1111111]$ , where '1' in the first three entries of the  $\tau$  represents the existence of blue roof, red roof and grey roof respectively. For non-building area superpixels, they are labeled as  $\tau = [0001111]$ .

As is shown in proportion maps in Fig 4.15, even though the grey roof is assigned with the individual labels, the grey roof is not estimated as an individual endmember as expected. However, it's interesting to note that the grey roof regions show high proportions of several different endmembers other than the desired asphalt. For instance, some pixels in grey roof regions have high proportions of blue roof or red roof. The main point of this experiment is that the labels are more imprecise compared with second experiment, resulting in worse estimated endmember means. So some pixels in this region are better represented by a linear combination of several other endmembers than just by asphalt. The quantitative results, both the entropy and likelihood, show that unmixing performance suffers when regarding grey roof as an individual endmember.

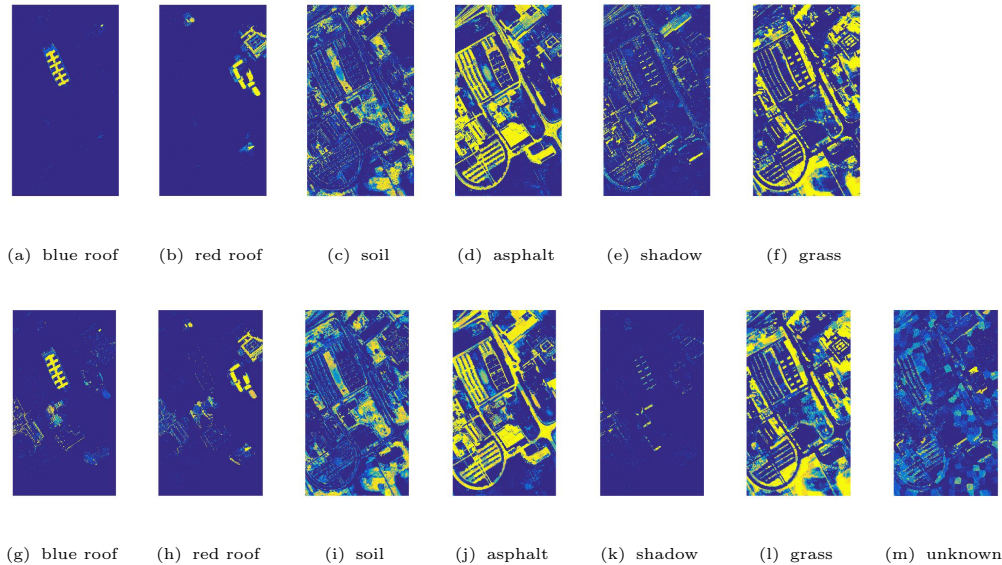


Figure 4.15: Estimated proportion maps on Pavia using (a)-(f) semi-supervised PM-LDA without grey roof, (g)-(i) semi-supervised PM-LDA with grey roof.

Labels	Imprecise	Imprecise&grey roof
Entropy	<b>8.81e4</b>	9.67e4
Likelihood	<b>6.46e7</b>	6.26e7

Table 4.4: Overall proportion map entropy and log-likelihood for semi-supervised PM-LDA with or without labeling grey roof on Pavia.

### Semi-supervised PM-LDA with different initializations

In these experiments, the default initialization setup for PM-LDA and proposed semi-supervised PM-LDA algorithms is VCA. However, different initializations can be applied resulting in the similar unmixing performance. Thus, two initialization approaches are also tested in this section and compared with the original results using VCA.

The unmixing results initialized by  $k$ -means and random selection show similar estimated endmember means and proportion maps, except for shadow, shown in Figure 4.16 and 4.17. Among these three initialization approaches, VCA is the only

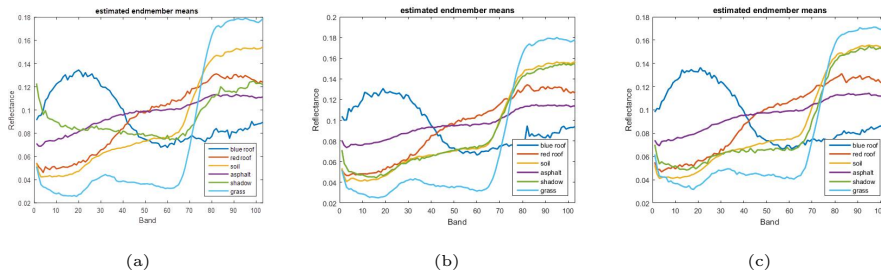


Figure 4.16: Endmember means on Pavia estimated by semi-supervised PM-LDA using (a) VCA, (b)  $k$ -means, (c) random selection.

one that is capable of estimating an accurate endmember mean signature of shadow. This is because shadow is a challenge endmember having a large covariance and large noise range. Therefore, a good initialization approach for example, VCA, can provide a good approximated shadow signature at the beginning so that the algorithm could converge fast to the underlying shadow distribution. On the contrary, some initialization approaches can't generate a good endmember mean for the challenge endmember, which may result in the longer converging time or even failing to estimate. The data likelihood results show that algorithm initialized with VCA can achieve the highest data likelihood among the three initialization methods. Random selection initialization yields the lowest proportion entropy since in this case the algorithm learns some other materials with very binary proportion values instead of the underlying shadow. In conclusion, semi-supervised PM-LDA is not sensitive to the endmembers with relatively well-separated clusters and small covariances, but sensitive to the endmembers with ill-separated clusters and large covariances (i.e. the challenging endmembers). A good initialization is also an essential step to ensure the good unmixing results for all underlying endmember distributions.



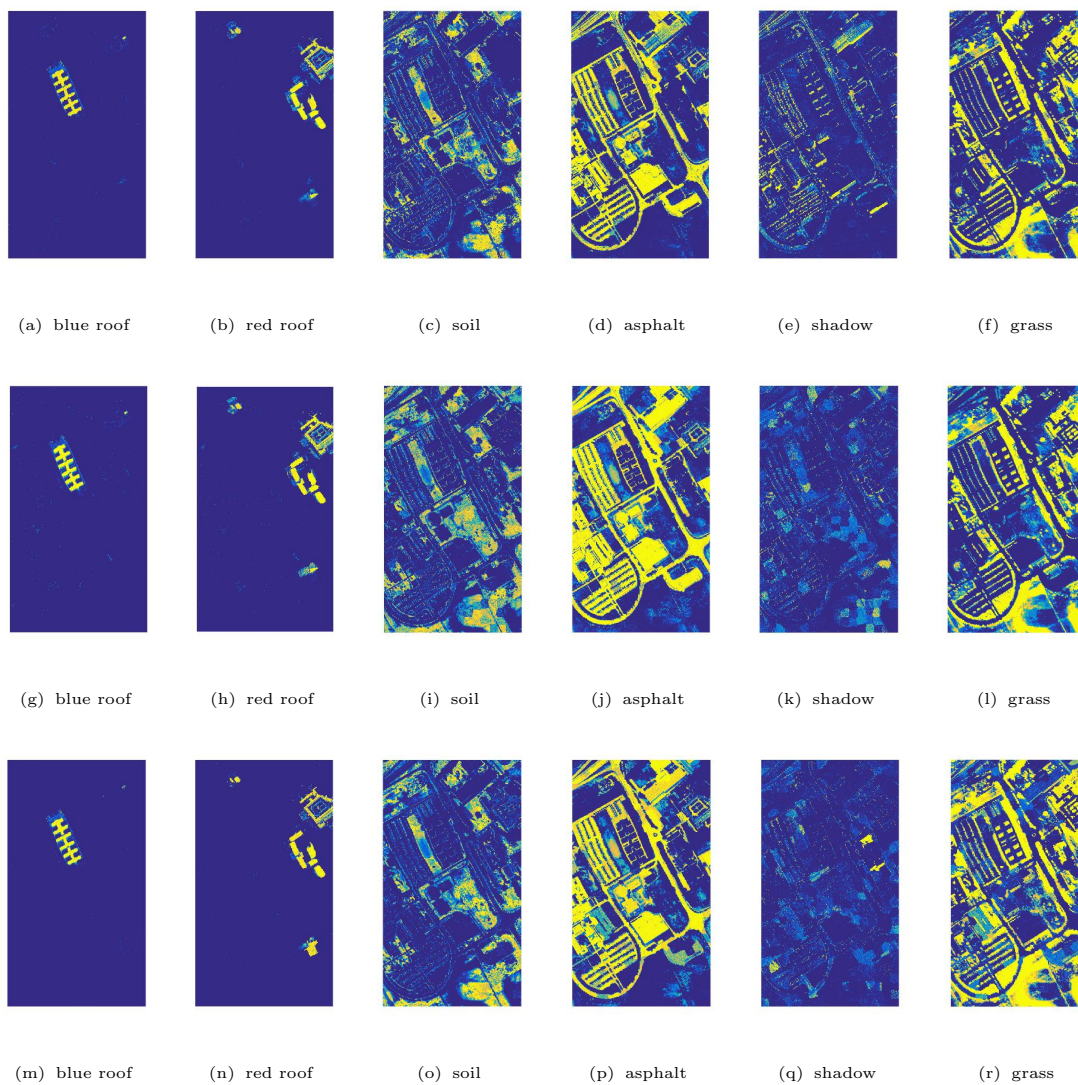


Figure 4.17: Estimated proportion maps on Pavia using semi-supervised PM-LDA initialized by (a)-(f) VCA, (g)-(i) *k*-means, (m)-(r) random selection.

Initialization	VCA	$k$ -means	Random
Entropy	8.39e4	8.55e4	<b>7.94e4</b>
Likelihood	<b>6.53e7</b>	6.39e7	5.88e7

Table 4.5: Overall proportion map entropy and log-likelihood for semi-supervised PM-LDA using different initializations on Pavia.

### 4.3.2 Hyperspectral Unmixing on Gulfport

#### Comparsion between semi-supervised PM-LDA with other methods

The test image is also segmented into superpixels using the SLIC algorithm as is shown in Fig 4.18 (a). The parameters for SLIC are the desired number of superpixels,  $k = 500$ , and the weighting parameter between spetral and spatial difference,  $m = 20$ . The number of endmembers is set as 7 for both proposed method and comparison methods. As a pre-processing step, all the pixel signatures in the data set are normalized to have unit length as the input for all methods. The initialization step for semi-supervised PM-LDA and PM-LDA is Vertex Component Analysis (VCA), which is the endmember extraction method applied in NCM-Bayes. The parameter setting for all methods are selected as follows to yield the best performance: For supervised PM-LDA, endmember red roof, light grey roof and beach sand are selected to be the semi-supervised endmembers in this study, note that the open street map doesn't contain beach sand polygons but it's very feasible to be manually selected because it covers only a few superpixels. The other parameters are  $K = 7$ ,  $\lambda = 1$ ,  $\alpha = 0.3$ ,  $\epsilon = 10\%$  and  $T = 200$ . For NCM-Bayes, the Markov chain length was set to 250, the length of the burn-in period was set to 1000,  $\delta = 0.001$ , and the initial endmember variance was set to 0.001. For PM-LDA,  $K = 7$ ,  $\lambda = 1$ ,  $\alpha = 0.3$  and  $T = 200$ .

Similar to experiments on Pavia, seven endmember distribution parameters includ-

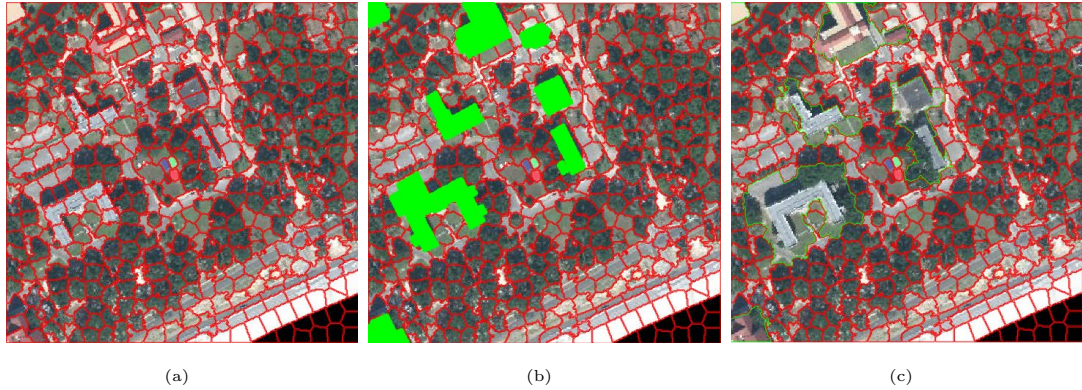


Figure 4.18: Superpixels on Gulfport: (a) superpixels by SLIC; (b) red: superpixels by SLIC, green: open street map; (c) merge of superpixels and open street map.

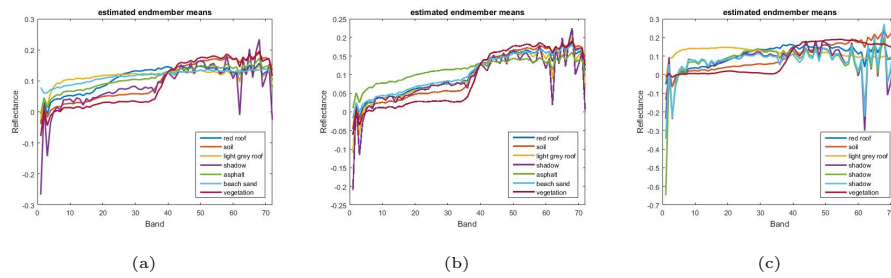


Figure 4.19: Endmember means on Gulfport estimated by (a) supervised PM-LDA, (b) PM-LDA, (c) VCA.

ing endmember means and covariance matrices are learned. The 7 desired endmembers are red roof, soil, light grey roof, shadow, asphalt, beach sand and vegetation. These estimated endmember means are shown in Fig 4.19 and proportion maps are shown in Fig 4.20. All desired materials are found in proportion maps of sPM-LDA. For PM-LDA, the red roof and light grey roof are not learned. For NCM-Bayes, soil is mixed with most materials in the dataset as is shown in (p) and shadow is over estimated to 3 endmembers from (r) to (t).

Then, the quantitative evaluation for three methods are performed. sPM-LDA

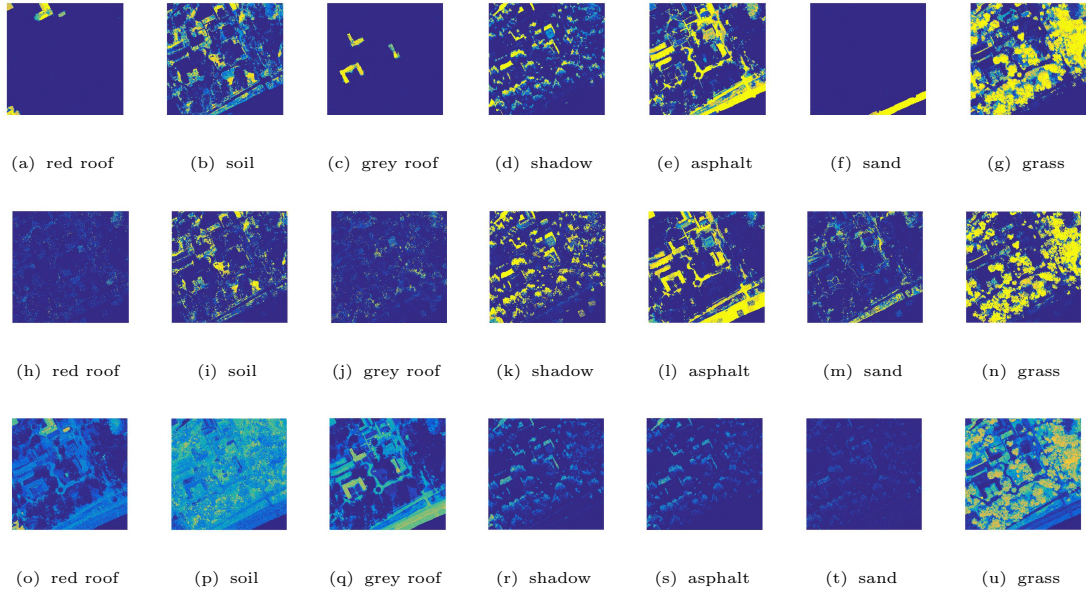


Figure 4.20: Estimated proportion maps on Gulfport using (a)-(g) supervised PM-LDA, (h)-(n) PM-LDA, (o)-(u) NCM-Bayes.

achieves the best entropy performance and PM-LDA has the largest log-likelihood. These evaluation results are very similar to that of Pavia.

Dataset	NCM-Bayes	PM-LDA	sPM-LDA
Pavia	1.15e5	4.18e4	<b>3.96e4</b>

Table 4.6: Overall proportion map entropy for three methods on Gulfport.

Dataset	NCM-Bayes	PM-LDA	sPM-LDA
Pavia	3.28e6	<b>2.09e7</b>	1.97e7

Table 4.7: Overall log-likelihood for three methods on Gulfport.

### Robustness analysis using imprecise labels

From the RGB image of Gulfport data set, it proved to be more reasonable to assign the superpixels in red roof endmember regions two building endmember labels because they contain two moderate different materials, which can be defined as red

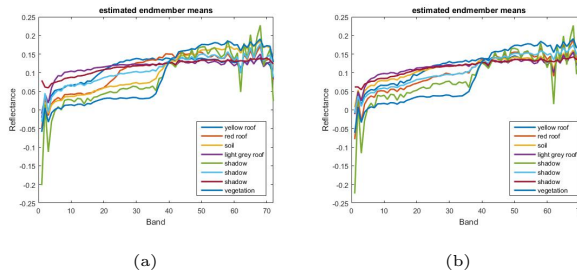


Figure 4.21: Endmember means on Gulfport estimated by (a) semi-supervised PM-LDA with precise labels, (b) semi-supervised PM-LDA with imprecise labels.

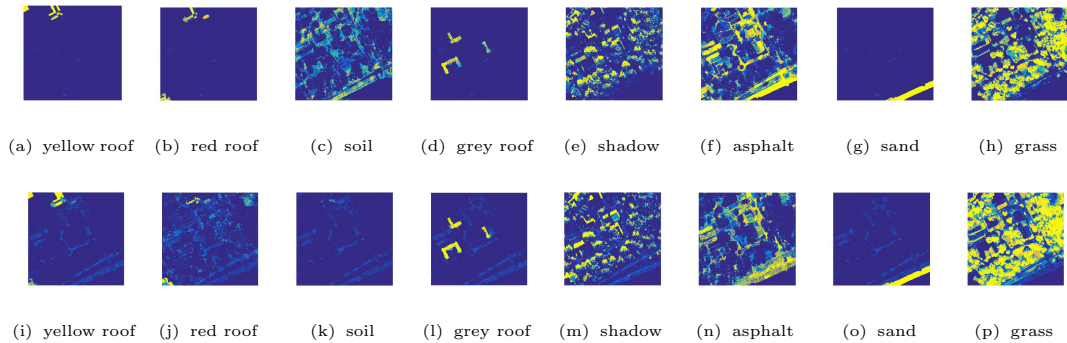


Figure 4.22: Estimated proportion maps on Gulfport using (a)-(f) semi-supervised PM-LDA with precise labels, (g)-(i) semi-supervised PM-LDA with imprecise labels.

roof and yellow roof. The experiment using precise versus imprecise labels is designed as: (1) For experiment with precise labels, the superpixels in the original red roof regions are manually relabeled as red or yellow roof individually; (2) For experiment with imprecise labels, these superpixels are all assigned with two undeclared building endmembers. Thus the number of endmembers is  $k = 8$ .

Labels	Precise	Imprecise
Entropy	<b>4.66e4</b>	4.93e4
Likelihood	<b>1.92e7</b>	1.78e7

Table 4.8: Overall proportion map entropy and log-likelihood for semi-supervised PM-LDA with precise and imprecise labels on Gulfport.

The proportion maps of yellow roof and red roof are well estimated as are shown in Fig 4.22, if the precise labels are used. However, in the case of imprecise labels, the red roof pixels show high yellow roof proportions. Therefore, the estimation and separation of yellow roof and red roof can be achieved perfectly by precise labels better than the imprecise labels. Meanwhile, the estimated endmembers using imprecise labels are not as good as precise labels, which may result in some estimation error on some other endmembers. For example, the soil is not well estimated using imprecise labels in this experiment.

The quantitative results, both the entropy and likelihood, show that unmixing performance suffers when using imprecise labels.

# Chapter 5

## Summary and Concluding Remarks

This thesis has reviewed literature around different types of hyperspectral unmixing approaches. Both the *e*FUMI based instance influence estimation and semi-supervised PM-LDA based hyperspectral unmixing that are discussed in this thesis can be regarded as semi-supervised learning approaches with human interaction.

*e*FUMI has the advantage of learning target concept signature with uncertain bag-level labels. The proposed instance influence estimation approach can effectively indicate the most influential points in a single run. Thanks to the influence estimation, human interaction on the bag-level instance labels can be intelligently modified based on their individual influences. The performance of the proposed algorithm has been examined on two real hyperspectral images from different perspectives including single instance influence and instance set (superpixel) influence estimation.

The hyperspectral unmixing methods with spectral variability are also reviewed. One of the effective ways to handle spectral variability is to assume endmembers follow the Normal Compositional Model. Partial Membership Latent Dirichlet Allocation

is applied on NCM-based hyperspectral unmixing. Then, a semi-supervised version of PM-LDA is suggested which removes unnecessary endmembers in superpixels with semi-supervised labels. The performance of the proposed method is also tested on the same data sets. Endmember and proportion estimation results demonstrate the improved accuracy of this novel hyperspectral unmixing approach.



# Bibliography

- [1] David Landgrebe. Hyperspectral image data analysis. *IEEE Signal Processing Magazine*, 19(1):17–28, 2002.
- [2] Dimitris Manolakis and Gary Shaw. Detection algorithms for hyperspectral imaging applications. *IEEE Signal Processing Magazine*, 19(1):29–43, 2002.
- [3] Shefali Aggarwal. Principles of remote sensing. *Satellite remote sensing and GIS applications in agricultural meteorology*, 23, 2004.
- [4] Dimitris Manolakis, David Marden, and Gary A Shaw. Hyperspectral image processing for automatic target detection applications. *Lincoln Laboratory Journal*, 14(1):79–116, 2003.
- [5] Nirmal Keshava. A survey of spectral unmixing algorithms. *Lincoln Laboratory Journal*, 14(1):55–78, 2003.
- [6] José M Bioucas-Dias, Antonio Plaza, Nicolas Dobigeon, Mario Parente, Qian Du, Paul Gader, and Jocelyn Chanussot. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE Journal*

- of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2):354–379, 2012.
- [7] Michael E Winter. N-findr: an algorithm for fast autonomous spectral end-member determination in hyperspectral data. In *SPIE's International Symposium on Optical Science, Engineering, and Instrumentation*, pages 266–275. International Society for Optics and Photonics, 1999.
- [8] José MP Nascimento and José MB Dias. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE transactions on Geoscience and Remote Sensing*, 43(4):898–910, 2005.
- [9] Jun Li and José M Bioucas-Dias. Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data. In *IGARSS 2008-2008 IEEE International Geoscience and Remote Sensing Symposium*, volume 3, pages III–250. IEEE, 2008.
- [10] José M Bioucas-Dias. A variable splitting augmented lagrangian approach to linear spectral unmixing. In *2009 First Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, pages 1–4. IEEE, 2009.
- [11] David Stein. Application of the normal compositional model to the analysis of hyperspectral imagery. In *Advances in Techniques for Analysis of Remotely Sensed Data, 2003 IEEE Workshop on*, pages 44–51. IEEE, 2003.
- [12] Olivier Eches, Nicolas Dobigeon, Corinne Mailhes, and Jean-Yves Tourneret. Bayesian estimation of linear mixtures using the normal compositional model.

- application to hyperspectral imagery. *IEEE Transactions on Image Processing*, 19(6):1403–1413, 2010.
- [13] Alina Zare, Paul Gader, and George Casella. Sampling piecewise convex unmixing and endmember extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 51(3):1655–1665, 2013.
- [14] Sheng Zou and Alina Zare. Hyperspectral unmixing with endmember variability using partial membership latent dirichlet allocation. *arXiv preprint arXiv:1609.03500*, 2016.
- [15] Xiaoxiao Du, Alina Zare, Paul Gader, and Dmitri Dranishnikov. Spatial and spectral unmixing using the beta compositional model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6):1994–2003, 2014.
- [16] Yuan Zhou, Anand Rangarajan, and Paul D Gader. A gaussian mixture model representation of endmember variability for spectral unmixing.
- [17] José M Bioucas-Dias and Antonio Plaza. An overview on hyperspectral unmixing: geometrical, statistical, and sparse regression based approaches. In *2011 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 1135–1138. IEEE, 2011.
- [18] Daniel C Heinz et al. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE transactions on geoscience and remote sensing*, 39(3):529–545, 2001.

- [19] Nicolas Dobigeon, Jean-Yves Tournet, and Chein-I Chang. Semi-supervised linear spectral unmixing using a hierarchical bayesian model for hyperspectral imagery. *IEEE Transactions on Signal Processing*, 56(7):2684–2695, 2008.
- [20] Joseph W Boardman. Automating spectral unmixing of aviris data using convex geometry concepts. 1993.
- [21] Yoann Altmann, Abderrahim Halimi, Nicolas Dobigeon, and Jean-Yves Tournet. Supervised nonlinear spectral unmixing using a postnonlinear mixing model for hyperspectral imagery. *IEEE Transactions on Image Processing*, 21(6):3017–3025, 2012.
- [22] Konstantinos E Themelis, Athanasios A Rontogiannis, and Konstantinos D Koutroumbas. A novel hierarchical bayesian approach for sparse semisupervised hyperspectral unmixing. *IEEE Transactions on Signal Processing*, 60(2):585–599, 2012.
- [23] Konstantinos Themelis, Athanasios A Rontogiannis, and Konstantinos Koutroumbas. Semi-supervised hyperspectral unmixing via the weighted lasso. In *ICASSP*, pages 1194–1197, 2010.
- [24] Oded Maron and Tomás Lozano-Pérez. A framework for multiple-instance learning. *Advances in neural information processing systems*, pages 570–576, 1998.
- [25] Alina Zare, Paul Gader, Jeremy Bolton, Seniha Yuksel, Thierry Dubroca, Ryan Close, and Rolf Hummel. Sub-pixel target spectra estimation and detection using functions of multiple instances. In *2011 3rd Workshop on Hyperspectral Image*

- and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–4. IEEE, 2011.
- [26] Alina Zare and Changzhe Jiao. Extended functions of multiple instances for target characterization. In *6th IEEE 2014 Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–4, 2014.
- [27] Changzhe Jiao and Alina Zare. Functions of multiple instances for learning target signatures. *IEEE Transactions on Geoscience and Remote Sensing*, 53(8):4670–4686, 2015.
- [28] Ben Somers, Gregory P Asner, Laurent Tits, and Pol Coppin. Endmember variability in spectral mixture analysis: A review. *Remote Sensing of Environment*, 115(7):1603–1616, 2011.
- [29] Alina Zare and KC Ho. Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing. *IEEE Signal Processing Magazine*, 31(1):95–104, 2014.
- [30] John B Adams, Donald E Sabol, Valerie Kapos, Raimundo Almeida Filho, Dar A Roberts, Milton O Smith, and Alan R Gillespie. Classification of multispectral images based on fractions of endmembers: Application to land-cover change in the brazilian amazon. *Remote sensing of Environment*, 52(2):137–154, 1995.
- [31] Philip E Dennison, Kerry Q Halligan, and Dar A Roberts. A comparison of error metrics and constraints for multiple endmember spectral mixture analysis and spectral angle mapper. *Remote Sensing of Environment*, 93(3):359–367, 2004.

- [32] J-Ph Combe, S Le Mouelic, C Sotin, A Gendrin, JF Mustard, L Le Deit, P Launeau, J-P Bibring, B Gondet, Y Langevin, et al. Analysis of omega/mars express data hyperspectral data using a multiple-endmember linear spectral unmixing model (melsum): Methodology and first results. *Planetary and Space Science*, 56(7):951–975, 2008.
- [33] Dar A Roberts, M Gardner, R Church, S Ustin, G Scheer, and RO Green. Mapping chaparral in the santa monica mountains using multiple endmember spectral mixture models. *Remote Sensing of Environment*, 65(3):267–279, 1998.
- [34] C Ann Bateson, Gregory P Asner, and Carol A Wessman. Endmember bundles: A new approach to incorporating endmember variability into spectral mixture analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 38(2):1083–1094, 2000.
- [35] Fereidoun A Mianji and Ye Zhang. Svm-based unmixing-to-classification conversion for hyperspectral abundance quantification. *IEEE Transactions on Geoscience and Remote Sensing*, 49(11):4318–4327, 2011.
- [36] Francesca Bovolo, Lorenzo Bruzzone, and Lorenzo Carlin. A novel technique for subpixel image classification based on support vector machine. *IEEE Transactions on Image Processing*, 19(11):2983–2999, 2010.
- [37] Ben Somers, Maciel Zortea, Antonio Plaza, and Gregory P Asner. Automated extraction of image-based endmember bundles for improved spectral unmixing. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2):396–408, 2012.

- [38] Alexey Castrodad, Zhengming Xing, John B Greer, Edward Bosch, Lawrence Carin, and Guillermo Sapiro. Learning discriminative sparse representations for modeling, source separation, and mapping of hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 49(11):4263–4281, 2011.
- [39] Kelly Canham, Ariel Schlamm, Amanda Ziemann, Bill Basener, and David Messinger. Spatially adaptive hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 49(11):4248–4262, 2011.
- [40] Miguel A Goenaga, Maria C Torres-Madronero, Miguel Velez-Reyes, Skip J Van Bloem, and Jesus D China. Unmixing analysis of a time series of hyperion images over the guánica dry forest in puerto rico. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(2):329–338, 2013.
- [41] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.
- [42] Chao Chen. *Partial Membership Latent Dirichlet Allocation*. PhD thesis, University of Missouri-Columbia, 2016.
- [43] Chao Chen, Alina Zare, and J. Tory Cobb. Partial membership latent dirichlet allocation for image segmentation. In *Int. Conf. Pattern Recognition (ICPR)*, 2016.
- [44] Hao Sun. Map-guided hyperspectral image superpixel segmentation using semi-supervised partital membership latent dirichlet allocation, 2016.
- [45] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. Slic superpixels compared to state-of-the-art super-

- pixel methods. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2274–2282, 2012.
- [46] Carrick Devine and Michael Dikeman. *Encyclopedia of meat sciences*. Elsevier, 2014.
- [47] Inmaculada Dópido, Jun Li, Prashanth R Marpu, Antonio Plaza, Jose M Bioucas Dias, and Jon Atli Benediktsson. Semisupervised self-learning for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 51(7):4032–4044, 2013.
- [48] Hyperspectral remote sensing scenes, 2002. [http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral\\_Remote\\_Sensing\\_Scenes](http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes).
- [49] Paul Gader, Alina Zare, Ryan Close, Jen Aitken, and Grady Tuell. Muufi gulfport hyperspectral and lidar airborne data set. *Univ. Florida, Gainesville, FL, USA, Tech. Rep. REP-2013-570*, 2013.
- [50] David B Gillis and Jeffrey H Bowles. Hyperspectral image segmentation using spatial-spectral graphs. In *SPIE Defense, Security, and Sensing*, pages 83901Q–83901Q. International Society for Optics and Photonics, 2012.