Hyperspectral Unmixing

Using Total Variation and Sparse Methods

Jakob Sigurðsson

Dissertation submitted in partial fulfillment of a *Philosophiae Doctor* degree
in
Electrical and Computer Engineering

Advisors
Professor Magnús Örn Úlfarsson
Professor Jóhannes R. Sveinsson

PhD Committee
Professor Magnús Örn Úlfarsson
Professor Jóhannes R. Sveinsson
Professor Jocelyn Chanussot

Opponents
Professor José M. Bioucas-Dias
Professor Mario Parente

Faculty of Electrical and Computer Engineering
School of Engineering and Natural Sciences
University of Iceland
Reykjavik, July 2015

Hyperspectral Unmixing Using Total Variation and Sparse Methods
Dissertation submitted in partial fulfillment of a *Philosophiae Doctor* degree in
Electrical and Computer Engineering

Faculty of Electrical and Computer Engineering
School of Engineering and Natural Sciences
University of Iceland
Hjarðarhagi 2-6
107, Reykjavik
Iceland

Telephone: 525-4000

Bibliographic information:
Jakob Sigurðsson, 2015, *Hyperspectral Unmixing Using Total Variation and Sparse Methods*,
PhD dissertation, Faculty of Electrical and Computer Engineering, University of Iceland

ISBN 978-9935-9243-4-6

Printing: Háskólaprent
Reykjavik, Iceland, July 2015

Abstract

The main focus of this thesis is blind hyperspectral unmixing. The proposed unmixing methods seek to exploit the characteristics of remote sensing hyperspectral data. This is achieved by using regularization methods such as total variation, ℓ_q sparsity, and roughness penalties. The total variation regularization is used to capture the spatial piecewise smoothness of the data. Sparsity regularization reflects the inherent sparsity of abundance maps, and a roughness penalty is used to promote smooth endmembers. Due to the high spectral correlation of hyperspectral images, the methods assume a low rank model for the data.

The effects of the tuning and model parameters on the unmixing solution are examined in detail, with respect to the mean square error, spectral angle distance, and also classification accuracy.

The thesis concludes by investigating semi-blind unmixing methods, two methods that combine supervised and unsupervised unmixing are proposed and two methods that make the pure pixel assumption.

The main contributions are the following;

- A new sparse unmixing method using a novel combination of regularization terms. The regularization terms are a first order roughness penalty, and an ℓ_q sparsity regularizer. The first order roughness penalty is able to promote smooth endmembers while retaining known discontinuities in the endmember spectrum. The ℓ_q regularizer promotes sparsity in the abundance maps.

- An improved sparse unmixing method that combines ℓ_q and total variation regularization. Hyperspectral images are spatially piecewise smooth and the addition of the TV regularizer improves the unmixing by incorporating spatial information into the sparse unmixing.

- Two new variations of semi-supervised unmixing, where *a priori* information about endmembers that are known to be in the image is incorporated into unmixing. The first method uses hard regularization to constrain some of the endmembers to be identical to known library endmembers, while the latter method uses soft regularization to constrain some of the endmembers to be similar to known library endmembers. The unknown endmembers are estimated along with the abundances for all endmembers.

- Two unmixing methods that make the assumption that pure pixels are present in the image. The endmembers will be selected from a dictionary constructed from the image itself. This guarantees that the endmembers can be fully explained by material

seen in the image. The two methods use different regularization terms to promote sparsity in the solutions, the first method uses ℓ_2 regularization and the second uses vector ℓ_0 regularization.

Ágrip

Megin áheyrsla þessarar ritgerðar er á aðgreiningu fjölrása fjarkönnunarmynda (*e. blind hyperspectral unmixing*). Aðferðirnar sem eru kynntar leitast við að nýta sér einkenni fjarkönnunarmynda. Það er gert með því að nota reglun, t.d. heildarviksreglun (TV) (*e. total variation*), ℓ_q reglun og fyrstu gráðu reglun. Notkun á heildarviksreglun endurspeglar að gögnin eru þjál á köflum í rúmi (*e. spatially piecewise smooth*). Hlutfallslegt magn (*e. abundances*) efna í fjarkönnunarmyndum er rýrt (*e. sparse*), og rýra ℓ_q reglunin er notuð til að ná fram þeim eiginleika. Fyrsta gráðu refsifall er notað til að þvinga lausnina til að hafa þjál grunnlitróf (*e. endmembers*). Sökum þess að fylgni litrófa myndanna er mikil, þá gera aðferðirnar ráð fyrir því að gögnin lifi í rúmi af lágri gráðu (*e. low rank model*).

Áhrif mismunandi stika (*e. model and tuning parameters*) á lausnina eru grandskoðuð, með tilliti til meðalfervikskekkju (*e. mean square error*), hornfjarlægðar (*e. spectral angle distance*) og flokkunar nákvæmni.

Í lok ritgerðarinnar er fjallað um stýrðar aðgreiningar aðferðir (*e. supervised unmixing*). Tvær aðferðir sem eru stýrðar að hluta (*e. semi-superised unmixing*) eru kynntar, svo og tvær aðferðir sem gera ráð fyrir því að grunnlitrófin séu að finna í gögnunum (*e. pure pixel assumption*).

Megin framlag ritgerðarinnar er eftirfarandi:

- Ný rýr aðgreiningaraðferð sem notar frumlega samsetningu reglunarliða. Reglunarliðirnir eru ℓ_q liður og fyrstu gráðu refsiliður. Fyrstu gráðu refsiliðurinn þvingar grunnlitrófin til þess að vera þjál og getur einnig varðveitt ósamfellur á skilgreindum stöðum í litrófinu. ℓ_q reglunarliðurinn stuðlar að rýrum hlutfallskortum (*e. abundance maps*).

- Ný endurbætt aðgreiningaraðferð sem sameinar ℓ_q og TV refsiliði. Fjölrása fjarkönnunarmyndir eru þjálar á köflum í rúmi. Með því að bæta við TV liðnum, er sá eiginleiki nýttur til að endurbæta rýru ℓ_q aðgreiningaraðferðina.

- Tvær úgáfur að aðgreiningaraðferðum sem eru stýrðar að hluta eru kynntar. Fyrri aðferðin notar harða reglun (*e. hard regularization*) til að þvinga ákveðin grunnlitróf til að vera nákvæmlega eins og fyrirfram ákveðin litróf. Sú senni notar mjúka reglun til að þvinga ákveðin grunnlitróf til að vera áþekk fyrirfram ákveðnum litrófum. Óþekktu litrófin eru fundin samhliða því að hlutfallskort allra litrófana eru fundin.

- Tvær aðgreiningaraðferðir sem gera ráð fyrir því að grunn-litrófin séu að finna í gögnunum eru kynntar. Grunn-litrófin eru valin úr gögnunum sjálfum. Aðferðirnar nota mismunandi reglun til að þvinga fram rýra lausn. Fyrri aðferðin notar ℓ_2 reglun en sú seinni notar ℓ_0-vigur reglun.

Acknowledgments

I would like to thank my supervisors, professors Magnús and Jóhannes, for their patience, guidance and support throughout this work. I consider myself fortunate to have had the opportunity to work with them. I am also grateful to professor Jocelyn Chanussot for accepting to be part of my PhD committee.

I would also like to thank the opponents, professors José M. Bioucas-Dias and Mario Parente for their part in the defense, and excellent comments and observations.

To my friends who have worked at VR-II; Behnood, Eysteinn, Frosti, Gabriele, Hildur, Kyriaki, Nicola, Pedram and Sigurjón, I thank you for your friendship and interesting discussions during these years.

Last but certainly not least, I am grateful to my family: my wife, Sigrún, and my daughters, Iðunn and Urður, who brighten up each and every day.

Contents

List of Figures

List of Tables

Notations

$\boldsymbol{A}$	all matrices are in upper case bold
$\boldsymbol{A}^\mathsf{T}$	transpose of $\boldsymbol{A}$
$\boldsymbol{A}_{\text{-}n}$	$\boldsymbol{A}$ with its nth column removed
$\boldsymbol{a}_{(j)}$	column j in matrix $\boldsymbol{A}$
$\boldsymbol{a}_i^\mathsf{T}$	row i in matrix $\boldsymbol{A}$
a_{ij}	element in row i and column j, in matrix $\boldsymbol{A}$
a_i	element i in vector $\boldsymbol{a}$
$\mathbf{1}_p$	a $p \times 1$ vector with all elements equal to one
$\mathbf{1}_{pm}$	a $p \times m$ matrix with all elements equal to one
$\odot$	elementwise multiplication (Hadamard product)
$\oslash$	elementwise division
$\otimes$	Kronecker product
$\text{tr}(\boldsymbol{A})$	trace of $\boldsymbol{A}$
$\text{diag}(\boldsymbol{a})$	diagonal matrix with $\boldsymbol{a}$ on its main diagonal, other values are zero
$\text{vec}(\boldsymbol{A})$	vectorization of $\boldsymbol{A}$ where the columns are stacked on top of one another
$\text{sgn}(a)$	sign of a and $\text{sgn}(0) = 0$

Chapter 1

Introduction

This chapter begins with introducing remote sensing hyperspectral images. The concept of hyperspectral unmixing is explained and main approaches used in unmixing are reviewed. Dimensionality reduction and its relationship to hyperspectral unmixing is also discussed. The chapter concludes with the goals and novelties of the work presented here, along with an overview of the thesis.

1.1 Hyperspectral Images

Hyperspectral images (HSI) are acquired using sensors that measure reflected light in hundreds of narrow spectral bands from across the electromagnetic spectrum [1]. These sensors cover the visible to the near infrared spectral region. For each pixel in a hyperspectral image, a spectrum is obtained. Different materials reflect light differently, so the reflected spectra of one material may be different from the reflected spectra of another material. Since the spectral resolution of hyperspectral sensors is high, it is possible to identify different material in the HSI using spectroscopic analysis [2].

The primary advantage of hyperspectral imaging is the fact that an entire spectrum is acquired for each pixel in the scene. This advantage comes at a cost, the complexity and sensitivity of hyperspectral systems is large and the computational processing power needed to analyze HSI is considerable. These factors greatly increase the cost of acquiring and analyzing hyperspectral data.

Hyperspectral data is often presented as a three dimensional data cube. One such data cube is shown in Figure 1.1. The three dimensions of the cube are the spatial (X and Y) and spectral (Z) dimensions. The data in Figure 1.1 is a remote sensing hyperspectral image of an urban landscape and is created with the HYDICE hyperspectral sensor. The spatial resolution of this sensor is 20m$\times$20m and the spectral resolution is 10nm.

HSI were originally used in mining and geology, but have spread into other fields such as oil industries, manuscript research, medical imaging, forensic applications, food processing, astronomy, quality control and and remote sensing [2].

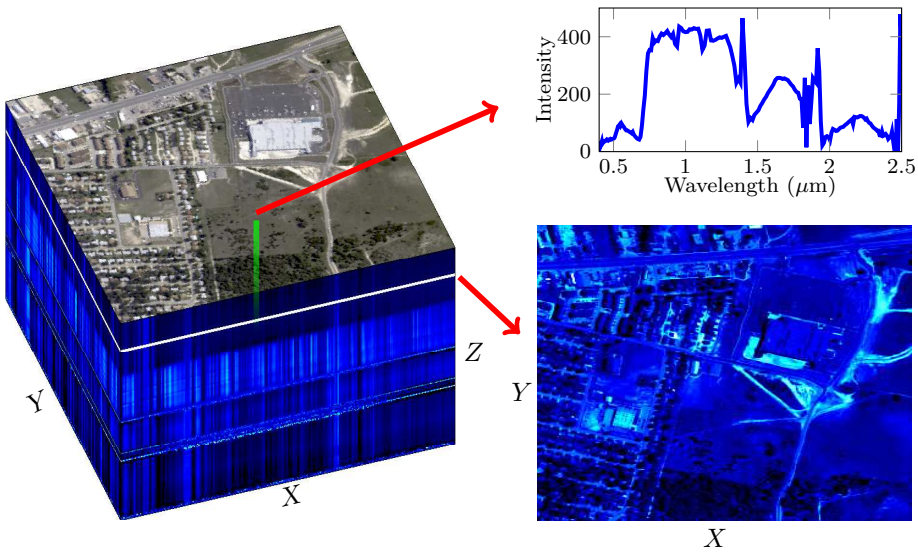


Figure 1.1: A visual representation of hyperspectral data. On the left is the hyperspectral data cube with an RGB image of the scene on top. At the top right is the spectral reflectance from one location. At the bottom right is the reflectance from all locations, as a map, at one specific wavelength, shown as a white line on the cube. X and Y represent the spatial dimensions and Z represents the spectral dimension.

1.2 Remote Sensing Hyperspectral Images

Remote sensing dates back to the 1950s and refers to observing and acquiring information about an object or a scene without coming into contact with it. In modern terms remote sensing involves using airborne sensors that remotely detect objects on earth. Remote hyperspectral imaging has been defined as measuring, analyzing, and interpreting the spectra acquired from a given scene (or specific object) at a short, medium or long distance by an airborne or satellite sensor [3].

Remote sensing HSI are created using airborne sensors. These sensors measure the sunlight (radiance) that reflects from the ground level. The HSI used here, are remote sensing HSI that are created by sensors that cover the visible, near-infrared and shortwave spectral bands (0.3μm to 2.5μm). These HSI are described in Appendix B. Both the spatial and spectral resolution varies between sensors, and developers are constantly improving these sensors so the quality is improved and the resolution is increased.

Remote sensing HSI are corrupted in a number of ways. The radiance that is measured by the sensor has been degraded by atmospheric effects. The light that travels to the

sensor is subject to absorption and scattering from atmospheric gases and aerosols. These effects must be removed from the data if it is to be used for qualitative remote sensing. Atmospheric correction is an active research topic and it is an important step in hyperspectral imaging. Many different methods have been developed to tackle this problem [4–6].

The instruments and circuits of the hyperspectral sensors are susceptible to noise, such as thermal, quantization, and shot noise. Thermal and quantization noise are typically modeled as signal independent Gaussian additive noise, while shot noise can be modeled by variance dependent additive noise [7–10].

There are spatial artifacts in HSI that originate from the way that the sensors capture the images. Non periodic partially deterministic disturbance patterns typically appear in HSI [11]. These artifacts can be horizontal or vertical lines, if the imaging method is the push-broom method [12]. Other acquisition methods may have different types of spatial artifacts dependent on the imaging method.

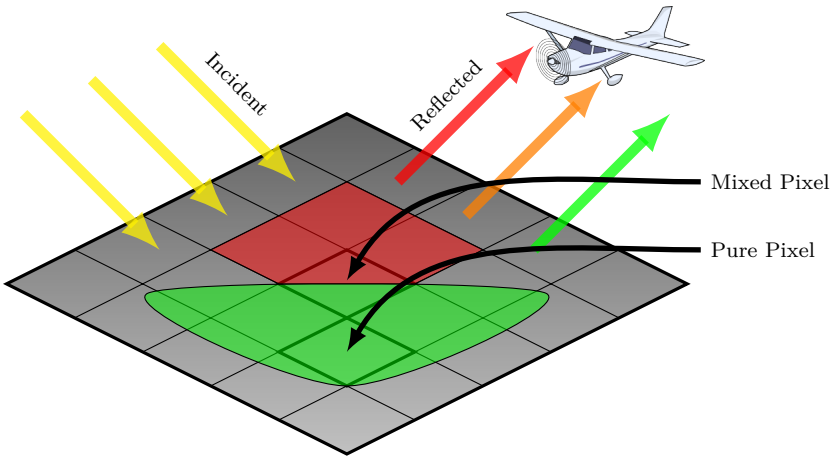


Figure 1.2: An illustration of the hyperspectral mixing process. The green, red and gray areas represent different material on the ground surface. The grid represents the resolution of the sensor. Indecent light reflects off the surface and is captured by an airborne hyperspectral sensor. Some of the pixels may be mixed pixels while others may be pure.

The spatial resolution of remote sensing HSI is such that more than one material can be within the spatial boundaries of one pixel. The pixels that contain multiple material are called *mixed pixels*, in contrast to *pure pixels* (PP) that are pixels containing only one material [13]. Figure 1.2 illustrates this. Each pixel in a HSI is thus a sum of spectral reflectances from the different materials within the spatial boundaries of the pixel.

1.3 Hyperspectral Unmixing

The spectral reflectance of one specific material is called an *endmember*. A measured reflectance cannot be negative, so most unmixing methods constrain the endmembers to be nonnegative. The endmembers in a HSI represent the pure material in the image. The number and types of endmembers in a HSI can vary quite a bit depending on the type of scene. In an urban area the endmembers may be rooftops, asphalt, trees, grass, etc., while a mining district may have endmembers that represent different types of minerals.

An abundance vector states how much of each material is in a pixel and an *abundance map* specifies the amount of one specific material in all the pixels. The abundances are also constrained to be nonnegative, since a negative abundance is not logical. This constraint is termed *abundance nonnegativity constraint* (ANC). An abundance is often viewed as a percentage value, and that the abundance represents the cross-regional relative area of the corresponding endmember in the pixel. This has been confirmed in laboratory experiments [2]. In addition to viewing the abundances as percentages, researchers have also constrained the sum of the abundances for each pixel to be one. This constraint is called the *abundance sum constraint* (ASC). If this constraint is applied, then the unmixing method must account entirely for each pixel in the image. The ASC has also been enforced softly in some unmixing methods [14–18].

However, small highly reflective surfaces that dominate other less reflective material may give misleading abundances. Incident angles may also cause the reflected light to differ for the same material, e.g., slanted rooftops may have the same endmember but the measured intensities may differ greatly between surfaces that are not parallel. The measured spectra of a material in a shaded area will also be much lower than of the spectra of the same material not in shade. For these reasons, viewing the abundances as percentages may not be entirely correct and applying the ASC may not be appropriate for all the pixels in a HSI.

Hyperspectral unmixing is the task of identifying the endmembers and abundances in a HSI [13]. Most classical approaches to unmixing have been based on the low rank model (LRM). This model is often referred to as the linear mixing model (LMM). These models assume that the signal lives in a low dimensional subspace. LRM assume that the spectral reflectance is a linear function of the cross-sectional area of the material. This is in essence a nonnegative matrix factorization (NMF) problem where the hyperspectral image is represented by a matrix, which is factorized into two nonnegative low rank matrices. These two matrices are composed of the endmembers and abundances, respectively.

The LRM is considered valid and sufficient when the incident light only reflects off one surface and the scale is macroscopic. When this is not the case, and the reflected light reflects off more than one surface, or if the mixture is intimate, then the LRM is not considered acceptable. Figure 1.3 illustrates the difference between an intimate nonlinear mixture and a linear mixture. In this thesis, only the LRM is considered.

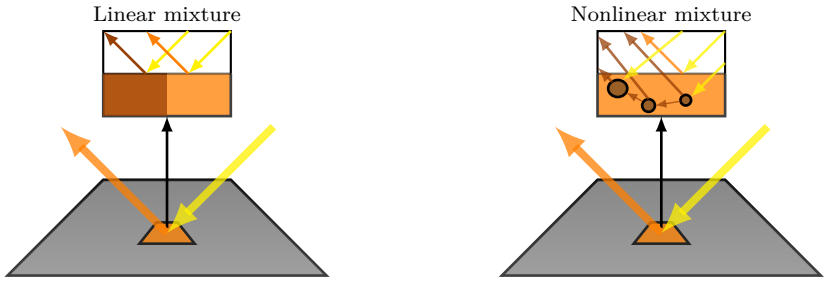


Figure 1.3: Mixture scenarios. Linear (left) and nonlinear (right) mixtures.

Readers interested in nonlinear unmixing are referred to [19–22].

Unmixing methods are often categorized into geometrical, statistical, and sparse regression based methods:

1. *Geometrical approaches* assume that the spectral vectors are positioned in a simplex set or in a positive cone.
2. *Statistical approaches* use parameter estimation methods to identify the endmember and abundance parameters.
3. *Sparse regression* methods assume that the abundances are sparse and formulate the task as a sparse regression problem. These methods rely on spectral dictionaries.

Early approaches to determining endmembers in HSI were mainly manual [23,24], but there have been many recent methods that have estimated the endmembers either automatically or semi-automatically. Most methods associate only one spectral signature to each endmember, some researchers have used multiple signatures to describe an endmembers [25–29]. These endmember bundles account for variability between pixels that represent the same material. These variations are referred to as endmember variability [30].

1.3.1 Geometrical Approaches

The geometrical approach to unmixing can be further split into two categories: PP approaches, and minimum volume (MV) approaches.

Pure Pixel Approaches

Many methods have made the PP assumption, which assumes that every material in the image is represented by at least one PP. This assumption facilitates the design of

very computationally efficient methods. Having each endmember represented by at least one representative pixel in the data can make the validation process simple. It can also be easier for practitioners in the field to interpret the endmembers if they are constructed from actual data elements. This behavior is also important in other fields of research [31, 32]. Endmember determination algorithms that make the PP assumption have been widely used in remote sensing [33–53].

Well known methods that use this approach are (among others) the pixel purity index (PPI) [33], N-FINDR [34] and VCA [37]. PPI begins by reducing the dimensionality of the data using maximum noise fraction (MNF) [54]. A large set of random vectors (skewers) is defined, and then every spectral vector in the data is projected onto these skewers. Following this, the pixels are scored and the pixels with the highest score are defined to be the purest pixels in the data. N-FINDR inflates a simplex inside the data, and finds the set of pixels that maximize the volume of the simplex. VCA is an iterative procedure that uses the subspace already defined by the endmembers and projects the data onto a direction orthogonal to this subspace. The extreme of this projection (the pixel corresponding to the maximum value) corresponds to the new endmember. This procedure is repeated until the endmembers are exhausted.

Minimum Volume Approaches

This approach does not assume that the endmembers are represented by pure pixels in the data, which makes the optimization nonconvex, and thus is much harder to solve. This approach attempts to find a set of endmembers that minimize the volume of the simplex that the endmembers define. This is done, subject to constraint that the observed data be confined within the simplex [2]. This constraint may be soft or hard. These methods rely on the presence of at least $n-1$ spectral vectors on each facet of the simplex. When the data is highly mixed, this condition may however not hold. Methods that adopt this approach are [15, 55–62].

Three well known methods using the minimum volume approach are the minimum volume simplex analysis (MSVA) [58], the simplex identification via variable splitting and augmented Lagrangian (SISAL) [59] and minimum volume transform-nonnegative matrix factorization (MVC-NMF) [15]. MSVA and SISAL allow the nonnegativity constraint to be violated and thus are able to gain a certain amount of robustness. MVC-NMF does not violate the nonnegativity constraint and the minimization is done by alternately estimating the abundances and endmembers in an iterative manner. MVC-NMF does not reduce the dimensionality of the data while SISAL and MSVA do.

1.3.2 Statistical Approaches

When the scenarios are highly mixed, the number of spectral vectors on the simplex facets decreases and this will cause the geometrical methods to yield worse results.

In these scenarios, researchers have looked towards statistical approaches as alternatives. These approaches are very powerful but the computational complexity of these methods is higher than the complexity of the geometrical methods. Using a statistical framework, the unmixing can be formulated into a statistical inference problem and the properties of the underlying distribution can be deduced by analyzing the data.

When both, the abundances, and endmembers themselves need to be estimated, the unmixing is an BSS problem. Independent Component Analysis (ICA) [63, 64], a well known method for solving BSS problems, has been proposed for unmixing hyperspectral data [65–69].

ICA makes the assumptions that the endmembers are linear mixtures weighted by the correspondent abundance fractions and that the abundances are independent. For HSI, the former of these assumptions is true, but the latter is not valid, due to physical constraints on the acquisition process [70]. In [70], both simulated and real hyperspectral data are considered and the authors come to the conclusion that ICA will always have endmembers that are incorrectly mixed and that the performance of ICA tends to increase, both with signal variability and with increasing number of endmembers. NMF is another BSS method that has been extensively used for hyperspectral unmixing. Following the publication of the Lee-Seung algorithm for NMF [71, 72], NMF became widely used in many fields. Hyperspectral unmixing using NMF was first used in 2005 [73][1], and again in 2006 [74]. In the following years many variants or extensions of NMF for hyperspectral unmixing have been published, see for example [15–18, 52, 75–78]. All NMF based unmixing methods are iterative. They alternate between estimating the abundances and the endmembers and this is repeated until convergence is achieved. NMF based unmixing methods often apply a sparsity regularization on the abundances with the ℓ_1 norm being the most common regularizer [79].

Two well known methods that use NMF are MVC-NMF [15] and $L_{1/2}$-NMF [76]. MVC-NMF was briefly explained in Section 1.3.1. $L_{1/2}$-NMF incorporates an $\ell_{1/2}$ sparsity term that promotes sparse abundances.

Bayesian approaches are among the statistical approaches that have been applied to unmixing. These methods treat the endmembers as distributions and are thus able to account for spectral variability. However, exact knowledge about the distributions may not be available in practice. When this is the case, the unknown distributions and the parameters, along with the abundances are jointly estimated during the unmixing. A prominent statistical distribution used for endmembers is the normal compositional model (NCM) [80]. In [81], a Markov chain Monte Carlo sampling approach is taken, where the mean values of endmembers are assumed to be known. In [82], a similar approach is taken, but the endmember covariance values are assumed to be known, but not the endmember means. Other models have also been used in hyperspectral unmixing, such as the Beta composition model [83], and the method of higher moments [84].

[1] No hyperspectral unmixing methods that use NMF, prior to this one, are known to the author.

1.3.3 Sparse Regression Approaches

When a library of spectral signatures is available, the unmixing may be achieved by identifying the endmembers in a HSI from this library of endmembers. The endmembers that compose the library can be collected from the location of the image using field-sensors or obtained from a laboratory. There may however be calibration mismatches between the real image spectra and the spectra available from a library [85]. The HSI may be acquired under different circumstances than the library and may also be degraded by atmospheric effects.

This type of unmixing has been achieved by using linear regression based on sparsity inducing regularizers. The optimal set of endmembers is identified by reducing the support of endmembers in the library that are not in the HSI. This is done recursively, and in each iteration the library is reduced, until an optimal set of endmembers is left.

Classical methods have been used to solve similar problems, such as orthogonal matching pursuit (OMP) [86] which uses an ℓ_0 norm as a sparsity inducing regularizer, basis pursuit (BP) and basis pursuit denoising (BPDN) [87] which use an ℓ_1 norm. OMP and BPDN were not specifically designed for hyperspectral unmixing, but there have been many sparse regression methods published that are designed specifically for hyperspectral unmixing [85, 88–97].

Well known hyperspectral unmixing methods that use sparse regression are SUnSAL (sparse unmixing by variable splitting and augmented Lagrangian) [90, 91] and SUnSAL-TV [85, 94]. SUnSAL and its variants are based on the alternating direction method of multipliers (ADMM) [98] and solve the constrained least squares, the constrained BP, and the constrained BPDN problems. SUnSAL-TV is similar to SUnSAL, with the addition of a total variation (TV) [99, 100] regularization term which exploits the fact that abundance maps are piecewise smooth.

Along with the availability of spectral libraries, methods to prune the libraries have emerged. Pruning the spectral library before unmixing, can both, reduce the complexity dramatically, and improve the unmixing results. In [88], a spectral library is pruned prior to unmixing. First, a low dimensionality subspace is found using HySime (hyperspectral signal identification by minimum error) [101], then the distance from each member of the library to the estimated subspace is calculated. This distance measure is used to prune the library. Collaborative sparse regression [89] is then used to find the endmembers. In [92], a slight extension to [88] is proposed and a framework is presented for monitoring plant production systems.

1.4 Dimensionality Reduction

Hyperspectral images are composed of hundreds of spectral bands while the number of endmembers in each image is typically much lower. Hyperspectral unmixing is synonymous to dimensionality reduction (DR) and subspace identification in the sense

that it tries to find a low rank representation of the data. In hyperspectral unmixing, the rank of this representation is equal to the number of endmembers. Having a low rank representation of the data can be very advantageous, operating on the data in the signal subspace is much less computationally intensive, the data storage required is less, and the signal-to-noise ratio (SNR) can also be increased. DR for hyperspectral data can be split into two approaches, i.e., unsupervised and supervised.

There are differences between DR and hyperspectral unmixing. While DR is a way to achieve a low rank representation of the data, the goal may not specifically be to find the endmembers and abundances, DR may attempt, e.g., to remove noise, maximize SNR, improve classification results. The constraints made on the bases of the vectors that span the low rank subspace may not result in endmembers that are physically meaningful. Hyperspectral unmixing attempts a DR with the constraints that the (bases) endmembers represent physical material in the scene. This is often achieved by incorporating prior knowledge about the data into the unmixing method, such as assuming the endmembers are nonnegative and that the abundances be both sparse and piecewise smooth.

Estimating the number of endmembers in HSI is a very important and challenging problem in hyperspectral data analysis. Classical methods such as Akaike's information criterion (AIC) [102], Bayesian information criterion (BIC) [103] and Stein's unbiased risk estimation (SURE) [104, 105] have been used for low rank model estimation [106–111].

Principal component analysis (PCA) [112] has also been used for rank estimation. This is often done by using the cumulative proportion of the variance accounted for by the current and all preceding principal components (PC). When the ith component retains a specified percentage of the cumulative variance explained by the PCs, i is set as the rank. Estimating the number of endmembers in HSI is a very active research topic.

1.4.1 Unsupervised Dimensionality Reduction

The bands in hyperspectral images are highly correlated and this has been exploited by methods to select a few informative spectral bands [113, 114]. Techniques that project the data into another subspace have been very popular for DR. PCA minimizes the sum of squares, which can be efficiently computed using singular value decomposition (SVD) [115]. MNF, and equivalently noise adjusted principal components (NAPC) [116] try to maximize the SNR. HySime [101], which is unsupervised and eigen decomposition based, begins by estimating the signal and noise correlation. In the next step, the least squares error is used to select subset of eigenvalues that best represent the signal subspace. In [117], a framework for DR using graph embedding is presented. ICA has also been used and studied as a DR method prior to classification [118].

1.4.2 Supervised Dimensionality Reduction

If classification ground-truth is available for the HSI, then some of the pixels in the image are labeled as belonging to a specific class. This ground-truth information has been used to supervise DR. Well known supervised DR methods are, e.g., Decision boundary feature extraction (DAFE) [119, 120] and nonparametric weighted feature extraction (NWFE) [121].

1.5 Thesis Contributions and Organization

The main topic of this thesis is blind hyperspectral unmixing. All of the unmixing methods presented here are based on the LRM. The LRM is used since HSI are highly spectrally correlated.

The different characteristics of the HSI are exploited using appropriate regularization terms. The usage of sparse regularization reflects the fact that abundance maps are inherently sparse. HSI are also piecewise spatially smooth and TV regularization is used to incorporate this knowledge into the unmixing. The spectral resolution of HSI is high and the endmember spectra is thus a smooth function. Roughness penalties are used to promote smooth endmembers.

In Chapter 2, the low rank model is detailed along with the classical constraints that have been proposed for unmixing in the literature.

In Chapter 3, sparse ℓ_q unmixing is introduced, the novelties within this chapter are the use of sparse regularization terms that reflect and exploit the characteristics of HSI, and the methods used to solve the blind unmixing problem. The regularization terms that are used, are the ℓ_0 [122], the ℓ_q $(0<q<1)$ [123] and the ℓ_1 [124] terms, respectively. These sparsity inducing regularization terms are applied on the abundance maps. A first order roughness penalty is applied alongside the sparsity term and it promotes smooth endmembers. A new majorization-minimization unmixing algorithm is developed that minimizes a novel cost function which combines these regularization terms. Extensive evaluations of the method are performed, w.r.t. the tuning parameters, using both simulated and real hyperspectral data. The effects of imposing the ASC are also examined using a real data set.

In Chapter 3, the sparse properties of the abundance maps are exploited using sparse regularization. HSI do have other properties that can be exploited to further improve the unmixing. One such property is the spatial piecewise smoothness of the abundance maps. In Chapter 4, TV regularization is introduced into the unmixing. The TV regularizer exploits the spatial piecewise smoothness of the abundance maps. The unmixing method [125] combines the ℓ_q and TV regularizers and is thus able to promote sparse and spatially piecewise smooth abundances. A new unmixing method is developed using the dyadic expansion, ℓ_q regularization and a split Bregman implementation of TV denoising. The effects of the tuning parameters on the unmixing solution are thoroughly examined. The chapter concludes by using the abundance

maps generated by the proposed method as features for classification [126], where the tuning parameters are selected using cross-validation.

The first two methods in Chapter 5 partially rely on spectral libraries, or *a priori* information about the HSI. These methods are semi-supervised hyperspectral unmixing methods. The first method is an endmember constrained semi-supervised method [127]. *A priori* information about the endmembers is incorporated into the objective function with hard regularization. This information can be acquired from a spectral library or from the data itself. The second method uses soft regularization to incorporate *a priori* information about the endmembers into the objective function [128].

It can however be appealing to have the endmembers coincide with certain pixels in the image. This is the PP assumption, and guarantees that the endmembers can be fully explained by material seen in the image.

The latter two methods proposed in Chapter 5, make the PP assumption and the unmixing is done using the sparse regression approach. The methods [129] decompose the data matrix into nonnegative endmembers and abundance maps. The endmembers will be selected from a dictionary constructed from the data matrix. Each endmember will coincide with certain columns of the data matrix.

1.6 Publications

Chapter 3 is based on the following publications:

[a] J. Sigurdsson, M.O. Ulfarsson, and J.R. Sveinsson, "Hyperspectral unmixing with ℓ_q regularization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 11, pp. 6793–6806, Nov 2014.

[b] J. Sigurdsson, M.O. Ulfarsson, and J.R. Sveinsson, "Smooth and sparse hyperspectral unmixing using an l_0 penalty," *Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, June 2013.

[c] J. Sigurdsson, M.O. Ulfarsson, J.R. Sveinsson, and J.A. Benediktsson, "A smooth hyperspectral unmixing method using cyclic descent," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2012, pp. 3082–3085.

Chapter 4 is based on the following publications:

[d] J. Sigurdsson, M.O. Ulfarsson, and J.R. Sveinsson, "Hyperspectral image unmixing using dyadic cyclic descent, total variation and ℓ_q sparse regularization," submitted.

[e] J. Sigurdsson, M.O. Ulfarsson, and J.R. Sveinsson, "Total variation and ℓ_q based hyperspectral unmixing for feature extraction and classification," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2015.

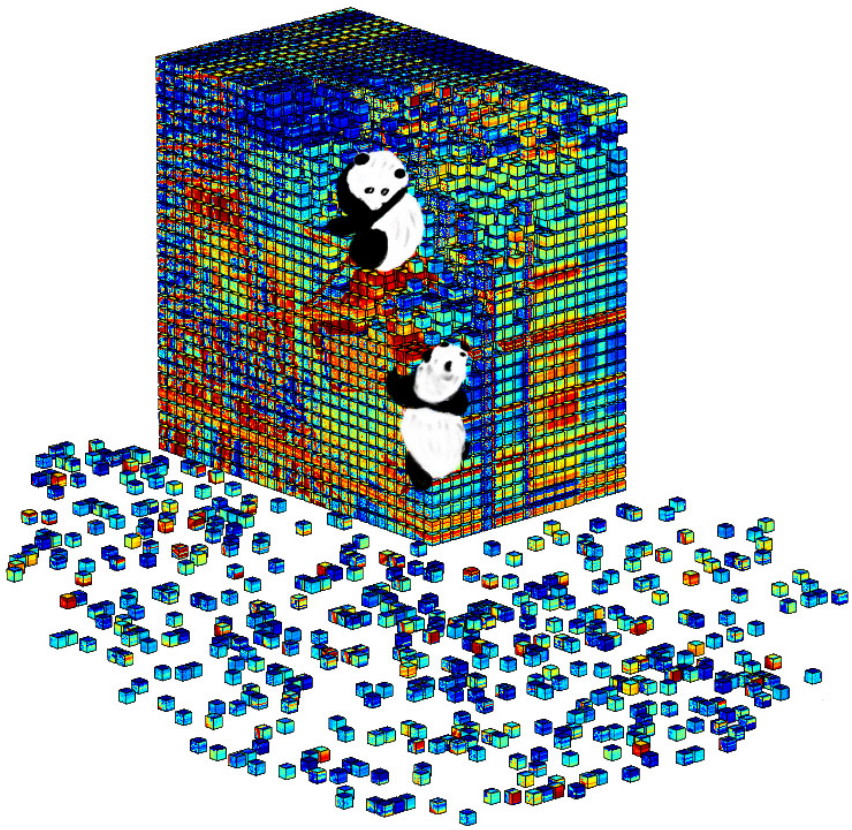
[f] J. Sigurdsson, M.O. Ulfarsson, J.R. Sveinsson, and J.A. Benediktsson, "Smooth spectral unmixing using total variation regularization and a first order roughness penalty," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2013, pp. 2160–2163.

Chapter 5 is based on the following publications:

[g] J. Sigurdsson, M.O. Ulfarsson, and J.R. Sveinsson, "Semi-supervised hyperspectral unmixing," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2014, pp. 3458–3461.

[h] J. Sigurdsson, M.O. Ulfarsson, and J.R. Sveinsson, "Endmember constrained semi-supervised hyperspectral unmixing," *Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, June 2014.

[i] J. Sigurdsson, M.O. Ulfarsson, J.R. Sveinsson, and J.A. Benediktsson, "Sparse representation of hyperspectral data using CUR matrix decomposition," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2013, pp. 433–436.



Chapter 2

Low Rank Models

In this chapter, low rank models are considered. The constraints and regularization terms that are typically used in hyperspectral unmixing are given, along with some well known solutions. Next, two methods that include sparse regularization terms to promote sparsity in the low rank model are given. The chapter concludes with by showing that the ASC cannot be used with the ℓ_1 norm.

2.1 The Low Rank Model and Constraints

Let us assume that a HSI has M spectral bands, r endmembers, and P pixels. For one given pixel, the measured spectral signature, given by the low rank model (LRM) is

$$y_{ip} = \sum_{n=1}^{r} a_{in} s_{np} + w_{ip}, \quad i = 1, \dots, M, \ p = 1 \dots P, \tag{2.1}$$

where $a_{in} \geq 0$ is the measured signature of endmember n at spectral band i, $s_{np} \geq 0$ is the fractional abundance of endmembers n, and w_i denotes noise or modeling error. The model is assumed to be of low rank, i.e., $\min(P, M) > r$. If the abundance should represent a true fraction and the model should account for every endmember in the HSI[1], the the abundance sum constraint (ASC) is added to the model. Applying the ASC, the abundance nonnegativity constraint (ANC) and endmember nonnegativity gives

$$\begin{aligned} &\text{ANC} \quad s_{np} \geq 0, \ n = 1, \dots r, \\ &\text{ASC} \quad \sum_{n=1}^{r} s_{np} = 1, \end{aligned} \tag{2.2}$$

and the endmember nonnegativity constraint is given by

$$a_{in} \geq 0, \quad n = 1, \dots r. \tag{2.3}$$

Define the $r \times 1$ abundance vector as $\boldsymbol{s} = [s_1, \dots, s_r]^\mathsf{T}$, with $\boldsymbol{s}_{(p)}$ being the abundance vector at pixel p. Let $\boldsymbol{A}$ be an $M \times r$ unknown mixing matrix where each column represents one unknown endmember, and $\boldsymbol{y}_{(p)} = [y_{1p}, \dots, y_{Mp}]^\mathsf{T}$ is an $M \times 1$ vector representing the spectral vector at pixel p. The LRM for one pixel can now be written as

$$\boldsymbol{y}_{(p)} = \boldsymbol{A}\boldsymbol{s}_{(p)} + \boldsymbol{n}_{(p)}, \quad p = 1, \dots, P, \tag{2.4}$$

where $\boldsymbol{n}_{(p)} = [w_{1p}, \dots, w_{Mp}]^\mathsf{T}$ is a noise or model error vector.

[1] See Section 1.3 for a discussion about the validity of the ASC.

Given a HSI with p pixels, we denote the image as the $M \times P$ matrix $\boldsymbol{Y} = [\boldsymbol{y}_{(1)}, \ldots, \boldsymbol{y}_{(P)}]$, and the $r \times P$ abundance matrix as $\boldsymbol{S} = [\boldsymbol{s}_{(1)}, \ldots \boldsymbol{s}_{(P)}]$. If the HSI is in the form of an image cube as presented in Section 1.1, then the data is transformed into a matrix where each column is composed spectral signature of one pixel. The LRM for the whole image can be written neatly in matrix formation as

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{S} + \boldsymbol{N}, \tag{2.5}$$

where the noise matrix is $\boldsymbol{N} = [\boldsymbol{n}_{(1)}, \ldots, \boldsymbol{n}_{(P)}]$.

A cost function that is often used for finding a solution to the LRM is the Frobenius norm,

$$J(\boldsymbol{A}, \boldsymbol{S}) = \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{S}\|^2 = \frac{1}{2}\sum_{i=1}^{M}\sum_{p=1}^{P}(y_{ip} - \boldsymbol{a}_i^{\mathsf{T}}\boldsymbol{s}_{(p)})^2. \tag{2.6}$$

The solution is found by minimizing (2.6) w.r.t. $\boldsymbol{A}$ and $\boldsymbol{S}$. If the ASC and nonnegativity constraints are used, then the minimization problem is,

$$\begin{aligned} &\min_{\boldsymbol{A},\boldsymbol{S}} J(\boldsymbol{A}, \boldsymbol{S}) \\ &\text{s.t. } \boldsymbol{A}, \boldsymbol{S} \geq 0,\ \mathbf{1}_r^{\mathsf{T}}\boldsymbol{s}_{(p)} = 1,\ p = 1, \ldots, P. \end{aligned} \tag{2.7}$$

This minimization problem is not uniquely determined, which mean that there are many different values of $\boldsymbol{A}$ and $\boldsymbol{S}$, resulting in the same value of the cost function, but the quality of these solutions with respect to the true abundances and endmembers may differ greatly.

In the next section a well known method to softly enforce the ASC will be discussed, and following that, methods to solve (2.7) will be detailed.

2.1.1 Softly constraining the ASC

The ASC can be softly constrained using matrix augmentation when minimizing (2.6). A early use of this method in hyperspectral unmixing was in [131], but the methodology was published earlier [132, 133]. This is very simple and computationally easy to implement. An additional term is added to (2.6), which softly enforces the ASC using a tuning parameter,

$$J_2(\boldsymbol{A}, \boldsymbol{S}) = \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{S}\|^2 + \frac{\delta^2}{2}\|\mathbf{1}_p^{\mathsf{T}} - \mathbf{1}_r^{\mathsf{T}}\boldsymbol{S}\|^2. \tag{2.8}$$

The larger that δ is set, the more forcefully the ASC is enforced. With $\delta = 0$, the ASC is not enforced. Continuing with (2.8),

$$\begin{aligned}
J_2(\boldsymbol{A}, \boldsymbol{S}) &= \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{S}\|^2 + \frac{\delta^2}{2}\|\mathbf{1}_p^\mathsf{T} - \mathbf{1}_r^\mathsf{T}\boldsymbol{S}\|^2 \\
&= \frac{1}{2}\mathrm{tr}\Big(\boldsymbol{Y}^\mathsf{T}\boldsymbol{Y} - 2\boldsymbol{S}^\mathsf{T}\boldsymbol{A}^\mathsf{T}\boldsymbol{Y} + \boldsymbol{S}^\mathsf{T}\boldsymbol{A}^\mathsf{T}\boldsymbol{A}\boldsymbol{S}\Big) + \frac{\delta^2}{2}\Big(\mathbf{1}_p\mathbf{1}_p^\mathsf{T} - 2\mathbf{1}_p\mathbf{1}_r^\mathsf{T}\boldsymbol{S} + \boldsymbol{s}^\mathsf{T}\mathbf{1}_r\mathbf{1}_r^\mathsf{T}\boldsymbol{S}\Big) \\
&= \mathrm{tr}\frac{1}{2}\Big((\boldsymbol{Y}^\mathsf{T}\boldsymbol{Y} + \delta^2\mathbf{1}_p\mathbf{1}_p^\mathsf{T}) - 2\big(\boldsymbol{S}^\mathsf{T}(\boldsymbol{A}^\mathsf{T}\boldsymbol{Y} + \delta^2\mathbf{1}_r\mathbf{1}_p^\mathsf{T})\big) + \boldsymbol{S}^\mathsf{T}(\boldsymbol{A}^\mathsf{T}\boldsymbol{A} + \delta^2\mathbf{1}_r\mathbf{1}_r^\mathsf{T})\boldsymbol{S}\Big) \\
&= \mathrm{tr}\frac{1}{2}\left(\begin{bmatrix}\boldsymbol{Y}\\ \delta\mathbf{1}_p^\mathsf{T}\end{bmatrix}^\mathsf{T}\begin{bmatrix}\boldsymbol{Y}\\ \delta\mathbf{1}_p^\mathsf{T}\end{bmatrix} - 2\boldsymbol{S}^\mathsf{T}\begin{bmatrix}\boldsymbol{A}\\ \delta\mathbf{1}_r^\mathsf{T}\end{bmatrix}^\mathsf{T}\begin{bmatrix}\boldsymbol{Y}\\ \delta\mathbf{1}_p^\mathsf{T}\end{bmatrix} + \boldsymbol{S}^\mathsf{T}\begin{bmatrix}\boldsymbol{A}\\ \delta\mathbf{1}_r^\mathsf{T}\end{bmatrix}^\mathsf{T}\begin{bmatrix}\boldsymbol{A}\\ \delta\mathbf{1}_r^\mathsf{T}\end{bmatrix}\boldsymbol{S}\right) \\
&= \frac{1}{2}\mathrm{tr}\Big(\boldsymbol{Y}_f^\mathsf{T}\boldsymbol{Y}_f - 2\boldsymbol{A}_f^\mathsf{T}\boldsymbol{S}^\mathsf{T}\boldsymbol{Y}_f + \boldsymbol{S}^\mathsf{T}\boldsymbol{A}_f^\mathsf{T}\boldsymbol{A}_f\boldsymbol{S}\Big) \\
&= \frac{1}{2}\|\boldsymbol{Y}_f - \boldsymbol{A}_f\boldsymbol{S}\|^2,
\end{aligned}$$

where $\mathrm{tr}(\cdot)$ is the trace operator and the augmented matrices are given by

$$\boldsymbol{A}_f = \begin{bmatrix}\boldsymbol{A}\\ \delta\mathbf{1}_r^\mathsf{T}\end{bmatrix} \qquad \text{and} \qquad \boldsymbol{Y}_f = \begin{bmatrix}\boldsymbol{Y}\\ \delta\mathbf{1}_p^\mathsf{T}\end{bmatrix}. \tag{2.9}$$

Softly enforcing the ASC can be achieved replacing $\boldsymbol{Y}$ and $\boldsymbol{A}$ in (2.6) with $\boldsymbol{Y}_f$ and $\boldsymbol{A}_f$ from (2.9), respectively. This method is used many NMF-type unmixing methods, e.g. [14–18, 76, 78].

2.2 Nonnegative Matrix Factorization

NMF became popular for unmixing, following the update rules published in [71, 72]. In [72], two measures between two nonnegative matrices, $\boldsymbol{C}$ and $\boldsymbol{D}$, are given, with the first being the Euclidean distance,

$$D_1 = \|\boldsymbol{C} - \boldsymbol{D}\|^2, \tag{2.10}$$

and the second being

$$D_2(\boldsymbol{C}||\boldsymbol{D}) = \sum_{ij}\left(c_{ij}\log\frac{c_{ij}}{d_{ij}} - c_{ij} + d_{ij}\right). \tag{2.11}$$

Both of these metrics are zero only when $\boldsymbol{C} = \boldsymbol{D}$. The second metric is not a distance measure, since it is not symmetric in $\boldsymbol{C}$ and $\boldsymbol{D}$, and it reduces to the Kullback-Leibler divergence when $\sum_{ij} c_{ij} = \sum_{ij} d_{ij} = 1$. Two formulations for NMF are considered:

$$\min_{A,S} \|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{S}\|^2, \text{ s.t. } \boldsymbol{A}, \boldsymbol{S} \geq 0. \tag{2.12}$$

$$\min_{A,S} D_2(\boldsymbol{Y}||\boldsymbol{A}\boldsymbol{S}), \text{ s.t. } \boldsymbol{A}, \boldsymbol{S} \geq 0. \tag{2.13}$$

Both (2.12) and (2.13) are convex in $\boldsymbol{A}$ only, or $\boldsymbol{S}$ only, but they are not convex in both variables together, so finding a global minima is a challenging task.

The formulation in (2.12) has received much more attention than (2.13) for hyperspectral unmixing. To solve (2.12), [72] proposes elementwise multiplicative update rules[2], that are said to be a good compromise between speed and ease of implementation,

$$\boldsymbol{S} \leftarrow \boldsymbol{S} \odot (\boldsymbol{A}^{\mathsf{T}}\boldsymbol{Y}) \oslash (\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{S}) \quad \text{and} \quad \boldsymbol{A} \leftarrow \boldsymbol{A} \odot (\boldsymbol{Y}\boldsymbol{S}^{\mathsf{T}}) \oslash (\boldsymbol{A}\boldsymbol{S}\boldsymbol{S}^{\mathsf{T}}), \tag{2.14}$$

where $\odot$ and $\oslash$ are elementwise multiplication (Hadamard product) and division operators, respectively. The update rules in (2.14), along with matrix augmentation in (2.9) can be used to find a solution to (2.7) (with ASC softly constrained). An algorithm that solves this problem shown in Algorithm 1.

Algorithm 1: NMF Algorithm

Initialize $k = 1,\ \boldsymbol{A}^{(k)} \geq 0,\ \boldsymbol{S}^{(k)} \geq 0$

while $J(\boldsymbol{A}^{(k)}, \boldsymbol{S}^{(k)})$ *has not converged* **do**

$$\boldsymbol{S}^{(k+1)} = \boldsymbol{S}^{(k)} \odot (\boldsymbol{A}^{(k)})^{\mathsf{T}}\boldsymbol{Y}) \oslash ((\boldsymbol{A}^{(k)})^{\mathsf{T}}\boldsymbol{A}^{(k)}\boldsymbol{S}^{(k)}) \tag{2.15}$$

$$\boldsymbol{A}^{(k+1)} = \boldsymbol{A}^{(k)} \odot (\boldsymbol{Y}(\boldsymbol{S}^{(k)})^{\mathsf{T}}) \oslash (\boldsymbol{A}^{(k)}\boldsymbol{S}^{(k+1)}(\boldsymbol{S}^{(k+1)})^{\mathsf{T}}) \tag{2.16}$$

$$k = k + 1$$

end

The update rules in (2.14), and variants of them have been extensively used in hyperspectral unmixing. See Subsection 1.3.2 for references of methods that use or modify these update rules for hyperspectral unmixing.

In [134], theoretical issues such as convergence, correctness and uniqueness of the NMF solution are addressed. Geometrical conditions are derived under which the factorization is essentially unique in [134].

2.3 Other methods

The NMF method in Algorithm 1 is a special case in a framework which alternatively fixes one matrix and improves the other:

$$\text{Find } \boldsymbol{S}^{(k+1)} \text{ such that } J(\boldsymbol{A}^{(k)}, \boldsymbol{S}^{(k+1)}) \leq J(\boldsymbol{A}^{(k)}, \boldsymbol{S}^{(k)}).$$

$$\text{Find } \boldsymbol{A}^{(k+1)} \text{ such that } J(\boldsymbol{A}^{(k+1)}, \boldsymbol{S}^{(k+1)}) \leq J(\boldsymbol{A}^{(k)}, \boldsymbol{S}^{(k+1)}).$$

The alternative least squares method finds the best point in the scheme, and it is shown in Algorithm 2.

[2] These multiplicative update rules are derived in Appendix C.

Algorithm 2: Alternating nonnegative least squares

Initialize $k = 1,\ \boldsymbol{A}^{(k)} \geq 0,\ \boldsymbol{S}^{(k)} \geq 0$

while $J(\boldsymbol{A}^{(k)}, \boldsymbol{S}^{(k)})$ *has not converged* **do**

$$\boldsymbol{S}^{(k+1)} = \arg\min_{\boldsymbol{S} \geq 0} J(\boldsymbol{A}^{(k)}, \boldsymbol{S}) \tag{2.17}$$

$$\boldsymbol{A}^{(k+1)} = \arg\min_{\boldsymbol{A} \geq 0} J(\boldsymbol{A}, \boldsymbol{S}^{(k+1)}) \tag{2.18}$$

$$k = k + 1$$

end

This is referred to as the *block coordinate descent* or *cyclic descent* method in bound-constrained optimization [135], where one block of variables is minimized while all other blocks of variables are fixed. In the case of NMF, there are only two block variables, $\boldsymbol{S}$ and $\boldsymbol{A}$.

In [136], a study of using projected gradients for NMF is presented, several modifications are done that lead to efficient implementations, exhibit strong optimization properties and converge faster than the multiplicative update approach given in Algorithm 1.

Nonnegative quadratic programming can also be used to solve (2.6). By formulating (2.6) into a quadratic cost function of either $\boldsymbol{A}$ ($\boldsymbol{v} = \boldsymbol{a}_{(i)}$) or $\boldsymbol{S}$ ($\boldsymbol{v} = \boldsymbol{s}_{(p)}$),

$$J(\boldsymbol{v}) = \frac{1}{2}\boldsymbol{v}^\mathsf{T}\boldsymbol{C}\boldsymbol{v} + \boldsymbol{b}^\mathsf{T}\boldsymbol{v}, \tag{2.19}$$

and assuming that $\boldsymbol{C}$ is symmetric, the quadratic programming update rules proposed in [137] can be used. Let $\boldsymbol{C}^+$ and $\boldsymbol{C}^-$ denote the nonnegative matrices, with entries defined as

$$c_{ij}^+ = \begin{cases} c_{ij} & \text{if } c_{ij} > 0 \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad c_{ij}^- = \begin{cases} |c_{ij}| & \text{if } c_{ij} < 0 \\ 0 & \text{otherwise.} \end{cases}$$

The updates rules are then given by

$$v_i^{(k+1)} = v_i^{(k)} \left(\frac{-b_i + \sqrt{b_i^2 + 4(\boldsymbol{C}^+\boldsymbol{v})_i(\boldsymbol{C}^-\boldsymbol{v})_i}}{2(\boldsymbol{C}^+\boldsymbol{v})_i} \right). \tag{2.20}$$

2.4 Sparsity in The Low Rank Model

Inducing sparsity in low rank models has been studied extensively. Most NMF-based hyperspectral unmixing methods apply sparse regularization terms on the abundances to induce sparsity. In the next two subsections, well known methods that extend NMF to include sparsity are described.

2.4.1 Nonnegative Sparse Component Analysis

The goal of linear nonnegative sparse coding is similar to NMF, however with the important distinction that the components be sparse. This means that the original data can represented using only a few nonzero coefficients. To accomplish this, a sparsity inducing penalty term can been added to (2.6). In [138], the ℓ_1 norm is used in the following cost function

$$\begin{aligned} J(\boldsymbol{A}, \boldsymbol{S}) &= \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{S}\|^2 + \lambda\|\boldsymbol{S}\|_1 \\ &= \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{S}\|^2 + \lambda \sum_{ij} |s_{ij}|, \\ &\text{s.t. } \boldsymbol{A}, \boldsymbol{S} \geq 0, \|\boldsymbol{a}_{(i)}\| = 1, \lambda > 0. \end{aligned} \tag{2.21}$$

Optimizing for $\boldsymbol{S}$, given a fixed $\boldsymbol{A}$, yields

$$\boldsymbol{S}^{(k+1)} = \boldsymbol{S}^{(k)} \odot (\boldsymbol{A}^{(k)})^\mathsf{T}\boldsymbol{Y}) \oslash ((\boldsymbol{A}^{(k)})^\mathsf{T}\boldsymbol{A}^{(k)}\boldsymbol{S}^{(k)} + \lambda). \tag{2.22}$$

Projected gradient descent, with step size μ, is used to optimize for $\boldsymbol{A}$. Three steps are involved:

1. $\boldsymbol{A}' = \boldsymbol{A}^{(k)} + \mu(\boldsymbol{A}^{(k)}\boldsymbol{S} - \boldsymbol{Y})\boldsymbol{S}^\mathsf{T}$.
2. Negative values of $\boldsymbol{A}'$ are set to zero.
3. Rescale each column of $\boldsymbol{A}'$ to have unit norm, then set $\boldsymbol{A}^{(k+1)} = \boldsymbol{A}'$.

Using these steps, the algorithm, termed *nonnegative sparse coding* [138], is shown in Algorithm 3.

Algorithm 3: Nonnegative Sparse Coding

Initialize $k = 1,\ \boldsymbol{A}^{(k)} \geq 0,\ \boldsymbol{S}^{(k)} \geq 0$

while $J(\boldsymbol{A}^{(k)}, \boldsymbol{S}^{(k)})$ **has not converged do**

$$\boldsymbol{A}' = \max(0, \boldsymbol{A}^{(k)} + \mu(\boldsymbol{A}^{(k)}\boldsymbol{S}^{(k)} - \boldsymbol{Y})(\boldsymbol{S}^{(k)})^\mathsf{T}) \tag{2.23}$$

$$\boldsymbol{a}'_{(i)} = \frac{\boldsymbol{a}'_{(i)}}{\|\boldsymbol{a}'_{(i)}\|}, \forall i \tag{2.24}$$

$$\boldsymbol{S}^{(k+1)} = \boldsymbol{S}^{(k)} \odot (\boldsymbol{A}^{(k)})^\mathsf{T}\boldsymbol{Y}) \oslash ((\boldsymbol{A}^{(k)})^\mathsf{T}\boldsymbol{A}^{(k)}\boldsymbol{S}^{(k)} + \lambda) \tag{2.25}$$

$$k = k + 1$$

end

In [139], (2.21) is reformulated to work with normalized vectors as

$$J(\boldsymbol{A}, \boldsymbol{S}) = \frac{1}{2}\sum_p \left\| \boldsymbol{y}_{(p)} - \sum_{i=1}^{M} s_{ij} \frac{\boldsymbol{a}_i}{\|\boldsymbol{a}_{(i)}\|} \right\|^2 + \lambda \sum_{ij} |s_{ij}|. \tag{2.26}$$

The update rule for estimating $\boldsymbol{S}$ is the same as (2.25), but the update rule for $\boldsymbol{A}$ is

$$\begin{aligned}\boldsymbol{A}^{(k+1)} = \boldsymbol{A}^{(k)} \odot \Bigg(\boldsymbol{Y}(\boldsymbol{S}^{(k)})^\mathsf{T} + \boldsymbol{A}^{(k)} \odot \Big(\mathbf{1}_{MM}(\boldsymbol{A}^{(k)}(\boldsymbol{S}^{(k)}(\boldsymbol{S}^{(k)})^\mathsf{T}) \odot \boldsymbol{A}^{(k)}) \Big) \Bigg) \\ \oslash \Bigg(\boldsymbol{A}^{(k)}\boldsymbol{S}^{(k)}(\boldsymbol{S}^{(k)})^\mathsf{T} + \odot \Big(\mathbf{1}_{MM}(\boldsymbol{Y}(\boldsymbol{S}^{(k)})^\mathsf{T} \odot \boldsymbol{A}^{(k)}) \Big) \Bigg).\end{aligned} \tag{2.27}$$

This update rule can be using instead of (2.23) in Algorithm 3. This method has the advantage that no step size (μ in (2.23)) parameter is needed.

2.4.2 $\ell_{1/2}$-NMF

In [76], NMF is extended by incorporating an $\ell_{1/2}$ regularizer in a method the authors term $L_{1/2}$-NMF. The authors claim that the $\ell_{1/2}$ regularizer induces sparsity and that it is also a good choice among ℓ_q $(0<q\leq 1)$ regularizers. The cost function used is

$$J(\boldsymbol{A}, \boldsymbol{S}) = \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{S}\|^2 + \lambda \|S\|_{1/2}, \tag{2.28}$$

where $\|S\|_{1/2} = \sum_{ij} s_{ij}^{1/2}$. The update rules proposed in [76] for minimizing (2.28), are

$$\boldsymbol{A}^{(k+1)} = \boldsymbol{A}^{(k)} \odot (\boldsymbol{Y}(\boldsymbol{A}^{(k)})^\mathsf{T}) \oslash (\boldsymbol{A}^{(k)}\boldsymbol{S}^{(k)}(\boldsymbol{S}^{(k)})^\mathsf{T}), \tag{2.29}$$

$$\boldsymbol{S}^{(k+1)} = \boldsymbol{S}^{(k)} \odot (\boldsymbol{A}^{(k+1)})^\mathsf{T}\boldsymbol{Y}) \oslash \Big((\boldsymbol{A}^{(k+1)})^\mathsf{T}\boldsymbol{A}^{(k+1)}\boldsymbol{S}^{(k)} + \frac{\lambda}{2}(\boldsymbol{S}^{(k)})^{-1/2} \Big), \tag{2.30}$$

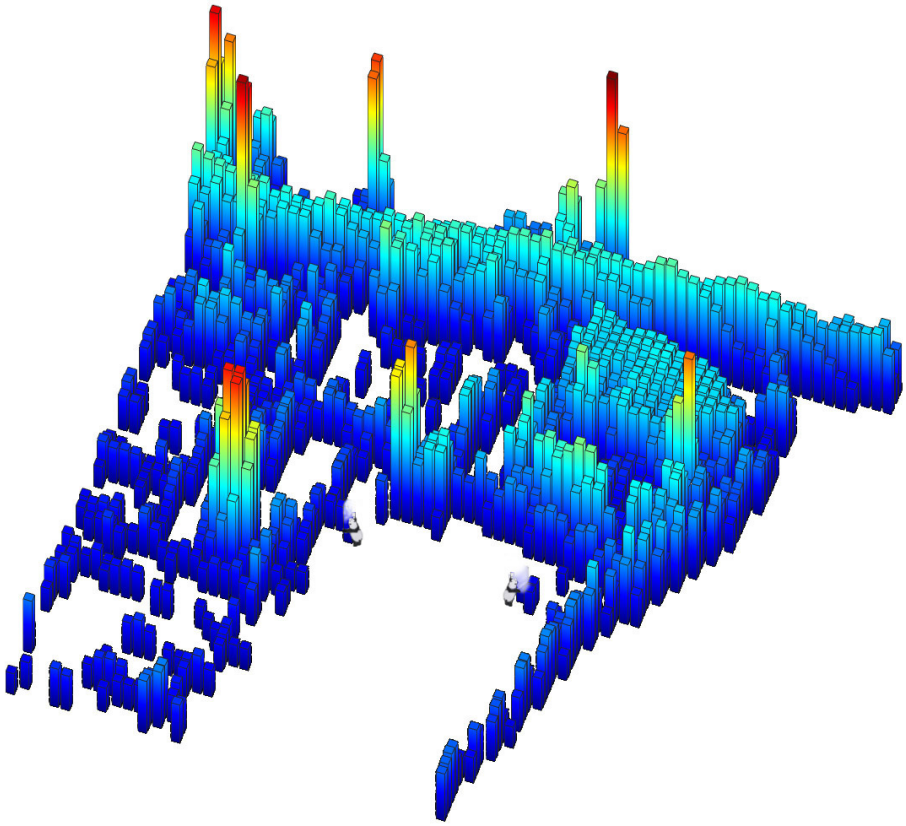
with $\boldsymbol{S}^{-1/2}$ being the elementwise square roots of $\boldsymbol{S}$.

2.4.3 The ASC and the ℓ_1 norm

Using the ℓ_1 penalty along with the ASC will not further induce sparsity. To understand why, let us consider (2.21) with the ASC softly constrained as in (2.8). Note that when $\boldsymbol{S}$ is nonnegative, we can write $\mathbf{1}_r^\mathsf{T}\boldsymbol{s}_{(p)} = \|\boldsymbol{s}_{(p)}\|_1$. Now we write (2.21), discarding terms that do not depend on $\boldsymbol{S}$ as

$$\begin{aligned} J_S =& \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{S}\|^2 + \frac{\delta^2}{2}\|\mathbf{1}_P^\mathsf{T} - \mathbf{1}_r^\mathsf{T}\boldsymbol{S}\|^2 + \frac{h}{2}\sum_{p=1}^{P}\mathbf{1}_r^\mathsf{T}\boldsymbol{s}_{(p)} \\ =& \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{S}\|^2 + \frac{\delta^2}{2}\sum_{p=1}^{P}\Big\{1 - 2(\mathbf{1}_r^\mathsf{T}\boldsymbol{s}_{(p)}) + (\mathbf{1}_r^\mathsf{T}\boldsymbol{s}_{(p)})^2 + \frac{h}{\delta^2}\mathbf{1}_r^\mathsf{T}\boldsymbol{s}_{(p)}\Big\}, \\ =& \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{S}\|^2 + \frac{\delta^2}{2}\sum_{p=1}^{P}\big\{(b - \mathbf{1}_r^\mathsf{T}\boldsymbol{s}_{(p)})^2 + 1 - b^2\big\} \\ =& \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{S}\|^2 + \frac{\delta^2}{2}\|\boldsymbol{b}_p^\mathsf{T} - \mathbf{1}_r^\mathsf{T}\boldsymbol{S}\|^2 + C, \end{aligned} \tag{2.31}$$

where $b = 1 - h/(2\delta^2)$, $\boldsymbol{b}_p$ is a $p \times 1$ vector of with all entries equal to b and C is a constant. In (2.31), we can see that the sparsity term has been rendered ineffective. Applying the ℓ_1 penalty along with the ASC will neither further induce sparsity nor will it force the columns of $\boldsymbol{S}$ to sum to one. It will force the column of $\boldsymbol{S}$ to sum to b without further inducing sparsity.



Chapter 3

Sparse Hyperspectral Unmixing Using ℓ_q Regularization

In this chapter, a novel hyperspectral unmixing method using ℓ_q sparse regularization is detailed. A first order roughness penalty is applied to promote piecewise smooth endmembers. An iterative algorithm for simultaneously estimating the endmembers and the abundances is developed and tested both on simulated and two real hyperspectral data sets. An extensive simulation study is done where both the SNR and the sparsity of the simulated data are varied and the model parameters that minimize the reconstruction errors and the spectral angle distance are identified. The effects of the imposing the abundance sum constraint using a real hyperspectral data set are also examined.

3.1 Introduction

A high intensity pixel in an abundance map means that the corresponding material is abundant in the pixel. It is unlikely that all the materials in the image will be present in every pixel. Thus, there will be locations in the abundance maps that have low or zero value pixels. This means that the abundance maps are sparse in nature. To exploit this fact, sparsity promoting regularization has been widely used in hyperspectral unmixing [2, 140, 141]. The most common sparsity regularization term is the ℓ_1 regularizer. There are exceptions to this, and some researchers have used an ℓ_q regularizer, with $0 \leq q < 1$ [76, 122, 123].

Due to the high spectral resolution of HSIs, the spectrum of the endmembers varies smoothly in wavelength [142]. Noisy bands and water absorption bands are however often removed. This may lead to discontinuities at specific wavelengths in the spectrum of the endmembers.

The approach presented here solves the unmixing problem by optimizing a novel penalized cost function. Nonnegativity is enforced and the ASC can also be enforced. Piecewise smoothness of the endmembers is promoted, as is sparseness in the abundances. As was shown in Subsection 2.4.3, the l_1 penalty will not further promote sparsity while the ASC is enforced. The ℓ_q penalty is able to promote sparsity (albeit to a varying extent) while enforcing the ASC. A first order roughness penalty will be used to encourage piecewise smooth endmembers while maintaining known discontinuities in the spectrum.

Using simulated data, the effects of the ℓ_q penalty on the unmixing solution is evaluated, and the set of parameters that results in the best unmixing results are identified.

The metrics that use are the reconstruction error, the abundance reconstruction error and the spectral angle distance, as defined in Appendix A. Simulation data with different SNR is used, and the sparsity of the data is also varied. The method is applied on two real hyperspectral images, the first is the Urban data set and the second is the Cuprite data set. These data sets are detailed in Appendix B.

The Urban image is unmixed both when the ASC is enforced and not enforced, respectively. The results are evaluated using an RGB image of the scene and the differences between the solutions obtained are discussed. The Cuprite image is unmixed without enforcing the ASC and the endmembers found are compared to a spectral library.

3.2 Related Work

Unmixing methods that use the ℓ_1 penalty term to induce sparsity in the abundance matrices [79, 91, 124] have gained interest. Relatively few publications have examined hyperspectral unmixing using of penalties of lower degree, that is ℓ_q penalties when $0 \leq q < 1$. There are some exceptions to this, e.g., [122] where and ℓ_0 penalty term was used in a majorization-minimization (MM) approach and in [76] where the main focus is the $\ell_{1/2}$ penalty.

ℓ_q penalties and norms have been used in other fields, e.g., in signal reconstruction and in matrix completion problems. In [143], the matrix completion problem is addressed by using an ℓ_q penalty term, in MM approach, with the focus on recovering a low rank matrix from a given subset of its entries. In [144], it is shown that exact signal reconstruction is possible using substantially fewer measurements by replacing the standard ℓ_1 norm with an ℓ_p norm.

In [145], a smooth nPCA method is presented. By adding a first roughness penalty term to the likelihood function which is then maximized for the parameters of interest with an expectation maximization (EM) algorithm.

3.3 The Hyperspectral Model and Cost Function

Using the definitions given in Section 2.1, the hyperspectral LRM used here is

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{S} + \boldsymbol{N}. \tag{3.1}$$

The penalized cost function that is optimized to estimate $\boldsymbol{A}$ and $\boldsymbol{S}$ in (3.1), is given by

$$\begin{aligned} J(\boldsymbol{A},\boldsymbol{S}) =& \frac{1}{2}\|\boldsymbol{Y}-\boldsymbol{A}\boldsymbol{S}\|^2 + \frac{\delta^2}{2}\|\mathbf{1}_P^\mathsf{T} - \mathbf{1}_r^\mathsf{T}\boldsymbol{S}\|^2 \\ &+ \frac{\gamma}{2}\|\boldsymbol{D}\boldsymbol{A}\|_F^2 + \frac{h}{2}\sum_{p=1}^{P}\|\boldsymbol{s}_{(p)}\|_q \\ &\text{s.t. } \boldsymbol{S},\boldsymbol{A} \geq 0. \end{aligned} \tag{3.2}$$

The first term in the cost function is the fidelity term. The second term is controlled by δ and promotes the ASC. The third term encourages piecewise smoothness in the columns of $\boldsymbol{A}$. The last term will promote sparseness in $\boldsymbol{S}$. The term, $\|\boldsymbol{s}_{(p)}\|_q$ with $0 \leq q \leq 1$, is the l_q norm [143] of $\boldsymbol{s}_{(p)}$, and is defined as

$$\|\boldsymbol{s}_{(p)}\|_q = \begin{cases} \|\boldsymbol{s}_p\|_0 & \text{if } q = 0, \\ \sum_{j=1}^{r}|s_{jp}|^q & \text{if } 0 < q < 1, \\ \sum_{j=1}^{r}|s_{jp}| & \text{if } q = 1. \end{cases} \tag{3.3}$$

In (3.3), $\|\boldsymbol{s}_{(p)}\|_0$ is the number of non zero components in $\boldsymbol{s}_{(p)}$. A value of $q = 1$ will penalize entries in $\boldsymbol{S}$ relative to their values, while setting $q = 0$ will penalize all non zero entries the same, regardless of their values. Figure 3.1 shows, how the penalty changes with different values of q.

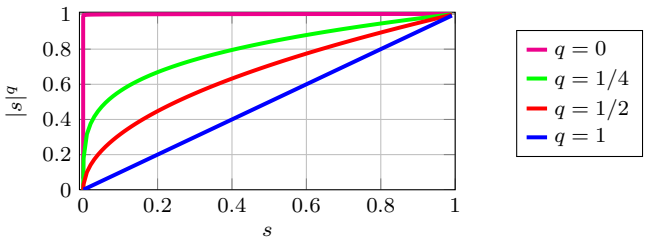


Figure 3.1: The ℓ_q penalty for different values of s and q. When $q = 0$, every nonzero value of s is penalized equally. When $q = 1$, s is penalized in proportion to its value.

The h parameter controls the sparsity of the abundance matrix. Choosing a high value for h will zero out many values in the abundance matrix. This will result in sparse abundance vectors which means that each pixel in the image is associated with few endmembers.

The $\boldsymbol{D}$ matrix is $(M-1) \times M$ and is defined as

$$\boldsymbol{D} = \begin{bmatrix} -1 & 1 & 0 & 0 & \dots & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 \\ & & \ddots & & & \\ 0 & 0 & \dots & 0 & -1 & 1 \end{bmatrix}. \tag{3.4}$$

By choosing the $\boldsymbol{D}$ matrix in this manner a first order roughness penalty is added to the cost function and this will encourage the solution to have smoother endmembers [146]. The parameter γ controls the smoothness of the endmembers. A large value enforces the endmember spectra to be very smooth, while a low value results in very little smoothing. If there are known discontinuities in the endmember spectra it is possible to negate the roughness penalty by replacing the appropriate row in the $\boldsymbol{D}$ matrix with zeros.

3.4 Estimation Method

The optimization approach used is a (block) cyclic descent algorithm [135]. The algorithm consists of two steps, an abundance estimation step and an endmember estimation step. The abundances are estimated while holding the endmembers fixed and then the endmembers are estimated while holding the abundances fixed. This process is repeated until the cost function has converged to its final value.

3.5 Abundance Estimation

To estimate the abundances, or equivalently minimizing (3.2) w.r.t. $\boldsymbol{S}$, a [147] MM approach is used. The MM approach works by finding and iteratively minimizing a majorizing function. In each iteration, the objective function is driven downward until a local minima is found. The terms that are dependent on $\boldsymbol{S}$ are,

$$J_S = \frac{1}{2}\|\boldsymbol{Y}_f - \boldsymbol{A}_f\boldsymbol{S}\|^2 + \frac{h}{2}\sum_{p=1}^{P}\|\boldsymbol{s}_{(p)}\|_q, \tag{3.5}$$

where $\boldsymbol{Y}_f$ and $\boldsymbol{A}_f$ are augmented using (2.9) to softly enforce the ASC. The majorizing function [148] that will be minimized is

$$\begin{aligned} F(\boldsymbol{S}, \boldsymbol{S}^{(k)}) =& J_S + \frac{1}{2}\mathrm{tr}(\boldsymbol{S} - \boldsymbol{S}^{(k)})^\mathsf{T}(\alpha\boldsymbol{I}_r - \boldsymbol{A}_f^\mathsf{T}\boldsymbol{A}_f)(\boldsymbol{S} - \boldsymbol{S}^{(k)}) \\ =& \frac{1}{2}\mathrm{tr}(\alpha\boldsymbol{S}^\mathsf{T}\boldsymbol{S} - 2\boldsymbol{B}^\mathsf{T}\boldsymbol{S} + \boldsymbol{C}) + \frac{h}{2}\sum_{p=1}^{P}\|\boldsymbol{s}_{(p)}\|_q, \end{aligned} \tag{3.6}$$

where $\boldsymbol{S}^{(k)}$ is an estimate of $\boldsymbol{S}$, $\boldsymbol{C}$ is a constant,

$$\boldsymbol{B} = \boldsymbol{A}_f^\mathsf{T}\boldsymbol{Y}_f + \alpha\boldsymbol{S}^{(k)} - \boldsymbol{A}_f^\mathsf{T}\boldsymbol{A}_f\boldsymbol{S}^{(k)},$$

and α is the maximum eigenvalue of $\boldsymbol{A}_f^\mathsf{T}\boldsymbol{A}_f$. Choosing α in this manner ensures that $(\alpha\boldsymbol{I}_r - \boldsymbol{A}_f^\mathsf{T}\boldsymbol{A}_f)$ is non-negative definite, so the following will hold,

$$F(\boldsymbol{S}^{(k)}, \boldsymbol{S}^{(k)}) = J_S \quad \text{and} \quad F(\boldsymbol{S}, \boldsymbol{S}^{(k)}) \geq J_S.$$

Eq. (3.6) is minimized for all three cases: $0 < q < 1$, $q = 0$ and $q = 1$.

3.5.1 $0 < q < 1$

The majorizer function, (3.6), can be written as

$$\begin{aligned} F(\boldsymbol{S}, \boldsymbol{S}^{(k)}) =& \frac{\alpha}{2} \sum_{j=1}^{r} \sum_{p=1}^{P} s_{jp}^2 - \sum_{j=1}^{r} \sum_{p=1}^{P} s_{jp} b_{jp} + \frac{h}{2} \sum_{p=1}^{P} |\boldsymbol{s}_{(p)}|^q \\ =& \sum_{j=1}^{r} \sum_{p=1}^{P} f(s_{jp}, s_{jp}^{(k)}), \end{aligned}$$

where

$$f(s_{jp}, s_{jp}^{(k)}) = \frac{\alpha}{2} s_{jp}^2 - s_{jp} b_{jp} + \frac{h}{2} |s_{jp}|^q. \tag{3.7}$$

Now, (3.7) is reformulated and multiplied with a constant to obtain

$$\begin{aligned} g(s_{jp}, s_{jp}^{(k)}) &= \frac{1}{\alpha} f(s_{jp}, s_{jp}^{(k)}) \\ &= \frac{1}{2} (s_{jp} - \frac{1}{\alpha} b_{jp})^2 + \frac{h}{2\alpha} |s_{jp}|^q + C \\ &= \frac{1}{2} (\beta - z)^2 + \lambda |\beta|^q + C, \end{aligned} \tag{3.8}$$

where C is a constant. The objective is to minimize (3.8), that is, find a solution to

$$\min_{\beta} \frac{1}{2} (\beta - z)^2 + \lambda |\beta|^q. \tag{3.9}$$

This problem is solved using Theorem 1 in [143]. The theorem states that if $0 < q < 1$ and with the following constants defined as

$$\beta_a = [2\lambda(1-q)]^{\frac{1}{2-q}} \quad \text{and} \quad h_a = \beta_a + \lambda q \beta_a^{q-1}.$$

Then the solutions, $\widehat{\beta} = \tau(z)$, of (3.9) are

$$\tau(z) = \begin{cases} 0 & \text{if } |z| < h_a \\ \{0, \operatorname{sgn}(z)\beta_a\} & \text{if } |z| = h_a, \\ \operatorname{sgn}(z)\beta_* & \text{if } |z| > h_a, \end{cases} \tag{3.10}$$

where for $|z| > h_a, \beta_* \in (\beta_a, |z|)$ solves

$$\beta + \lambda q \beta^{q-1} = |z| \quad \text{where} \quad \beta > 0. \tag{3.11}$$

When $|z| > h_a$ there are two solutions to (3.11) and β_* is the larger one which may be computed from the iterations,

$$\beta_{(k+1)} = \rho(\beta_{(k)}) \quad \text{where } \rho(\beta) = |z| - \lambda q \beta^{q-1}, \tag{3.12}$$

with the initial condition $\beta_{(0)} \in [\beta_a, |z|]$.

3.5.2 $q = 0$

In this case, the majorizer function is

$$F(\boldsymbol{S}, \boldsymbol{S}^{(k)}) = \frac{1}{2}\text{tr}(\alpha \boldsymbol{S}^\mathsf{T}\boldsymbol{S} - 2\boldsymbol{B}^\mathsf{T}\boldsymbol{S} + \boldsymbol{C}) + \frac{h}{2}\sum_{p=1}^{P}\sum_{j=1}^{r}\|s_{jp}\|_0. \tag{3.13}$$

Eq. (3.13) is first minimized, ignoring the ℓ_0 regularization (since it is a constant), the solution is $\widehat{\boldsymbol{S}} = \frac{1}{\alpha}\boldsymbol{B}$. Then, the nonzero solution, $F(\widehat{\boldsymbol{S}}, \boldsymbol{S}^{(k)})$, is compared to the zero solution, $F(0)$, and the solution that gives the lower value of F is chosen. The following is compared (ignoring constants)

$$F(\widehat{s}_{jp}, s_{jp}^{(k)}) = \frac{-b_{jp}^2}{\alpha} + h \quad \text{and} \quad F(0) = 0. \tag{3.14}$$

That is,

$$\widehat{s}_{jp} = \begin{cases} \frac{b_{jp}}{\alpha} & \text{if } F(0) - F(\widehat{s}_{jp}, s_{jp}^k) \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.15}$$

This minimization yields

$$\begin{aligned} \boldsymbol{B} &= [b_{jp}] = \boldsymbol{A}_f^\mathsf{T}\boldsymbol{Y}_f + \alpha \boldsymbol{S}^{(k)} - \boldsymbol{A}_f^\mathsf{T}\boldsymbol{A}_f\boldsymbol{S}^{(k)}, \\ s_{jp} &= b_{jp}\frac{1}{\alpha}I(b_{jp} \geq \sqrt{h\alpha}), \\ \boldsymbol{S}^{(k+1)} &= [s_{jp}], \end{aligned}$$

where $I(a > b)$ is 1 if $a > b$ and 0 otherwise and $\boldsymbol{S}^{(k+1)}$ is the new estimate of $\boldsymbol{S}$. The ℓ_0 regularization is incorporated into the solution using hard thresholding.

3.5.3 $q = 1$

The minimizing of (3.6) is now acheived with

$$\boldsymbol{S}^{(k+1)} = \text{soft}\Big(\boldsymbol{S}^{(k)} + \frac{1}{\alpha}\boldsymbol{A}^\mathsf{T}(\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{S}^{(k)}), \frac{h}{2\alpha}\Big), \tag{3.16}$$

where soft$(\cdot)$ is the soft thresholding function defined as

$$\text{soft}(x, T) = \begin{cases} x + T & \text{if} \quad x \leq -T, \\ 0 & \text{if} \quad |x| \leq T, \\ x - T & \text{if} \quad x > T. \end{cases} \tag{3.17}$$

3.6 Endmember Estimation

The nonnegative quadratic programing approach [137] shown in Section 2.3 is used to estimate the endmembers. Let $\boldsymbol{R} = \boldsymbol{D}^\mathsf{T}\boldsymbol{D} = \boldsymbol{R}^+ - \boldsymbol{R}^-$ where

$$R_{ij}^+ = \begin{cases} R_{ij} & \text{if } R_{ij} > 0 \\ 0 & \text{otherwise,} \end{cases} \quad \text{and } R_{ij}^- = \begin{cases} |R_{ij}| & \text{if } R_{ij} < 0 \\ 0 & \text{otherwise.} \end{cases}$$

Now, (3.2) can be rewritten as

$$J_A = -\operatorname{tr}(\boldsymbol{Y}^\mathsf{T}\boldsymbol{A}\boldsymbol{S}) + \frac{1}{2}(\boldsymbol{S}^\mathsf{T}\boldsymbol{A}^\mathsf{T}\boldsymbol{A}\boldsymbol{S}) + \frac{\gamma}{2}\operatorname{tr}(\boldsymbol{A}^\mathsf{T}\boldsymbol{R}^+\boldsymbol{A}) - \frac{\gamma}{2}\operatorname{tr}(\boldsymbol{A}^\mathsf{T}\boldsymbol{R}^-\boldsymbol{A}). \tag{3.18}$$

By defining $\boldsymbol{v} = \operatorname{vec}(\boldsymbol{A})$, $\boldsymbol{b} = -\operatorname{vec}(\boldsymbol{Y}\boldsymbol{S}^\mathsf{T})$ and using the rule [149]

$$\operatorname{tr}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{C}\boldsymbol{D}) = \operatorname{vec}^\mathsf{T}(\boldsymbol{D})(\boldsymbol{A}\otimes\boldsymbol{C}^\mathsf{T})\operatorname{vec}(\boldsymbol{B}^\mathsf{T}),$$

(3.18) can be rewritten as a quadratic cost function,

$$J_A = \frac{1}{2}\boldsymbol{v}^\mathsf{T}\boldsymbol{C}^+\boldsymbol{v} - \frac{1}{2}\boldsymbol{v}^\mathsf{T}\boldsymbol{C}^-\boldsymbol{v} + \boldsymbol{b}^\mathsf{T}\boldsymbol{v}, \tag{3.19}$$

where

$$\boldsymbol{C}^+\boldsymbol{v} = \operatorname{vec}(\boldsymbol{A}\boldsymbol{S}\boldsymbol{S}^\mathsf{T} + \gamma\boldsymbol{R}^+\boldsymbol{A}),$$
$$\boldsymbol{C}^-\boldsymbol{v} = \operatorname{vec}(\gamma\boldsymbol{R}^-\boldsymbol{A}).$$

In each iteration, the multiplicative update rule given in (2.20) is used to decrease the cost function.

3.7 The ℓ_q Unmixing Algorithm

The methods given in the previous sections are used for estimating $\boldsymbol{S}$ and $\boldsymbol{A}$. A cyclic descent algorithm is used where $\boldsymbol{A}$ and $\boldsymbol{S}$ are alternatively estimated until the cost function has converged to its final value. The algorithm is considered converged if

$$\frac{J(\boldsymbol{A}^{(k-1)}, \boldsymbol{S}^{(k-1)}) - J(\boldsymbol{A}^{(k)}, \boldsymbol{S}^{(k)})}{J(\boldsymbol{A}^{(k-1)}, \boldsymbol{S}^{(k-1)})} < 10^{-4}, \tag{3.20}$$

where k is the number of the current iteration and $\boldsymbol{A}^{(k)}$ and $\boldsymbol{S}^{(k)}$ are the current estimations of the endmembers and abundances, respectively. The inputs to the algorithm are the data ($\boldsymbol{Y}$), the initial estimates of $\boldsymbol{A}$ and $\boldsymbol{S}$ and the tuning parameters. Algorithm 4 details this procedure.

3.8 Experimental Results

The method is evaluated using real and simulated data. The simulated data is created using six spectral signatures as endmembers and the abundance maps are generated following a Dirichlet probability density function. The first real data set is the Urban data set detailed in Appendix B.1. The second real data set is the Cuprite data set detailed in Appendix B.2.

Algorithm 4: The ℓ_q Unmixing Algorithm.

Require $\boldsymbol{Y}$, r, h, γ, δ, $\boldsymbol{A}^{(0)}$ **and** $\boldsymbol{S}^{(0)}$
while *the cost function has not converged* **do**
  Abundance step:
  while $\boldsymbol{S}$ *has not converged* **do**
    $\boldsymbol{S}^{(k+1)} = \arg\min_{\boldsymbol{S}} J(\boldsymbol{A}^{(k)}, \boldsymbol{S}^{(k)})$
  end
  Endmember step: while $\boldsymbol{A}$ *has not converged* **do**
    $\boldsymbol{A}^{(k+1)} = \arg\min_{\boldsymbol{A}} J(\boldsymbol{A}^{(k)}, \boldsymbol{S}^{(k+1)})$
  end
  $k \leftarrow k + 1$
end

3.8.1 Simulated Data

A quantitative analysis of our methods is done using simulated data. Six spectral signatures from the United States Geological Survey (USGS) digital spectral library[1] make up the columns of the $\boldsymbol{A}$ matrix. To reduce the computational complexity, the signatures are resampled so the number of bands is 80. The spectral signatures are shown in Figure 3.2.

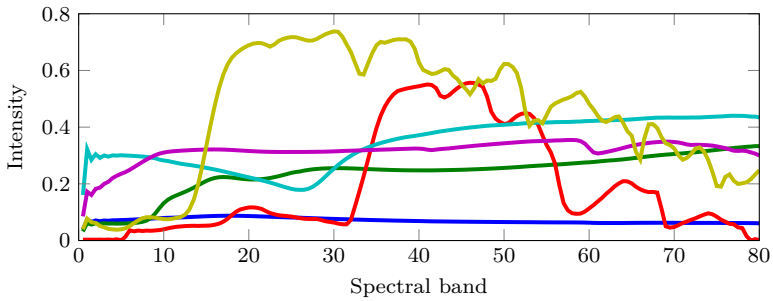


Figure 3.2: The six spectral signatures used in the simulations for ℓ_q unmixing.

Each column in the $\boldsymbol{S}$ matrix is generated following a Dirichlet probability density function, with parameter equal to one. The signal variability is also controlled to be within $[0.7, 1.3]$. This means that the sum of each column in $\boldsymbol{S}$ will be in the aforementioned range. The number of pixels in the data are $P = 300$, so $\boldsymbol{S}$ will be of size 6×300. Additive Gaussian white noise is added uniformly to all bands according to (3.1).

To evaluate the method, the SAD, $\text{nMSE}_{\boldsymbol{S}}$ and $\text{nMSE}_{\boldsymbol{AS}}$, as defined in Appendix A

[1] http://speclab.cr.usgs.gov/spectral.lib06

(A.2-A.4) are used. These evaluation metrics will be referred to as *the metrics*.

Four scenarios will be examined, the first two scenarios will be dedicated to examining how the metrics change while varying q and h. The third scenario will focus on the smoothness parameter, γ, and the fourth scenario examines the sparsity of $\boldsymbol{S}$.

In the first and third scenario, the effects of noise in the data will be examined. In the second scenario, the effects of sparsity of the $\boldsymbol{S}$ matrix will be examined.

The fourth scenario is dedicated to examining the sparsity of $\boldsymbol{S}$, i.e., how the sparsity changes while varying q and h.

To begin with, both δ and γ are set to zero in these simulations and the number of endmembers, r is set to the true number of endmembers used to create the data, that is, $r = 6$. $\boldsymbol{A}$ is initialized using vertex component analysis [37] (VCA). $\boldsymbol{S}$ is initialized with constants, $1/r$.

Varying noise - examine q and h

In this scenario the noise in the data is varied. Four different SNR will be considered: 25dB, 30dB, 35dB and 40dB. The sparsity of $\boldsymbol{S}$ is fixed at 40%. This means that 40% of all entries in $\boldsymbol{S}$ will be randomly set to zero.

The metrics are calculated for $0 \leq q \leq 1$ and $0 \leq h \leq 0.03$. A surface area with the metrics is calculated for all combinations of h and q and plotted as a function of q and h. In this manner the changes in the metrics with respect to the choices of h and q can be evaluated. The results are shown in Figures 3.6 and 3.7.

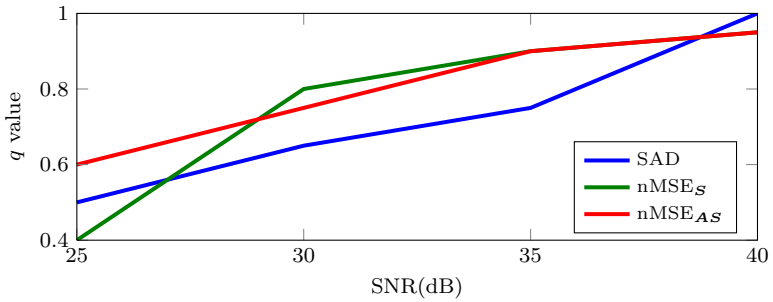


Figure 3.3: The values of q that minimize the metrics for SNR values of 25, 30, 35 and 40 (dB), respectively, in the ℓ_q unmixing simulations.

The minima of the metrics are marked in Figures 3.6 and 3.7. The minima for each SNR value occur within the same regions. E.g., when the SNR is 25dB, the minima of all the metrics are found when $0.4 < q < 0.6$, and $0.005 < h < 0.006$. It is also noted that when the SNR increases, lower values of h and higher values of q minimize the

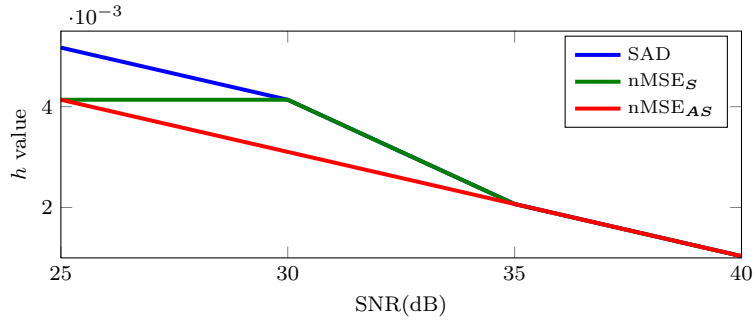


Figure 3.4: The values of h that minimize the metrics for SNR values of 25, 30, 35 and 40 (dB), respectively, in the ℓ_q unmixing simulations

metrics. When the SNR is 40dB, $q \approx 1$ gives the lowest values in the metrics.

However, the value of h at the minima is very low which indicates that the penalty term has a small impact on the solution. This behavior can be better observed in Figures 3.3 and 3.4. The value of q that minimizes the metrics increases when the SNR increases while the value of h decreases.

The endmember matrix was initialized using VCA. In Figure 3.5, the SAD of the initial values is compared to the optimal SAD found. Our method always improved the SAD, although the improvement is minor when the unmixing problem is not very hard, as is the case when SNR>35dB.

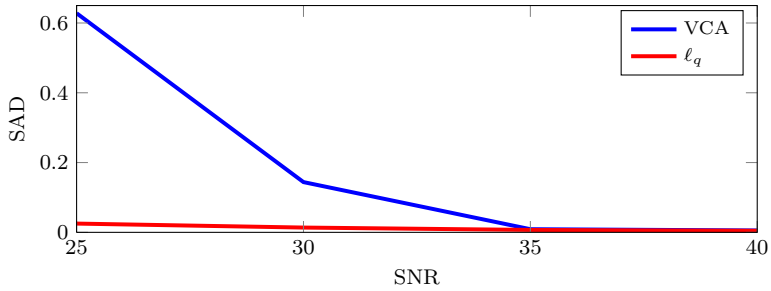


Figure 3.5: The initial SAD compared to the optimal SAD found by the proposed method while varying the SNR, in the ℓ_q unmixing simulations.

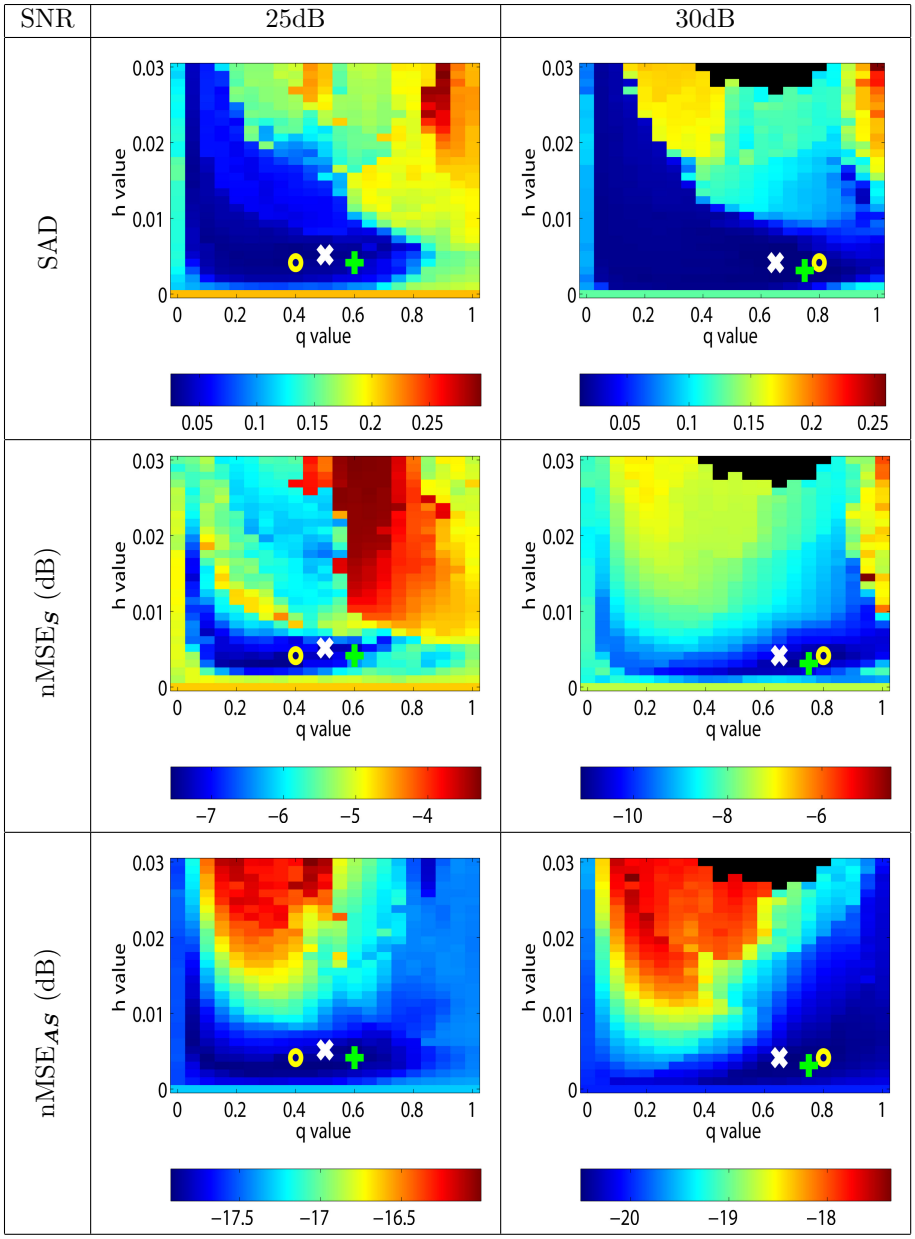


Figure 3.6: ℓ_q unmixing results. With the sparsity fixed at 40%, the SAD, nMSE$_{\boldsymbol{S}}$ and nMSE$_{\boldsymbol{AS}}$ are calculated for snr equal to 25dB and 30dB, respectively. The white '×' shows the location where the SAD is at a minima. The yellow '○' shows the minima of nMSE$_{\boldsymbol{S}}$ and the green '+' shows the minima of nMSE$_{\boldsymbol{AS}}$. The black areas show values of h and q where the number of endmembers were less than 6.

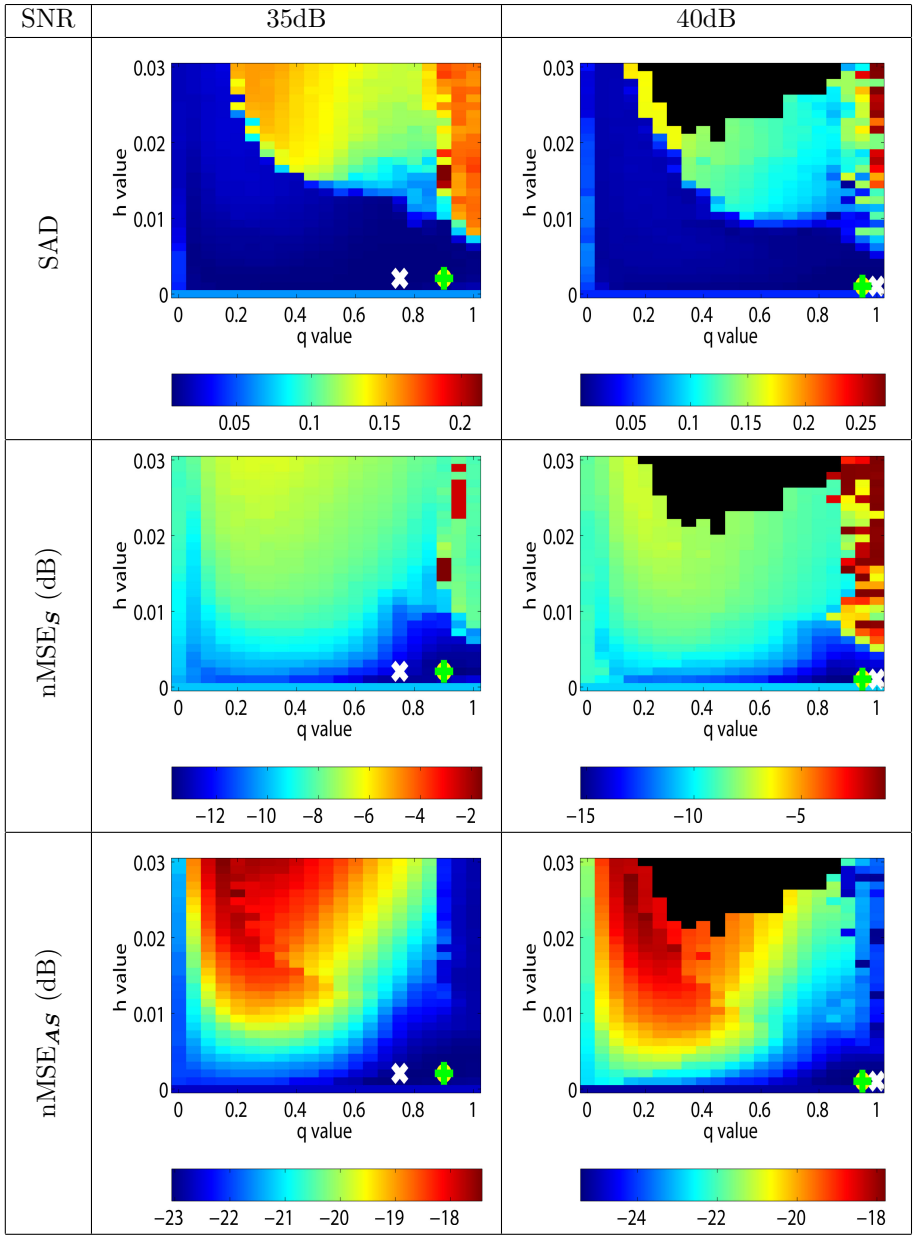


Figure 3.7: ℓ_q unmixing results. With the sparsity fixed at 40%, the SAD, nMSE$_{\boldsymbol{S}}$ and nMSE$_{\boldsymbol{AS}}$ are calculated for SNR equal to 35dB and 40dB, respectively. The white '×' shows the location where the SAD is at a minima. The yellow '○' shows the minima of nMSE$_{\boldsymbol{S}}$ and the green '+' shows the minima of nMSE$_{\boldsymbol{AS}}$. The black areas show values of h and q where the number of endmembers were less than 6.

Varying sparsity - examine q and h

The sparsity of $\boldsymbol{S}$ will be controlled to be 10%, 30%, 50% and 70%, respectively. The variance of the noise will be chosen fixed at a SNR of 30dB. As was done in the previous simulations, the metrics will be calculated for $0 \leq q \leq 1$ and $0 \leq h \leq 0.03$. The results are shown in Figures 3.10 and 3.11.

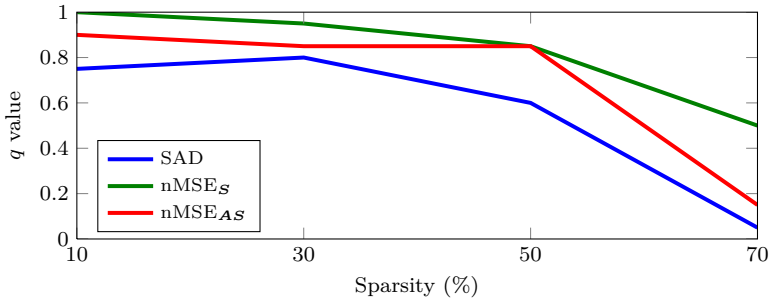


Figure 3.8: The values of q that result in the minima of the metrics when the sparsity of $\boldsymbol{S}$ is 10%, 30%, 50%, and 70%, respectively, in the ℓ_q simulations.

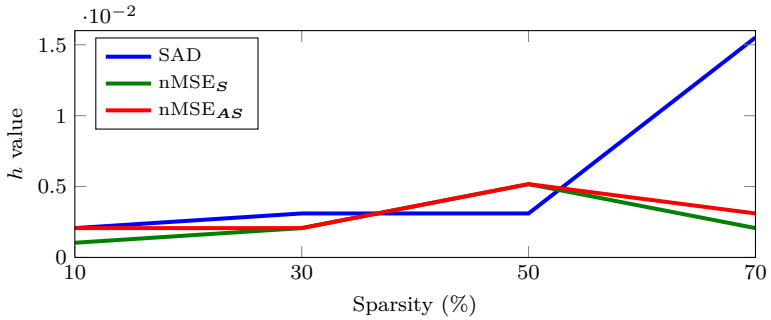


Figure 3.9: The values of h that result in the minima of the metrics when the sparsity of $\boldsymbol{S}$ is 10%, 30%, 50%, and 70%, respectively, in the ℓ_q simulations.

In Figure 3.10, when the sparsity of $\boldsymbol{S}$ is at 10% and 30%, it can be seen that the minima of the metrics occur in the same region. When the sparsity increases (Figure 3.11) the values of q and h that minimize the metrics are not confined to the same region.

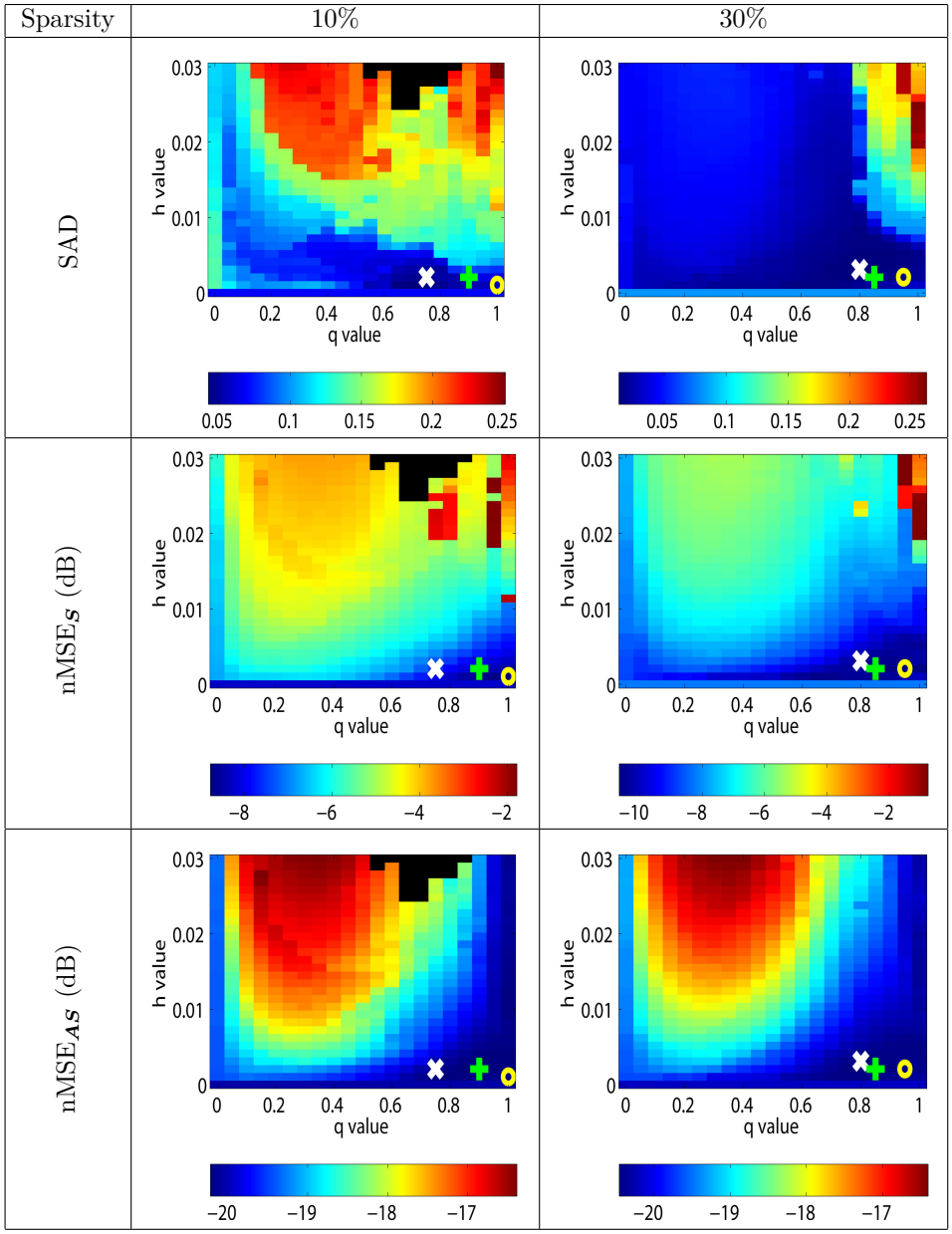


Figure 3.10: ℓ_q unmixing results. With the SNR fixed at 30dB, the SAD, nMSE$_{\boldsymbol{S}}$ and nMSE$_{\boldsymbol{AS}}$ are calculated for when the sparsity of $\boldsymbol{S}$ is 10%, 30%, respectively. The white '×' shows the location where the SAD is at a minima. The yellow '○' shows the minima of nMSE$_{\boldsymbol{S}}$ and the green '+' shows the minima of nMSE$_{\boldsymbol{AS}}$. The black areas show values of h and q where the number of endmembers were less than 6.

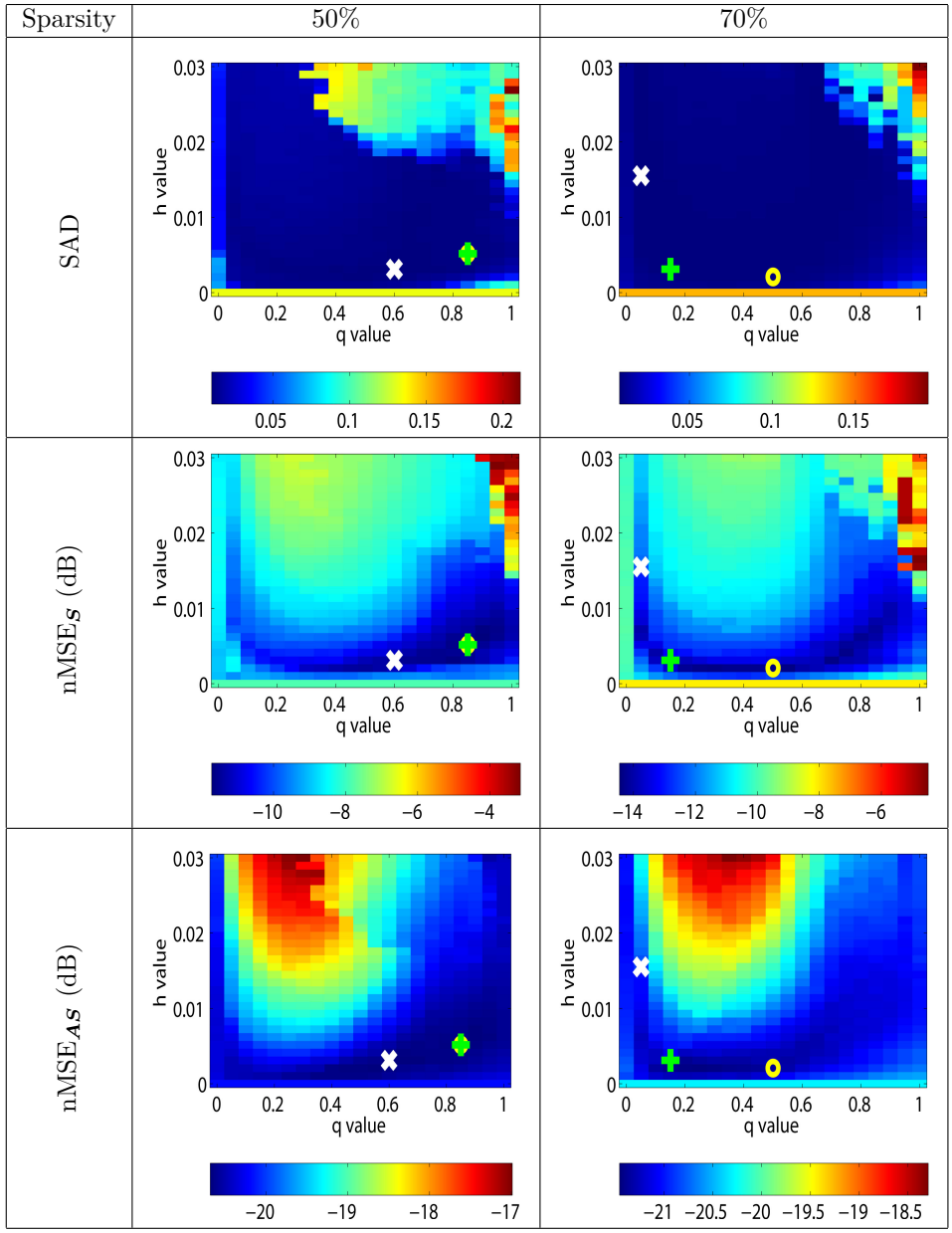


Figure 3.11: ℓ_q unmixing results. With the SNR fixed at 30dB, the SAD, nMSE$_{\boldsymbol{S}}$ and nMSE$_{\boldsymbol{AS}}$ are calculated when the sparsity of $\boldsymbol{S}$ is 50% and 70%, respectively. The white '$\times$' shows the location where the SAD is at a minima. The yellow '$\circ$' shows the minima of nMSE$_{\boldsymbol{S}}$ and the green '$+$' shows the minima of nMSE$_{\boldsymbol{AS}}$. The black areas show values of h and q where the number of endmembers were less than 6.

However, when $\boldsymbol{S}$ is very sparse, the surface areas of the metrics are relatively flat over large regions. This is due to the fact that when $\boldsymbol{S}$ is very sparse, purity of the pixels is higher and this makes the unmixing task easier. This is especially apparent in Figure 3.11, in the case when the sparsity is 70%. In this case, the SAD surface plot is very flat excluding the upper right corner. Most combinations of q and h result in a low SAD and a very good estimate of the endmembers.

In Figure 3.8, it can be seen that when the sparsity increases, lower values of q result in lower metric values. This is partially supported by [143] where the authors state that selecting $q = 0$ in sparse regression and matrix completion outperforms setting $q = 1$ when the underlying model is very sparse.

In Figure 3.9, the value of h that minimizes the metrics does rise with increasing sparsity, although it falls slightly for $\text{nMSE}_{\boldsymbol{S}}$ and $\text{nMSE}_{\boldsymbol{AS}}$ when the sparsity is 70%. In Figure 3.12, the SAD of the initial values is compared to the optimal SAD found.

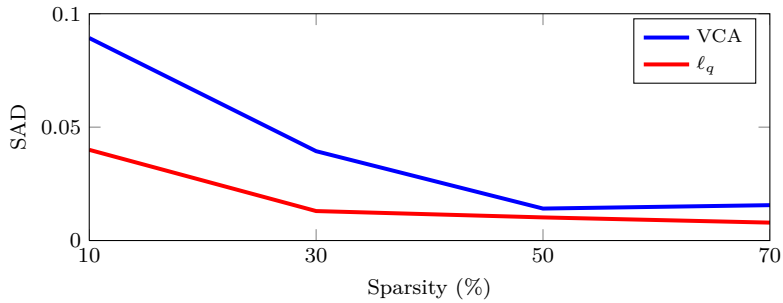


Figure 3.12: The initial SAD compared to the optimal SAD found by the proposed method while sparsity the sparsity, in the ℓ_q simulations.

Varying noise - examine γ

The noise will be varied to be from 25dB to 40dB. The sparsity of $\boldsymbol{S}$ will be fixed at 40%, $q = 0.8$ and $h = 2 \times 10^{-3}$. There is no guarantee that these values of q and h are optimal values for all the SNR values, but based on the results given in Figures 3.6 and 3.7, these parameters are expected to give good results.

For each SNR, the optimal values of γ is found. The values are optimal in the sense that they minimize the SAD and $\text{nMSE}_{\boldsymbol{AS}}$. The results are shown in Figures 3.13 and 3.14.

In Figures 3.13 and 3.14, it can be seen that the roughness penalty can improve the unmixing results when the SNR is lower than 32dB. Both the SAD and the reconstruction error can be lowered by choosing the optimal γ value.

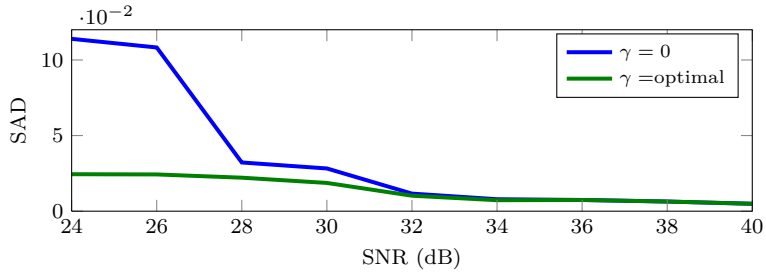


Figure 3.13: The SAD, calculated when the SNR varies from 25dB to 40dB, in the ℓ_q simulations. The blue plot shows the SAD when $\gamma = 0$ and the green plot shows the SAD when the optimal γ value is used.

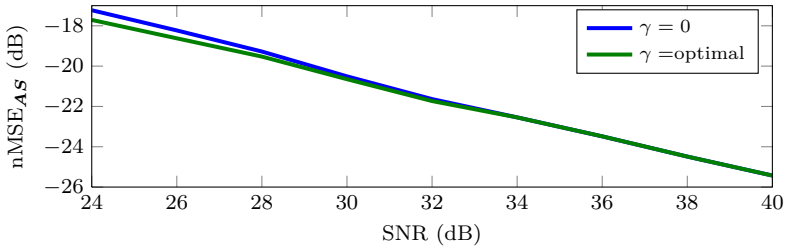


Figure 3.14: The nMSE$_{\boldsymbol{AS}}$, calculated when the SNR varies from 25dB to 40dB, in the ℓ_q simulations. The blue plot shows the nMSE$_{\boldsymbol{AS}}$ when $\gamma = 0$ and the green plot shows the nMSE$_{\boldsymbol{AS}}$ when the optimal γ value is used.

The SNR is now set to 30dB, and h and q are chosen so they are close to minimizing all the metrics. Using Figure 3.6 (right column), the parameters are set to $q = 0.75$ and $h = 0.0031$.

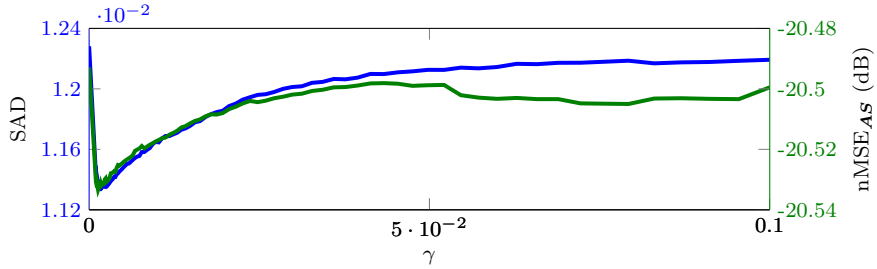


Figure 3.15: The SAD and nMSE$_{\boldsymbol{AS}}$, calculated when the SNR is 30dB, in the ℓ_q simulations. The blue plot shows the SAD and the green plot shows the nMSE$_{\boldsymbol{AS}}$.

In Figure 3.15, it can be seen that the roughness penalty improves the unmixing, even when the optimal q and h have been chosen. The SAD and nMSE$_{\boldsymbol{AS}}$ are at a minima

when $\gamma = 2 \times 10^{-3}$. This is however a marginal increase in performance.

Examine the sparsity of $\boldsymbol{S}$

The sparsity of $\boldsymbol{S}$ will be calculated for $0 \leq q \leq 1$ and $0 \leq h \leq 0.03$. This is done with $\delta = 0$ and also when $\delta = 2$. Setting $\delta = 2$ promotes the ASC forcefully in the sense that more than 90% of all columns of $\boldsymbol{S}$ will sum to 1 ± 0.05. The results are shown in Figure 3.16.

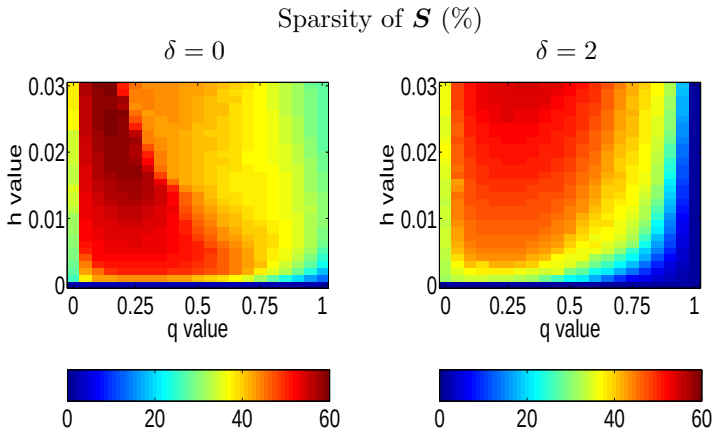


Figure 3.16: The sparsity (%) of $\boldsymbol{S}$ while varying h and q, in the ℓ_q simulations. The left image shows the sparsity when $\delta = 0$ and the right image shows the sparsity when the ASC is enforced by setting $\delta = 2$.

In Figure 3.16 it can be seen that when $\delta = 2$, the sparsity of $\boldsymbol{S}$ does not increase when q is decreased below ≈ 0.3, however when $\delta = 0$ this is not the case, having q lower than ≈ 0.3 can increase the sparsity of $\boldsymbol{S}$. When the ASC is enforced, the sparsity of $\boldsymbol{S}$ will be lower than when the ASC is not enforced for the same value of h, particularly when q is close to 1. It can also be seen that when $q = 1$ it is not possible to promote sparsity. When the ASC is enforced, the ℓ_1 penalty will not promote sparsity.

3.8.2 Real Data

Urban Data

The proposed method is now evaluated using the Urban data set. Low SNR bands, [89, 90, 103-109, 130-152, 204-210] are removed. This yields 171 usable bands out of the original 210. A RGB image is created by using spectral bands from the data to represent the red, green and blue channels in the RGB image. The RGB image of the data is shown in Figure 3.17.

Two unmixing results will be shown, when the ASC is enforced and not enforced, respectively. The two results will be compared and evaluated.

Figure 3.17: The RGB image of the Urban data set. The red x's labeled 1-8 show the locations of the spectra used to initialize the ℓ_q algorithm.

The first step is to estimate the number of endmembers in the data set. This is done using PCA and a visual inspection of the data. The principal components (PCs) and their corresponding eigenvalues are calculated. Let λ_j be the jth eigenvalue and the fraction of total variance retained be

$$\Lambda(n) = \frac{\sum_{j=1}^{n} \lambda_j}{\sum_{j=1}^{M} \lambda_j}. \tag{3.21}$$

In Figure 3.18, $\Lambda(n)$ is plotted vs. the number of eigenvalues. This shows the marginal explanatory power gained from including an additional PC in the model.

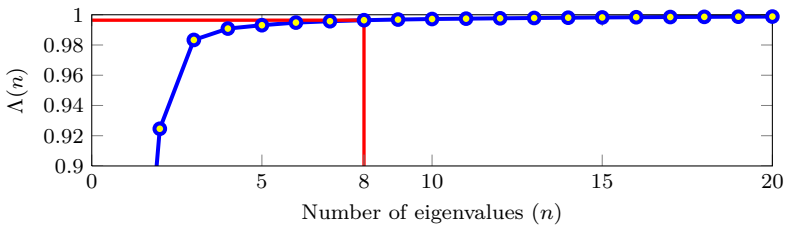


Figure 3.18: The cumulative sum of first 25 normalized eigenvalues ($\Lambda(n)$) found using the Urban data set. Retaining eight PCs explains 99.6% of the variance in the data. The red lines intersects with the 8th PC.

A visual inspection of the RGB image shows that there are at least eight different macroscopic materials in the image: three different colored rooftops, asphalt, dirt, shadow areas and two types of vegetation. Using PCA, and retaining eight PCs explains 99.6% of the variance in the data. The number of endmembers (r) will thus be set to eight.

To estimate suitable values of q and h for the Urban data set, when the ASC is not enforced, the SNR is evaluated, and compared to the simulation data in the previous section. The values of q and h that minimized the metrics in the simulation data with same SNR are chosen for the Urban data set. To estimate the SNR the noise estimation method in [101] is used. This method estimates the SNR to be 33.5dB.

Comparing this SNR to the simulation set (Figure 3.7), the parameters are set to $q = 0.8$ and $h = 0.002$. These values are close to minimizing all the metrics when the SNR in the simulation data is 35dB (left column in Figure 3.7). To make the two data sets comparable, the Urban data is scaled to have the same variance as the simulation data set.

Considering the estimated SNR of the Urban data set, it is unlikely that setting γ to a high value will result in a large performance gain.

The same value that minimized the SAD in Figure 3.15 is chosen, i.e., $\gamma = 0.002$. This is a very low value and will not have a large effect on the solution. The $\boldsymbol{D}$ matrix is modified to allow for discontinuities between bands (100,101) and (120,121). These discontinuities are a result of the removal of low SNR bands.

With the ASC enforced, the same parameter values will be used as when the ASC was not enforced, with the exception that $h = 0.008$. This choice of h along with the ASC will result in approximately the same sparsity of $\boldsymbol{S}$ as without the constraint.

The initial values for the $\boldsymbol{A}$ matrix are set by choosing 8 distinct unique macroscopic material seen in the RGB image. The locations of the initial values are marked with x's and labeled 1-8 in Figure 3.17.

The Urban data set is now unmixed without enforcing the ASC, using $r = 8$, $q = 0.8$, $h = 0.002$, $\gamma = 0.002$ and $\delta = 0$. Enforcing the ASC, the same parameters are used except that $h = 0.008$ and $\delta = 2$. The unmixing results are shown in Figures 3.19 and 3.20, respectively.

When creating the abundance maps, 0.25% of the pixels with the highest intensity values are ignored. Removing these outlier pixels makes visually evaluating the abundance maps, and comparing the two solutions easier.

A ground truth for the Urban data set is not available, so these two solutions are visually compared to each other and to the RGB image. By visually comparing the abundance maps to the RGB image, the endmembers are labeled as the material that

is present in the locations where the abundance maps have high intensity values. In the following text, the term "associated with" refers to high intensity pixels in an abundance map and the material seen in the RGB image at the same pixel locations. The ASC solution refers to the unmixing results when the ASC is enforced and likewise the non-ASC solution is when the ASC is not enforced.

1. The first endmember, labeled Asphalt, is associated with asphalt roads and parking lots. The abundance map, in the ASC solution is noisier than without the ASC, and it is also more noticeably associated with non-asphalt areas such as dirt and light rooftops.

2. The second endmember is labeled Light Rooftops and the third is labeled Grass. The endmember spectra and abundance maps are similar in both the ASC and non-ASC solutions.

3. The third endmember is labeled Grass and both ASC and non-ASC solutions are similar.

4. Endmember four is labeled Trees. The abundance maps in both solutions are similar, but the ASC solution shows some association to grass areas.

5, 6. The fifth and sixth endmembers show associations to both dirt areas and red rooftops.

7. The seventh endmember has some association with dirt areas and some rooftops. The abundance maps have however low intensity values and are noisy.

8. The eighth endmember in the non-ASC solution is not associated with any specific material and the abundance map has low intensity values while the same endmember in the ASC solution is associated with shadow areas.

The histograms of the sum of the columns if $\boldsymbol{S}$ are also shown in Figures 3.19 and 3.20. When the ASC is enforced, 99.5% of the columns of $\boldsymbol{S}$ sum to 1 ± 0.01. If the ASC is not enforced, the sum of the columns ranges from 0 to 4.

Both enforcing the ASC and not enforcing it, results in good unmixing and most abundance maps can be associated with few materials seen in the RGB image. A ground truth of this data is not available so it is not possible to definitively state which method is better. Not enforcing the ASC does though seem to give slightly clearer abundance maps, and these maps can each be associated with fewer materials in the RGB image, respectively.

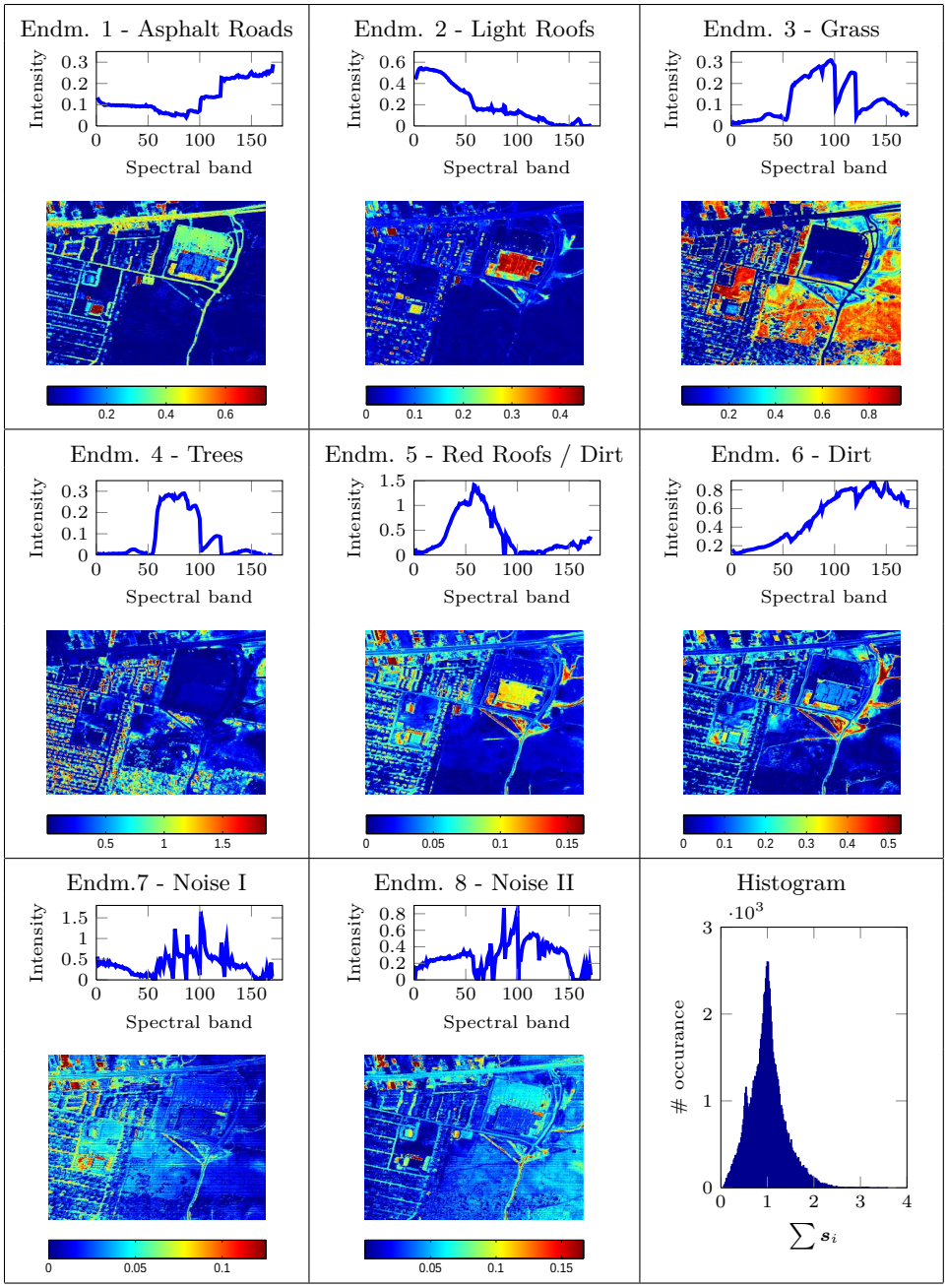


Figure 3.19: The ℓ_q unmixing results using the Urban data set with the ASC not enforced. The endmembers are labeled by visually comparing the abundance maps to the RGB image. The plot in the lower right corner shows a histogram of the sum of the columns of $\boldsymbol{S}$.

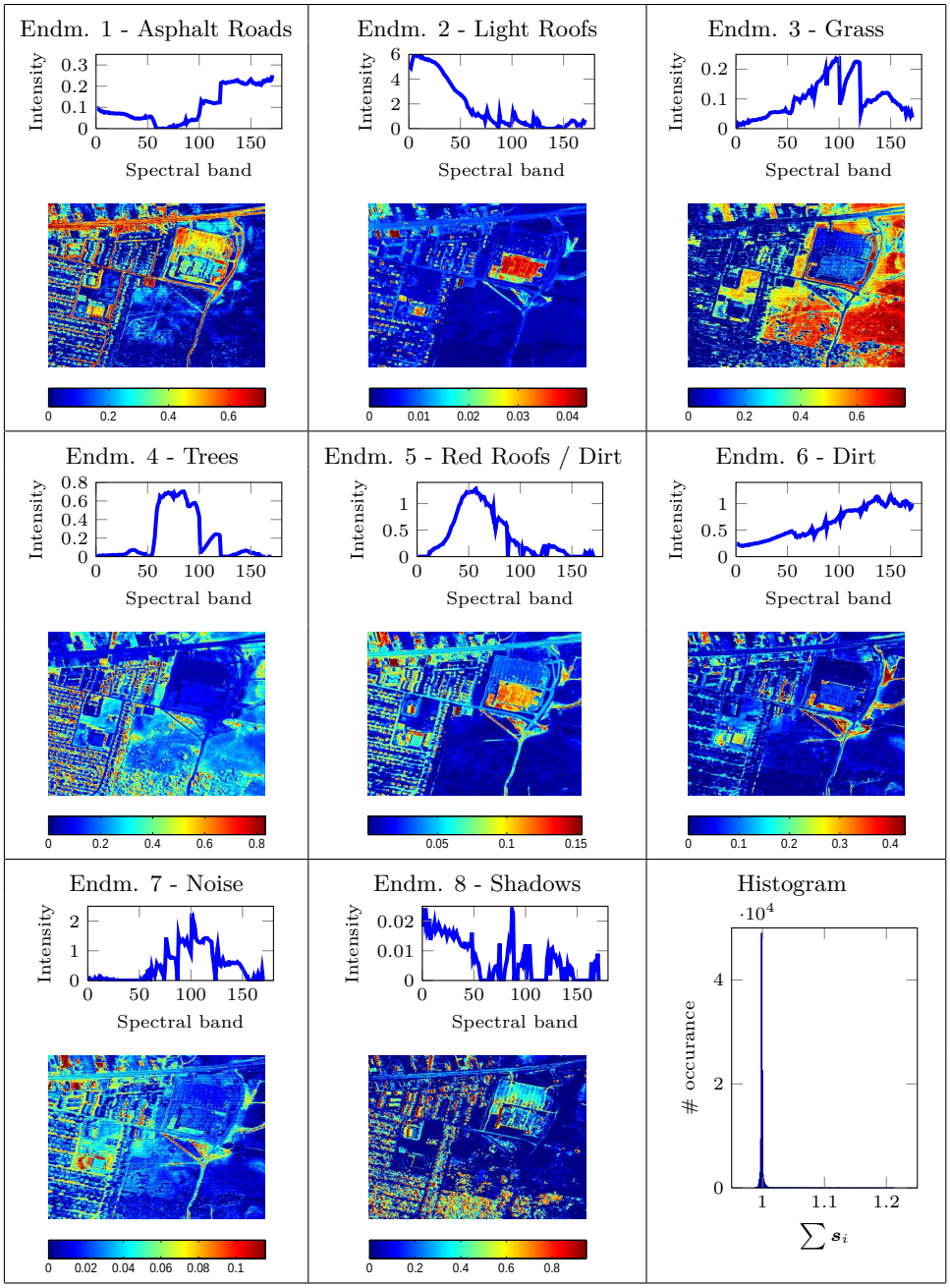


Figure 3.20: The ℓ_q unmixing results using the Urban data set with the ASC enforced. The endmembers are labeled by visually comparing the abundance maps to the RGB image. The plot in the lower right corner shows a histogram of the sum of the columns of $\boldsymbol{S}$.

Cuprite Data

Now the method is evaluated using the Cuprite data set. This site has received much attention by researchers and several well documented minerals are exposed in the landscape. These minerals are included in the USGS spectral library which holds several hundred spectral signatures. However, as with the Urban data set, a ground truth for this data set is not available.

The authors of [150] estimate 14 endmembers using the HySime [101] algorithm. The same number of endmembers are used here, that is, $r = 14$. The energy contained in the 14 first eigenvalues is 99.88% of the total signal energy.

The same methods as used in the previous section are used to estimate q and h. The SNR of the Cuprite data set is estimated to be 39 dB. The parameter are set to $q = 0.95$ and $h = 0.001$. These values minimize the metrics when the SNR of the simulation data is 40 dB. We will not enforce ASC so $\delta = 0$ and we set $\gamma = 0.002$.

To initialize $\boldsymbol{A}$ the image is first downsampled and then 14 endmembers are chosen from the downsampled image that well represent the original data, in the least squares sense. $\boldsymbol{S}$ is initialized with nonnegative regression. Low SNR bands, [1, 2, 105-115, 150-170, 223-224] are removed from the data set. This yield 188 usable bands out of the original 224.

In Figure 3.21, three pairs of abundance maps and their corresponding endmembers are shown. These are the abundance/endmember pairs that have the highest energy. The signature of the material in the USGS library that has the lowest SAD from the endmembers found is also shown. The signatures from the library are Muscovite, Kaolinite-Smectite and Sphene. All of these material have been identified to be in the area [37, 151, 152].

3.9 Conclusions

In this chapter an iterative cyclic descent algorithm, using an MM technique, for unmixing hyperspectral data is presented. An ℓ_q sparsity penalty, where $0 \leq q \leq 1$, is used to promote sparseness in the abundances. Since the endmembers vary smoothly in wavelength, a first order roughness penalty is used to promote smooth endmembers. Both the degree of sparseness and the roughness can be varied via parameters. The roughness penalty can also be canceled at known discontinuities in the endmember spectra. The ASC can also be softly enforced using matrix augmentation.

The algorithm was tested using simulated data and the optimal parameters that minimized the reconstruction errors, and the spectral angle distance were found. How the optimal parameters changed, when varying the SNR and the sparsity of the data was examined.

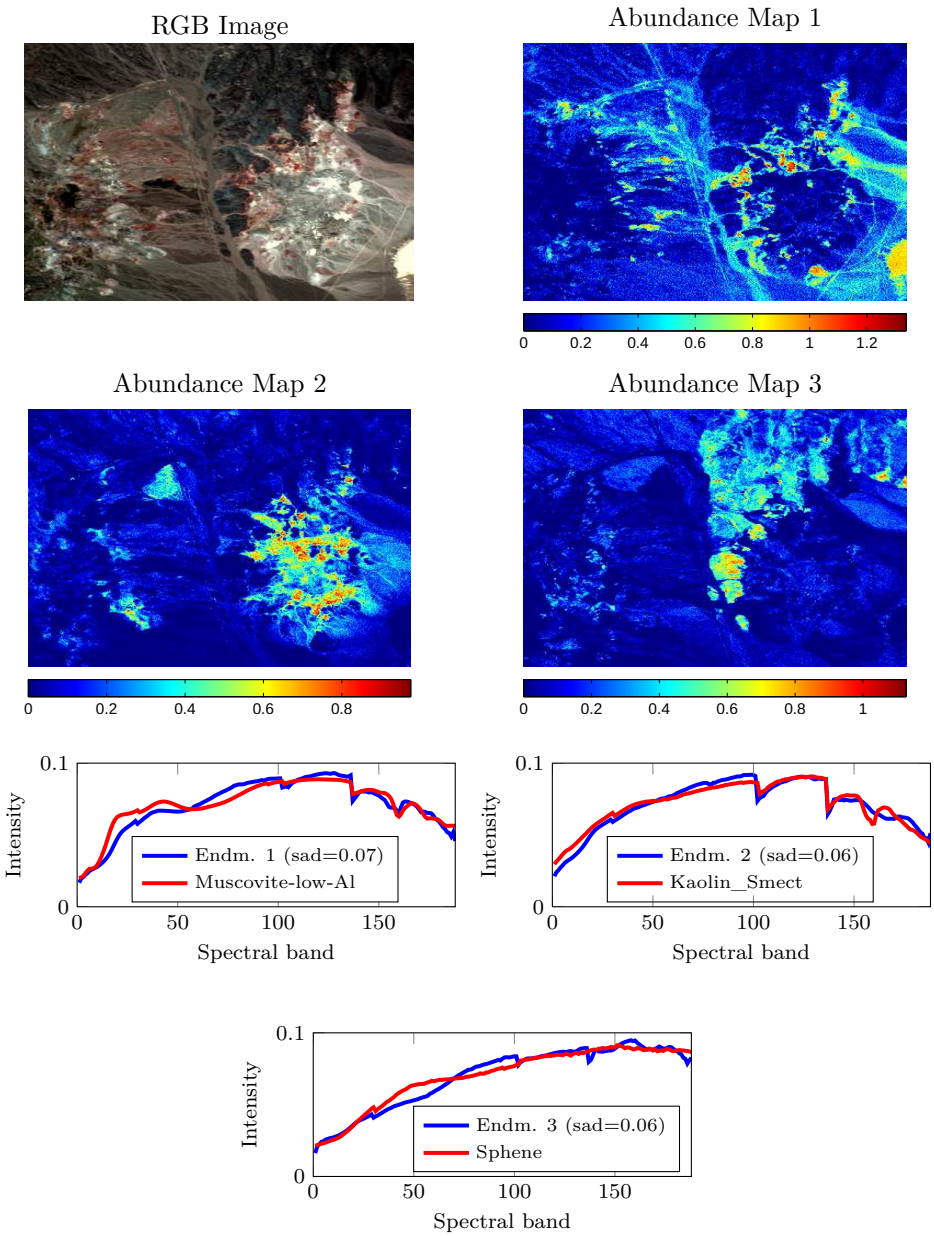
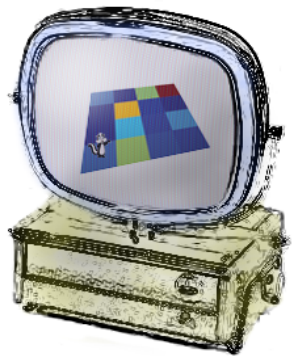


Figure 3.21: The ℓ_q unmixing results for the Cuprite data set. The top four images show the RGB image and three abundance maps. The three following graphs show the corresponding endmembers and the signatures of the material in the USGS library that had the lowest SAD from the endmembers.

The simulations showed that choosing $0 < q < 1$ could outperform the ℓ_1 norm, both in terms of reconstruction errors and spectral angle distance.

In the simulations, it was observed that low values of q resulted in slightly better unmixing results when the sparsity was increased. It was also noted that the value of q that gave the best unmixing results decreased when the SNR of that data decrease. Simulations also showed that in low SNR scenarios, the first order roughness penalty could improve the unmixing results.

The algorithm was also tested using the Urban and Cuprite data sets. The Urban data set was unmixed, both when the ASC was enforced and when it was not enforced. These two results were compared and it was concluded that not enforcing the ASC yielded better unmixing of this data set. The endmembers and their abundance maps in both solutions were easily associated with clearly visible distinct material in the RGB image. The Cuprite data set was unmixed and the most prominent endmembers found were compared to a spectral library.



Chapter 4

Hyperspectral Unmixing Using TV and ℓ_q Regularization

In the previous chapter the sparse nature of the abundance maps was exploited to aid the unmixing. In this chapter, the spatial characteristics of the abundance maps are also used to further aid in the unmixing. A new cyclic descent algorithm for blind hyperspectral unmixing using total variation (TV) and ℓ_q sparse regularization is developed. The TV regularizer is used to encourage piecewise smooth images and the ℓ_q regularizer promotes sparsity. The dyadic expansion decouples the problem making a simple cyclic descent procedure possible, where one abundance map is estimated, followed by the estimation of one endmember. The TV regularizer is used to improve the method from the previous chapter. The effects of the tuning parameters are examined using simulated data. The method is evaluated using real hyperspectral images. Finally, the proposed method is used for classification. The abundance maps are used as features for a Random Forest classifier, with the tuning and model parameters chosen using cross-validation.

4.1 Introduction

Natural images are often composed of large monotone objects, e.g., large buildings, vegetation fields. For this reason, neighboring pixels in hyperspectral images often represent the same material. To exploit this characteristic, researchers have introduced spatial regularization into the unmixing process. One such regularizer is the TV regularizer [99, 100], which has been widely used in image processing [153–158]. The TV regularizer promotes spatial homogeneity by penalizing differences in neighboring pixels while conserving discontinuities [159].

Spatially smooth hyperspectral images translate into smooth abundance maps. If only one specific material is locally abundant, then one of the abundance maps should have relatively high intensity value pixels in that area while the other abundance maps should have low or zero values in the same area.

As was discussed in the introduction of Chapter 3, the abundance maps are inherently sparse, so the ℓ_q sparsity inducing regularizer that was used in Chapter 3 is also used in the method presented here.

The ASC is often used in hyperspectral unmixing. The ASC forces the sum of the abundances for each pixel to be one. This means that each pixel in the image should be completely explained by the model. Adding the ASC to blind unmixing methods

also avoids the degenerate solution. Unmixing algorithms are however not always able to account for every material in an image and hyperspectral images contain noise. For these reasons it is not obvious whether it is more appropriate to relax the ASC or not use is at all and consider it as a part of the modeling error [2]. Also, the ASC does not take into account variations in reflectance from the same material. This can be caused by different incident reflectance angles or by shadowed areas. Despite these shortcomings, the ASC has been widely used in many unmixing methods.

Another constraint that also avoids the degenerate solution is the endmember norm constraint (ENC). This constraint puts a unit norm constraint on the endmembers. The ENC constrains the endmembers to have the same energy and also stabilizes the proposed algorithm.

4.2 Hyperspectral Unmixing

The hyperspectral model used is the low rank model (2.4), that is,

$$\boldsymbol{y}_{(p)} = \boldsymbol{A}\boldsymbol{s}_{(p)} + \boldsymbol{n}_{(p)}, \qquad p = 1, \ldots, P. \tag{4.1}$$

Using the definitions for all variables in Section 2.1, the matrix representation of the model is

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{S} + \boldsymbol{N}. \tag{4.2}$$

The matrices, $\boldsymbol{A}$ and $\boldsymbol{S}$, in (4.2), are found by solving,

$$\begin{aligned} &\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{S}} = \arg\min J(\boldsymbol{A}, \boldsymbol{S}) \\ &\text{s.t. } \boldsymbol{S} \geq 0, \boldsymbol{A} \geq 0 \text{ and } \|\boldsymbol{a}_{(n)}\|^2 = 1, \ n = 1, \ldots, r, \end{aligned} \tag{4.3}$$

where the cost function is given by

$$J(\boldsymbol{A}, \boldsymbol{S}) = \frac{1}{2} \sum_{p=1}^{P} \|\boldsymbol{y}_{(p)} - \boldsymbol{A}\boldsymbol{s}_{(p)}\|^2 + h_q \sum_{n=1}^{r} P_q(\boldsymbol{s}_n) + \gamma \sum_{n=1}^{r} \mathrm{TV}(\boldsymbol{s}_n). \tag{4.4}$$

Here, $0 \leq q \leq 1$, and the first penalty term is the ℓ_q regularizer, where

$$P_q(\boldsymbol{s}_n) = \begin{cases} \sum_p |s_{np}|_0 & \text{if } q = 0, \\ \sum_p |s_{np}|^q & \text{if } 0 < q \leq 1, \end{cases} \tag{4.5}$$

and

$$h_q = \begin{cases} \frac{h^2}{2} & \text{if } q = 0, \\ h^{2-q} & \text{if } 0 < q \leq 1. \end{cases}$$

In (4.5), $|s_{np}|_0$ is equal to one if s_{np} is nonzero, and zero otherwise. By adding the unit norm constraint on the columns of $\boldsymbol{A}$ (the ENC) we ensure that the degenerate solution is avoided.

The last term in (4.4) is the isotropic [159] TV regularizer, given by

$$\mathrm{TV}(\boldsymbol{s}_n^{\mathsf{T}}) = \|(\boldsymbol{z}_{xn}, \boldsymbol{z}_{yn})\| = \sum_{i,j} \sqrt{z_{xn,ij}^2 + z_{yn,ij}^2}, \tag{4.6}$$

where $\boldsymbol{z}_{xn} = \nabla_x \boldsymbol{s}_n$ and $\boldsymbol{z}_{yn} = \nabla_y \boldsymbol{s}_n$.

The operators ∇_x and ∇_y are linear operators corresponding to horizontal and vertical first order differences, respectively. The operation, $\nabla_x \boldsymbol{s}_n$, calculates the horizontal differences of the column vector $\boldsymbol{s}_n$, where $\boldsymbol{s}_n$ is first reshaped into a 2-dimensional array, i.e., the inverse operation of vec(). Then the horizontal differences are calculated for each pixel in the image.

4.3 Dyadic Cyclic Descent

There is no closed form solution available for solving (4.3), so a straight-forward approach could be to develop a two stage cyclic descent approach:

- Abundance Step: Fix $\boldsymbol{A}$ and estimate $\boldsymbol{S}$.
- Endmember Step: Fix $\boldsymbol{S}$ and estimate $\boldsymbol{A}$.

This approach has been used to solve similar optimization problems [76,122–124,130]. A different approach is taken here, where one column in $\boldsymbol{A}$ is estimated, followed by one row in $\boldsymbol{S}$. This is done by using the dyadic expansion [160], given by

$$\boldsymbol{A}\boldsymbol{S} = \sum_{n=1}^{r} \boldsymbol{a}_{(n)} \boldsymbol{s}_n^{\mathsf{T}}. \tag{4.7}$$

Using (4.7), an r-step cyclic descent procedure is possible. To describe the cyclic descent procedure, we first define the marginal cost function

$$J_n(\boldsymbol{a}, \boldsymbol{s}) = \frac{1}{2}\|\boldsymbol{R}_n - \boldsymbol{a}\boldsymbol{s}^{\mathsf{T}}\|^2 + h_q P_q(\boldsymbol{s}) + \gamma \mathrm{TV}(\boldsymbol{s}), \tag{4.8}$$

where $\boldsymbol{R}_n = \boldsymbol{Y} - \boldsymbol{A}_{\text{-}n}\boldsymbol{S}_{\text{-}n}$.

The main observation regarding the cyclic descent algorithm is that when each variable is fixed except an endmember $\boldsymbol{a}$, then (4.3) is equivalent to

$$\begin{aligned} &\boldsymbol{a} = \arg\min_{\boldsymbol{a}} J_n(\boldsymbol{a}, \boldsymbol{s}) \\ &\text{s.t. } \boldsymbol{a} \geq 0,\ \|\boldsymbol{a}\|^2 = 1,\ n = 1, \ldots, r. \end{aligned}$$

and when each variable is fixed except an abundance $\boldsymbol{s}$, then (4.3) is equivalent to

$$\begin{aligned} &\boldsymbol{s} = \arg\min_{\boldsymbol{s}} J_n(\boldsymbol{a}, \boldsymbol{s}) \\ &\text{s.t } \boldsymbol{s} \geq 0,\ n = 1, \ldots, r. \end{aligned}$$

4.3.1 Endmember Step

The endmember estimation step is the task of estimating $\boldsymbol{A}$, i.e., minimizing

$$\begin{aligned} \boldsymbol{a}_{(n)}^{(k+1)} &= \arg\min_{\boldsymbol{a}_{(n)}^{(k)}} \|\boldsymbol{R}_n - \boldsymbol{a}_{(n)}^{(k)}(\boldsymbol{s}_n^{(k)})^{\mathsf{T}}\|^2 \\ \text{s.t. } & \boldsymbol{a}_{(n)}^{(k)} \geq 0,\ \|\boldsymbol{a}_{(n)}^{(k)}\|^2 = 1,\ n = 1, \ldots, r. \end{aligned} \tag{4.9}$$

Using Lagrange multiplier theory, the solution is

$$\boldsymbol{a}_{(n)}^{(k+1)} = \max\left(0, \frac{\boldsymbol{R}_n \boldsymbol{s}_n^{(k)}}{\left\|\boldsymbol{R}_n \boldsymbol{s}_n^{(k)}\right\|}\right). \tag{4.10}$$

4.3.2 Abundance Step

Estimating the abundances is equivalent to minimizing

$$\begin{aligned} \boldsymbol{s}_n^{(k+1)} &= \arg\min_{\boldsymbol{s}_n^{(k)}} \frac{1}{2}\|\boldsymbol{R}_n - \boldsymbol{a}_{(n)}^{(k+1)}\boldsymbol{s}_n^{(k)}\|^2 + h_q P_q(\boldsymbol{s}_n^{(k)}) + \gamma \mathrm{TV}(\boldsymbol{s}_n^{(k)}) \\ &= \arg\min_{\boldsymbol{s}_n^{(k)}} \frac{1}{2}\|(\boldsymbol{a}_{(n)}^{(k+1)})^{\mathsf{T}}\boldsymbol{R}_n - \boldsymbol{s}_n^{(k)}\|^2 + h_q P_q(\boldsymbol{s}_n^{(k)}) + \gamma \mathrm{TV}(\boldsymbol{s}_n^{(k)}), \\ \text{s.t. } & \boldsymbol{s}_n^{(k)} \geq 0,\ n = 1, ..., r. \end{aligned} \tag{4.11}$$

To simplify the notation, let $\boldsymbol{u} = \boldsymbol{s}_n^{(k)}$ and $\boldsymbol{f} = (\boldsymbol{a}_{(n)}^{(k+1)})^{\mathsf{T}}\boldsymbol{R}_n$. The minimization problem can thus be written as

$$\begin{aligned} \boldsymbol{u}^{(k+1)} &= \arg\min_{\boldsymbol{u}} \frac{1}{2}\|\boldsymbol{f} - \boldsymbol{u}\|^2 + h_q P_q(\boldsymbol{u}) + \gamma \mathrm{TV}(\boldsymbol{u}) \\ &= \arg\min_{\boldsymbol{u}} \frac{\mu}{2}\|\boldsymbol{f} - \boldsymbol{u}\|^2 + h_q' P_q(\boldsymbol{u}) + \mathrm{TV}(\boldsymbol{u}), \\ &\text{s.t. } \boldsymbol{u} \geq 0, \end{aligned} \tag{4.12}$$

where $\mu = 1/\gamma$ and $h_q' = h_q/\gamma$. Having the minimization problem in the form given in (4.12) makes it easy to use a modified version the split Bregman algorithm for TV denoising from [158]. The method presented in [158] solves (4.12) without the sparsity term. To solve (4.12), we combine the TV denoising method from [158] with the method proposed in [143] for ℓ_q regularization.

The split Bregman formulation of the problem is

$$\begin{aligned} \min_{\boldsymbol{u},\boldsymbol{d}_x,\boldsymbol{d}_y} & \frac{\mu}{2}\|\boldsymbol{u} - \boldsymbol{f}\|^2 + \|(\boldsymbol{d}_x, \boldsymbol{d}_y)\| + \frac{\lambda}{2}\|\boldsymbol{d}_x - \nabla_x \boldsymbol{u} - \boldsymbol{b}_x\|^2 \\ & + \frac{\lambda}{2}\|\boldsymbol{d}_y - \nabla_y \boldsymbol{u} - \boldsymbol{b}_y\|^2 + h_q' P_q(\boldsymbol{u}). \end{aligned}$$

Defining $\boldsymbol{m}_x = \boldsymbol{d}_x - \boldsymbol{b}_x$ and $\boldsymbol{m}_y = \boldsymbol{d}_y - \boldsymbol{b}_y$, then estimating $\boldsymbol{u}$ requires solving

$$\begin{aligned}\boldsymbol{u}^{(k+1)} &= \arg\min_{\boldsymbol{u}} \frac{\mu}{2}\|\boldsymbol{u}-\boldsymbol{f}\|^2 + \frac{\lambda}{2}\|\nabla_x\boldsymbol{u}-\boldsymbol{m}_x\|^2 + \frac{\lambda}{2}\|\nabla_y\boldsymbol{u}-\boldsymbol{m}_y\|^2 + h_q' P_q(\boldsymbol{u}) \\ &= \arg\min_{\boldsymbol{u}} \frac{1}{2}\mathrm{tr}\Big\{\boldsymbol{u}^\mathsf{T}(\mu\boldsymbol{I}+\lambda\Delta)\boldsymbol{u} - 2\mu\boldsymbol{u}^\mathsf{T}\boldsymbol{f} - 2\lambda\boldsymbol{u}^\mathsf{T}(\nabla_x^\mathsf{T}\boldsymbol{m}_x + \nabla_y^\mathsf{T}\boldsymbol{m}_y)\Big\} \\ &\quad + h_q' P_q(\boldsymbol{u}), \end{aligned} \tag{4.13}$$

where $\Delta = (\nabla_x^\mathsf{T}\nabla_x + \nabla_y^\mathsf{T}\nabla_y)$.

To develop a fast iterative algorithm that solves this minimization problem, the Gauss-Seidel method, and thresholding is used. The Gauss-Seidel method involves estimating one entry in $\boldsymbol{u}$ while holding all other entries fixed. Using the following definitions,

$$\begin{aligned}[\Delta\boldsymbol{u}]_{i,j} &= 4u_{i,j} - (u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1}) \\ &= 4u_{i,j} - v_{i,j}, \\ \Delta_x^\mathsf{T}\boldsymbol{m}_{x,i,j} &= m_{x,i-1,j} - m_{x,i,j}, \\ \Delta_y^\mathsf{T}\boldsymbol{m}_{y,i,j} &= m_{y,i,j-1} - m_{y,i,j}, \\ w_{i,j} &= \Delta_x^\mathsf{T}\boldsymbol{m}_{x,i,j} + \Delta_y^\mathsf{T}\boldsymbol{m}_{y,i,j} \\ \boldsymbol{v} = [v],\ \boldsymbol{w} &= [w], \end{aligned}$$

the minimization problem can be written as

$$u_{i,j}^{(k+1)} = \arg\min_u \frac{1}{2}\|u_{i,j}^{(k)} - g_{i,j}^{(k)}\|^2 + \frac{h_q'}{\mu+4\lambda}P_q(u_{i,j}^{(k)}), \tag{4.14}$$

where

$$g_{i,j}^{(k)} = \frac{1}{\mu+4\lambda}\Big(\mu f_{i,j} + \lambda(w_{i,j}^{(k)} + v_{i,j}^{(k)})\Big). \tag{4.15}$$

- When $q = 0$, the solution to (4.14) is hard thresholding, that is

$$u_{i,j}^{(k+1)} = \begin{cases} g_{i,j}^{(k)} & \text{if } g_{i,j}^{(k)} > \sqrt{\frac{2h_q'}{\mu+4\lambda}}, \\ 0 & \text{otherwise.} \end{cases} \tag{4.16}$$

- When $0 < q < 1$ the solution is found using the thresholding method in [143], slightly modified to enforce nonnegativity.

- When $q = 1$, the solution is soft-thresholding,

$$u_{i,j}^{(k+1)} = \max\left(g_{i,j}^{(k)} - \frac{h_q'}{\mu+4\lambda}, 0\right). \tag{4.17}$$

The updates for $(\boldsymbol{d}_x^{(k+1)}, \boldsymbol{d}_y^{(k+1)})$ are [161]

$$\begin{aligned}\boldsymbol{d}_x^{(k+1)} &= \max(\boldsymbol{s}^{(k)} - 1/\lambda, 0)\frac{\nabla_x\boldsymbol{u}^{(k)} + \boldsymbol{b}_x^{(k)}}{\boldsymbol{s}^{(k)}}, \\ \boldsymbol{d}_y^{(k+1)} &= \max(\boldsymbol{s}^{(k)} - 1/\lambda, 0)\frac{\nabla_y\boldsymbol{u}^{(k)} + \boldsymbol{b}_y^{(k)}}{\boldsymbol{s}^{(k)}}, \end{aligned}$$

where

$$\boldsymbol{s}^{(k)} = \left\| \left(\nabla_x \boldsymbol{u}^{(k)} + \boldsymbol{b}_x^{(k)}, \nabla_y \boldsymbol{u}^{(k)} + \boldsymbol{b}_y^{(k)} \right) \right\|.$$

The update rules for $\boldsymbol{b}_x$ and $\boldsymbol{b}_y$ are

$$\begin{aligned} \boldsymbol{b}_x^{(k+1)} &= \boldsymbol{b}_x^{(k)} + (\nabla_x \boldsymbol{u}^{(k+1)} - \boldsymbol{d}_x^{(k+1)}), \\ \boldsymbol{b}_y^{(k+1)} &= \boldsymbol{b}_y^{(k)} + (\nabla_y \boldsymbol{u}^{(k+1)} - \boldsymbol{d}_y^{(k+1)}). \end{aligned}$$

The minimization algorithm is given in Algorithm 5.

Algorithm 5: Split Bregman ℓ_q sparse total variation algorithm. The type of thresholding done by the threshold$_q(\boldsymbol{g}^{(k)})$, depends on q.

Require $\boldsymbol{f}$, λ, q and h
Initialize $\boldsymbol{u}^{(0)} = \boldsymbol{f}$, $\boldsymbol{d}_x = \boldsymbol{d}_y = \boldsymbol{b}_x = \boldsymbol{b}_y = 0$, $k = 0$
while $\|\boldsymbol{u}^{(k+1)} - \boldsymbol{u}^{(k)}\| > tol$ **do**
  $\boldsymbol{u}^{(k+1)} = \text{threshold}_q(\boldsymbol{g}^{(k)})$
  $\boldsymbol{d}_x^{(k+1)} = \max(\boldsymbol{s}^{(k)} - 1/\lambda, 0) \frac{\nabla_x \boldsymbol{u}^{(k)} + \boldsymbol{b}_x^{(k)}}{\boldsymbol{s}^{(k)}}$
  $\boldsymbol{d}_y^{(k+1)} = \max(\boldsymbol{s}^{(k)} - 1/\lambda, 0) \frac{\nabla_y \boldsymbol{u}^{(k)} + \boldsymbol{b}_y^{(k)}}{\boldsymbol{s}^{(y)}}$
  $\boldsymbol{b}_x^{(k+1)} = \boldsymbol{b}_x^{(k)} + (\nabla_x \boldsymbol{u}^{(k+1)} - \boldsymbol{d}_x^{(k+1)})$
  $\boldsymbol{b}_y^{(k+1)} = \boldsymbol{b}_y^{(k)} + (\nabla_y \boldsymbol{u}^{(k+1)} - \boldsymbol{d}_y^{(k+1)})$
  $k = k + 1$
end

4.4 Proposed Cyclic Descent Algorithm

The proposed hyperspectral unmixing algorithm for minimizing (4.4) is given in Algorithm 6. The algorithm alternates between estimating one column in $\boldsymbol{S}$, followed by one column in $\boldsymbol{A}$. These estimations are done using (4.10) and the algorithm given in Algorithm 5, respectively.

The algorithm is terminated when the number of allowed iterations has been reached or if the cost function (4.4) has converged to its final value, that is if

$$\frac{J^{(k-1)} - J^{(k)}}{J^{(k)}} < tol, \tag{4.18}$$

where tol is a predefined value. Estimating $\boldsymbol{S}$ is the most computationally intensive part of the algorithm. This is done in Algorithm 5, and is called at line 6 in the main algorithm. This part of the algorithm is implemented in C++. The C++ (Matlab MEX) implementation provided by [158], is modified to handle the ℓ_q regularization and used to estimate $\boldsymbol{S}$. Approximately 80% of the computation time is spent running this code. The rest of the algorithm is implemented in Matlab.

Algorithm 6: The algorithm for sparse ℓ_q hyperspectral unmixing using the dyadic expansion and total variation.

Require $\boldsymbol{Y}, r, \gamma, q$ and h
Initialize $\boldsymbol{A}^{(0)}$ and $\boldsymbol{S}^{(0)}$, $k = 0$
while *the stopping criteria is not met* **do**
  for $n = 1, ..., r$ **do**
    $$\boldsymbol{s}_n^{(k+1)} = \arg\min_{\boldsymbol{s}_n^{(k)}} J_n(\boldsymbol{a}_{(n)}^{(k+1)}, \boldsymbol{s}_n^{(k)}) \quad \text{s.t. } \boldsymbol{s}_n^{(k)} \geq 0$$
    $$\boldsymbol{a}_{(n)}^{(k+1)} = \arg\min_{\boldsymbol{a}_{(n)}^{(k)}} J_n(\boldsymbol{a}_{(n)}^{(k)}, \boldsymbol{s}_n^{(k+1)}) \quad \text{s.t. } \boldsymbol{a}_{(n)}^{(k)} \geq 0,\ \|\boldsymbol{a}_{(n)}^{(k)}\|^2 = 1$$
  end
  $k = k + 1$
end

4.5 Simulations

4.5.1 Simulation Data

A simulation data set is used to evaluate the proposed algorithm. Spectral signatures (endmembers) from the United States Geological Survey (USGS) digital spectral library[1] are used to create the true endmembers.

Some endmembers in the library are highly correlated so the library is pruned to contain 62 endmembers. The minimum SAD between endmembers, after pruning, is 0.17rad (10°). Each endmember was resampled from a length of 224 bands to 100 bands. Six endmembers are randomly selected from the library for the simulations. The endmembers are scaled to have unit norm ($\|\boldsymbol{a}_{(n)}\|^2 = 1$).

Each abundance map is generated as a checkerboard using 16 (4×4) squares of size 18×18 pixels. For each square in the image, an abundance vector is generated using a Dirichlet distribution. Each square is thus spatially homogenous. Spectral variability is also introduced into the abundance vectors by allowing the sum of each abundance vector to be between 0.9 and 1.1. The sparsity of the abundance maps is set to be approximately 0.35, so 35% of all entries in the abundance maps are set to zero. Gaussian i.i.d. noise is added to the simulation data according to (4.1). Example endmembers and two abundance maps are shown in Figure 4.1. The algorithm is initialized with vertex component analysis (VCA) [37].

[1] http://speclab.cr.usgs.gov/spectral.lib06

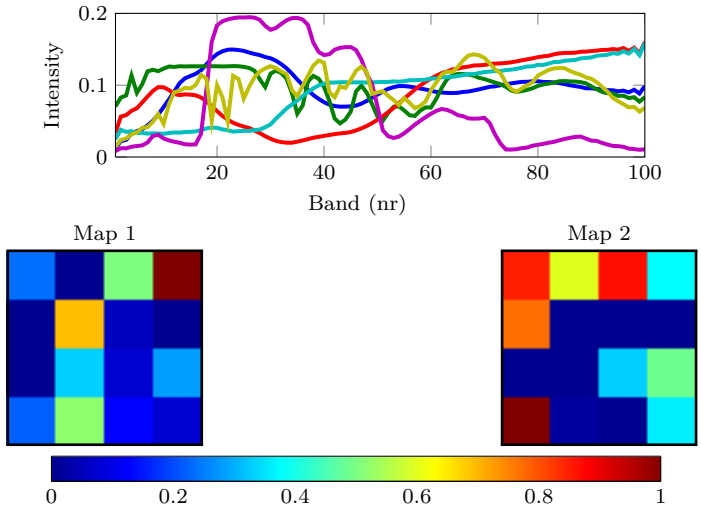


Figure 4.1: Examples of the signatures and two abundance maps used in the TV umixing simulations. Each checker in the images is 18×18 pixels and the number of checkers is 16. The horizontal bar shows the intensity values of the abundance maps.

4.5.2 Experiments

In the simulation experiments, the SNR is varied and the metrics defined in Appendix A are calculated for different values of the tuning parameters, q, h and γ. The number of endmembers is fixed at $r = 6$. For these experiments, it is assumed that the number of endmembers is known. Surface plot of the metrics are generated. Four values of q are considered for each SNR, $q \in \{0, 0.25, 0.5, 0.75, 1\}$. The values of h and γ are varied such that the minima of the metrics show up on the surface plots.

In figures 4.2 - 4.4, the surface plots of the metrics are shown when the SNR value is 20dB, 25dB and 30dB. The white '$\times$' marks the location of the surface minima. In Table 4.1, the values corresponding to the minima are shown. The bold values are the minimum values over all q. In the last two columns, the metrics calculated from the initial values of $\boldsymbol{A}$ and $\boldsymbol{S}$, and when both h and γ are set to zero are shown. Setting h and γ to zero amounts to using the method with no regularization.

The same values of h and γ are not guaranteed to minimize all three metrics, but the parameters that minimize the SAD and $\text{nMSE}_{\boldsymbol{S}}$ are often similar. Table 4.2 shows the metrics when h and γ are chosen to minimize the $\text{nMSE}_{\boldsymbol{S}}$. Choosing the parameter values that minimize $\text{nMSE}_{\boldsymbol{S}}$ is a good compromise, the SAD is very close to its minimum value and the $\text{nMSE}_{\boldsymbol{AS}}$ does not increase much, compared to the minimum value.

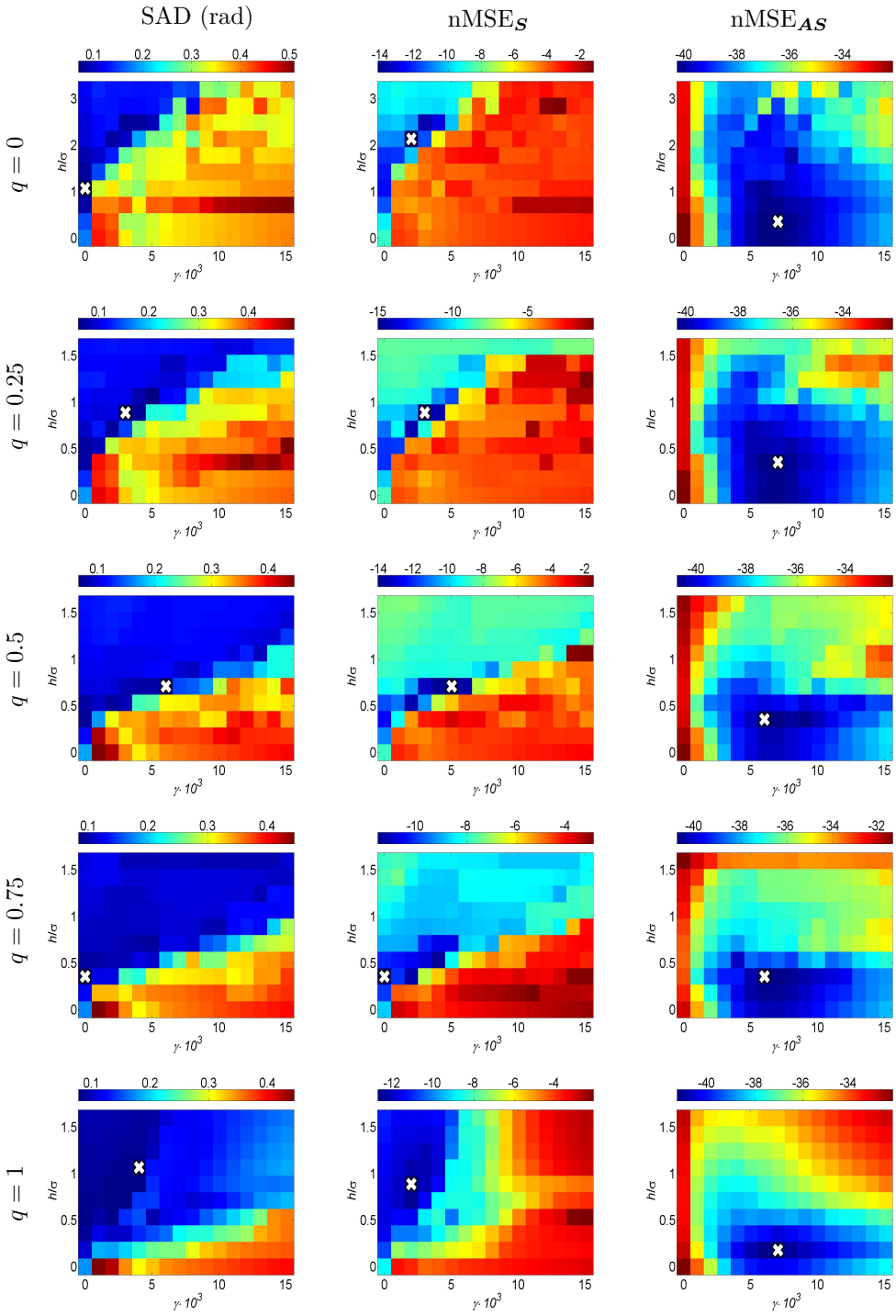


FIGURE 4.2: The metrics, calculated for various values of q, γ and h when the SNR is 20dB, in the TV unmixing simulations. The white '$\times$' marks the location of the surface minima.

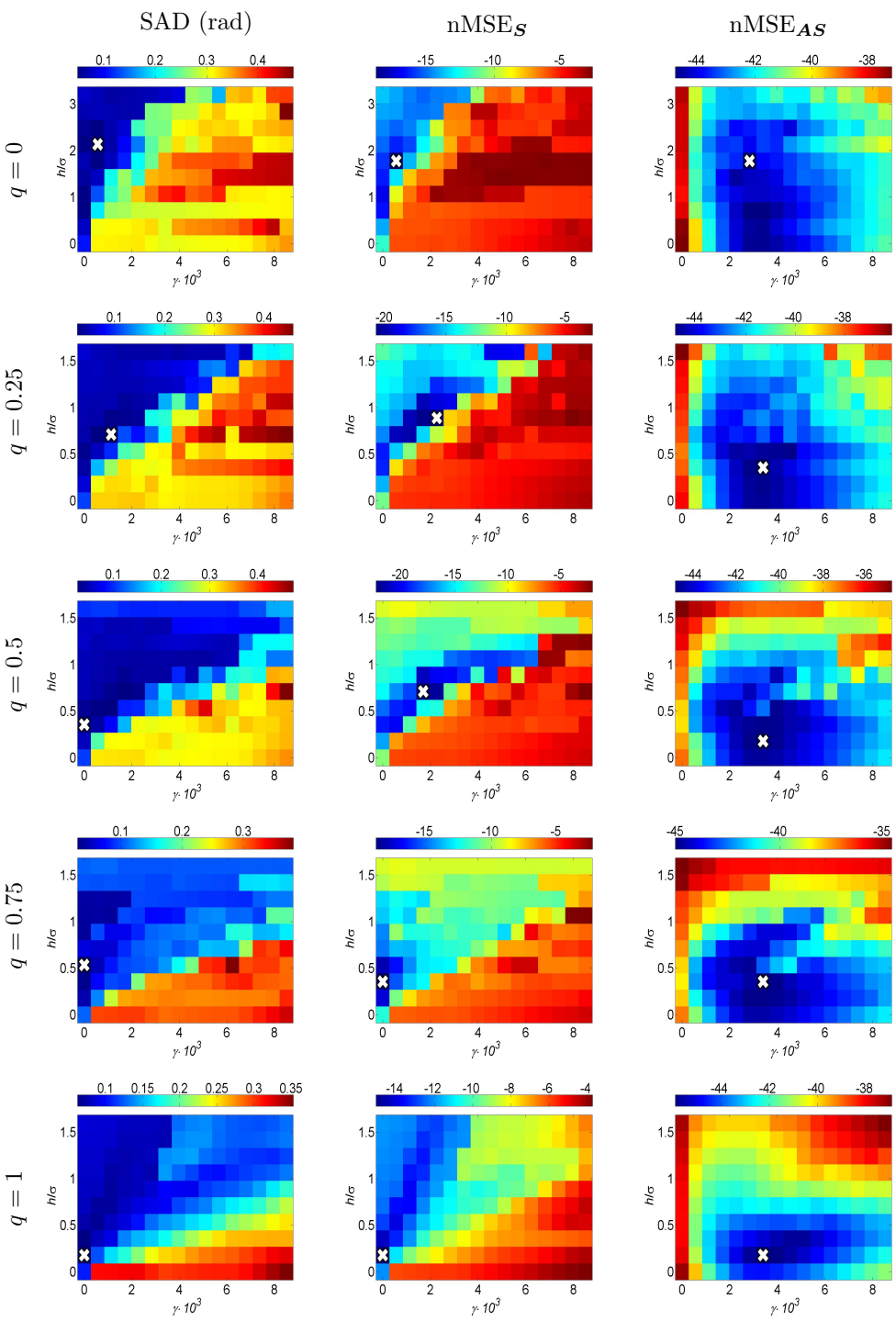


Figure 4.3: TV unmixing. The metrics, calculated for various values of q, γ and h when the SNR is 25dB, in the TV unmixing simulations. The white '×' marks the location of the surface minima.

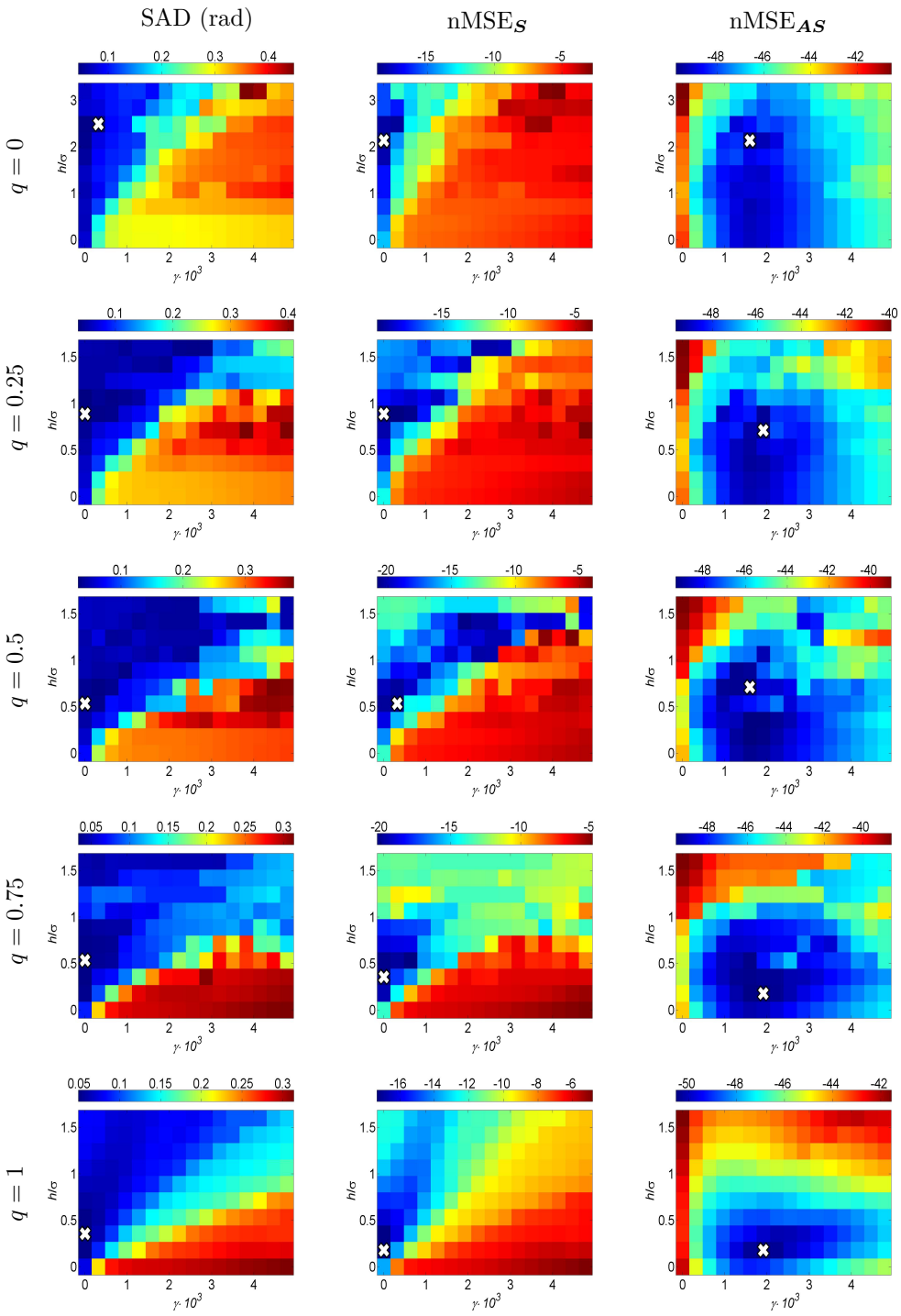


FIGURE 4.4: TV unmixing. The metrics, calculated for various values of q, k and h when the SNR is 30dB, in the TV unmixing simulations. The white '×' marks the location of the surface minima.

Table 4.1: The minimum values of the metrics, in the TV unmixing simulations. Bold values are the minimum values for each row.

SNR	Metric	q					Initial Values	$h=0$ $k=0$
		0	0.25	0.5	0.75	1		
20dB	SAD (rad)	0.068	**0.059**	0.066	0.083	0.078	0.111	0.176
	nMSE$_{\boldsymbol{S}}$ (dB)	-12.57	**-15.12**	-14.16	-11.58	-12.78	-9.77	-11.17
	nMSE$_{\boldsymbol{AS}}$ (dB)	-40.33	-40.40	-40.83	-40.81	**-41.32**	-29.16	-32.18
25dB	SAD (rad)	0.048	0.026	**0.025**	0.029	0.062	0.086	0.106
	nMSE$_{\boldsymbol{S}}$ (dB)	-18.86	-20.82	**-22.45**	-18.82	-15.06	-11.97	-11.97
	nMSE$_{\boldsymbol{AS}}$ (dB)	-44.80	-44.85	-44.92	-45.07	**-45.89**	-37.32	-37.19
30dB	SAD (rad)	0.044	**0.036**	0.032	0.033	0.049	0.081	0.076
	nMSE$_{\boldsymbol{S}}$ (dB)	-18.77	-19.98	**-20.84**	-20.27	-17.31	-12.30	-12.30
	nMSE$_{\boldsymbol{AS}}$ (dB)	-49.79	-49.64	-49.37	-49.72	**-50.53**	-41.56	-42.23

Table 4.2: The metrics when γ and h minimize nMSE$_{\boldsymbol{S}}$, in the TV unmixing simulations. Bold values are the minimum values for each row.

SNR	Metric	q				
		0	0.25	0.5	0.75	1
20dB	SAD (rad)	0.068	**0.059**	0.080	0.083	0.080
	nMSE$_{\boldsymbol{S}}$ (dB)	-12.57	**-15.12**	-14.16	-11.58	-12.78
	nMSE$_{\boldsymbol{AS}}$ (dB)	-32.89	-38.97	**-39.31**	-33.29	-37.33
25dB	SAD (rad)	0.051	0.036	**0.026**	0.033	0.062
	nMSE$_{\boldsymbol{S}}$ (dB)	-18.86	-20.82	**-22.45**	-18.82	-15.06
	nMSE$_{\boldsymbol{AS}}$ (dB)	-41.67	**-43.74**	-43.39	-38.44	-37.86
30dB	SAD (rad)	0.050	0.036	0.036	**0.035**	0.051
	nMSE$_{\boldsymbol{S}}$ (dB)	-18.77	-19.98	**-20.84**	-20.27	-17.31
	nMSE$_{\boldsymbol{AS}}$ (dB)	-42.57	-42.58	**-46.83**	-43.43	-42.80

By examining Table 4.2, it can be seen that choosing $q = 0.25$ gives the best results when the SNR is 20dB. Choosing $q = 0.5$ gives the best results when SNR is 25dB and 30dB, respectively. In all of these cases, the best result is obtained when both h and γ have nonzero values.

It is worth pointing out that choosing only sparsity, or only TV regularization, will not yield good values for the reconstruction error (nMSE$_{\boldsymbol{AS}}$), but a combination of sparsity and TV regularization ($h > 0$, $\gamma > 0$) can yield acceptable results for all metrics.

4.6 Real Hyperspectral Data

The proposed algorithm is evaluated using two real hyperspectral data sets. The first is the Urban data described in Appendix B.1 and the second is the Cuprite data set described in B.2.

Ground truth is not available for both of these data sets. However, the Cuprite data set has been examined extensively by researchers and several well documented minerals are exposed in the landscape. These minerals are included in the USGS spectral library, which holds several hundred spectral signatures. We compare the endmembers obtained to a subset of the spectral signatures in this library.

4.6.1 Urban

The Urban image is 307×307 pixels and the whole image is used. Bands numbered [89, 90, 103-109, 130-152, 204-210] are identified as water absorption or low SNR bands and are removed, resulting in 171 usable bands. An RGB image, generated using the hyperspectral data, is shown in Figure 4.5. The RGB image is split into two regions, marked as a and b. The RGB image is created by using specific spectral bands from the data set, to represent the red, green, and blue channels of the RGB image.

In the RGB image, at least eight different macroscopic materials are visible: asphalt, dirt, three different colored rooftops, two types of vegetation and shadow areas. The number of endmembers is therefore set to eight. The algorithm is initialized with VCA as was done in the simulation part.

To select the tuning parameters for this data set we observe that the parameter set $(q = 0.5, h/\sigma = 0.5, \gamma = 0.002)$ gave good results for all the simulations done in Section 4.5.2. These parameter values are used for the Urban data set.

No ground truth is available for this data set, so a qualitative evaluation is done. The obtained abundance maps are visually compared to the RGB image. The maps are also compared to the abundance maps obtained using no regularization. Five out of eight abundance maps are shown in Figure 4.5.

For the remainder of this paper the proposed sparse TV method is referred to as the SpTv method, while NMF refers to unmixing using no regularization.

The sparsity of the SpTv abundance maps is 17% compared to 6% using NMF. There are some important differences between the SpTv and NMF maps. The SpTv maps are sparser and can be better associated with specific material in the RGB image. Also, vertical artifacts that are clearly visible in the NMF maps are not as apparent in the SpTv maps.

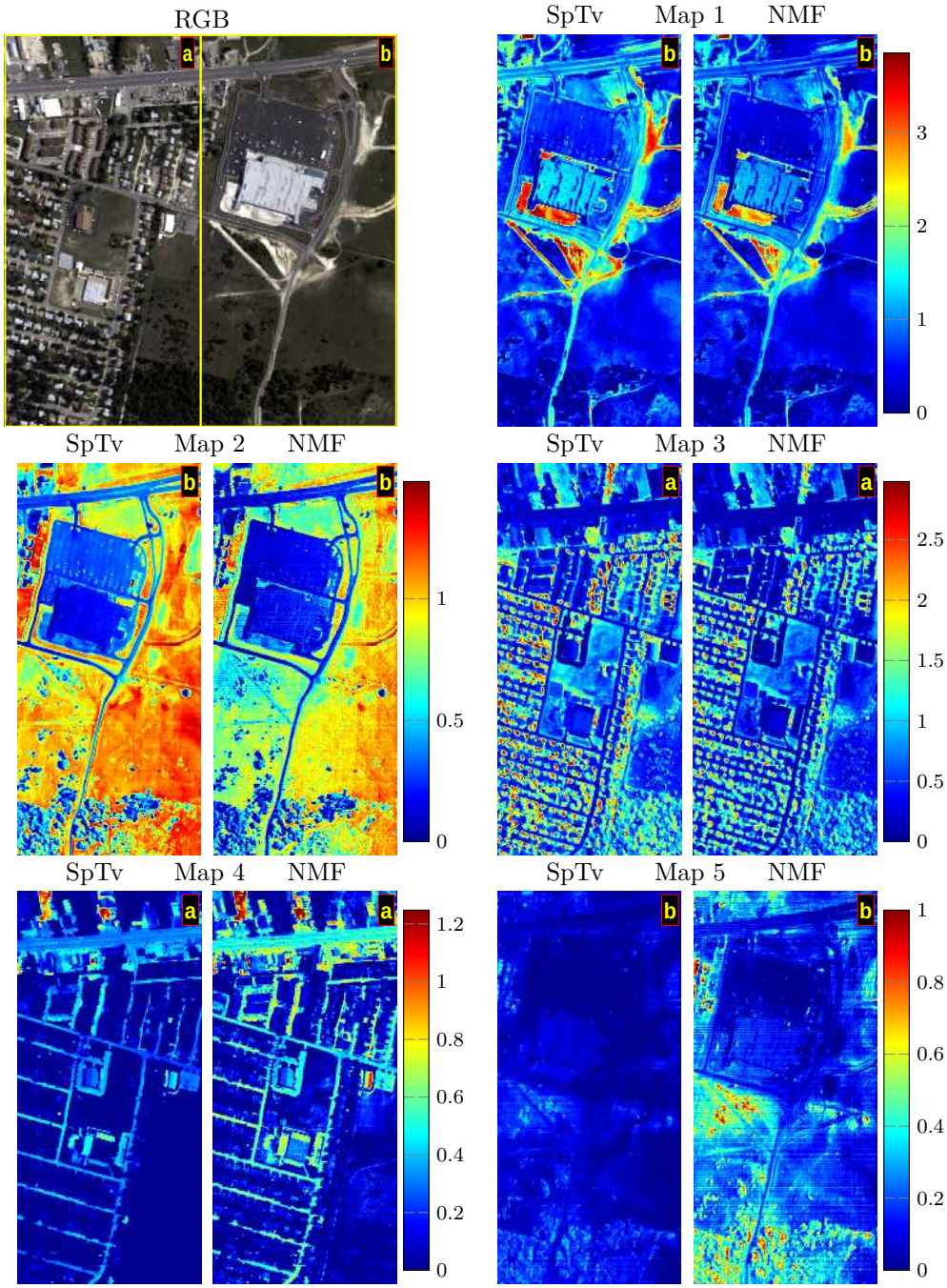


Figure 4.5: SpTv unmixing results. The top left image is the RGB generated image of the Urban data set, split into 2 regions marked 'a' and 'b'. The other images are five out of eight estimated abundance maps. The maps marked SpTv are the maps found with the proposed method, and the maps marked NMF are the maps found with no regularization.

The maps in Figure 4.5 labeled 1, represent dirt areas. The main difference between the SpTv and NMF map is that intensity of the SpTv map is higher at the locations were the dirt is visible in the RGB image. Horizontal artifacts that are visible in the NMF map are not visible in the SpTv map.

The maps labeled 2 are associated with grassy regions. The horizontal artifacts in the NMF map are much less apparent in the SpTv map and the intensity values at the grassy region are higher in the SpTv map.

The differences between the maps labeled 3 are not large but the intensity values of the SpTv map are slightly higher in the areas covered by trees in the RGB image.

The high intensity pixels in the map labeled 4 represent asphalt roads. The SpTv map is sparser than the NMF map and the high intensity pixels better reflect the asphalt roads seen in the RGB map.

The NMF map labeled 5 represents tree areas and it is highly corrupted with horizontal artifacts. The SpTv map however does not have these artifacts or any high value pixels. This SpTv map has therefore little impact on the whole SpTv solution and could be treated as noise while the NMF map cannot be treated as noise.

4.6.2 Cuprite

The method is now evaluated using the Cuprite data set detailed in Appendix B.2. A 300×150 pixel subset of the whole image is used here. In [37], the number of endmembers for this region is estimated to be 14 using the Virtual Dimensionality method [106]. The number of endmember is set to $r = 14$. As was done in with the Urban data set, the tuning parameters are set, $q = 0.5$, $h/\sigma = 0.5$ but the TV parameter is set to a lower value. The generated RGB image of the Cuprite data set does not exhibit the same spatially smooth characteristics of the Urban data set and the TV parameter is thus set as $\gamma = 2.5\text{e-}4$.

The algorithm is initialized with random values for the endmembers and the abundances are initialized using nonnegative regression. Low SNR bands numbered [1, 2, 105-115, 150-170, 222-224] are removed from the data set. This yields 187 usable bands out of the original 224 bands.

To evaluate the unmixing, the endmembers are compared to the USGS digital spectral library. This library contains laboratory spectra of minerals that are known to be in the area [151]. This library contains hundreds of spectra of different variations of minerals. Many of these spectra are very similar, so the library is pruned to contain only minerals that are known to be in the region [37, 151]. Different variations of the same mineral are also removed, to some extent, by constraining each spectra have at least a SAD of 0.05rad (3°) from all other spectra in the library. This leaves 245 signatures in the library.

Table 4.3: The SAD between the endmembers found, using SpTv unmixing, in the Cuprite data set and the library endmembers.

Endmember # / Mineral	SAD (rad)	
	SpTv	VCA
1. Kaolin_Smect	0.055	0.070
2. Kaol_wxl+others	0.040	0.065
3. Andradite WS487	0.053	(Andradite HS111.3B) 0.081
4. Andradite WS474	0.061	0.067
5. Montmorillonite+Illite	0.059	(Kaolin_Smect) 0.060
6. Andradite WS487	0.056	(Montmorillonite-Na) 0.042
7. Sphene HS189.3B	0.052	(Andradite WS474) 0.067
8. Andradite WS487	0.065	(Andradite WS474) 0.045
9. Alun_Na+Kaol+Hemat	0.044	(Kaolin_Smect) 0.066
10. Montmorillonite+Illite	0.047	0.048
11. Alunite.5+Musc	0.106	(Alunite_K CU98-5C) 0.065
12. Alunite0.35K+.65Na	0.110	(Kaol_wxl+others) 0.105
13. Dolomit+Calcite+Talc	0.141	(Alunite+Dickite) 0.048
14. Alunite HS295.3B	0.111	(Buddingtonite) 0.077

In Table 4.3, the SAD from each estimated endmember to its closest match in the library is shown. This is done for the proposed method (SpTv) and also for Vertex Component Analysis (VCA). If the closest library endmembers are not the same for both SpTv and VCA, the library endmembers that most resemble the VCA endmembers are shown within brackets. Both methods find library endmembers that are very close to the estimated endmembers. On four occasions, the methods find the same library spectra for an endmember (endmembers numbered 1, 2, 4, 10). On three occasions, the methods agree on the mineral but not on the spectra (endmembers numbered 3, 8, 11). On some occasions, the methods find the same library endmember for more than one estimated endmember, e.g., endmembers 3 and 6 both have the lowest SAD corresponding to *Andradite WS487*.

In Figures 4.6 and 4.7, six out of fourteen abundance maps and their corresponding endmembers are also shown. Figure 4.6 also shows a generated RGB image of the scene while Figure 4.7 also shows with the reconstruction error for each pixel in the image.

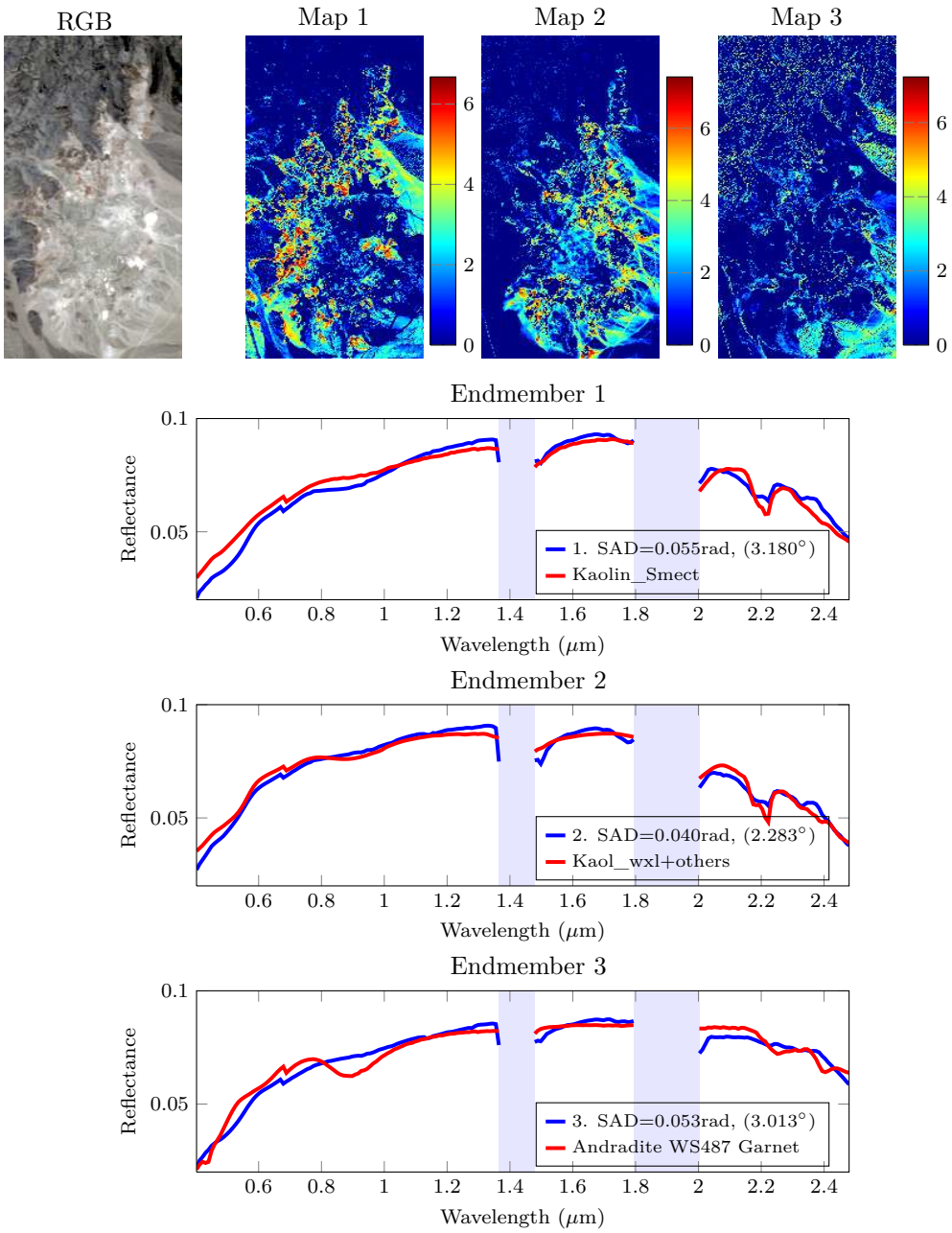


Figure 4.6: Unmixing results using SpTv. The first image in the top row is the RGB generated image of the Cuprite data set. The other three images in the top row are three out of fourteen abundance maps. The plots show the corresponding estimated endmembers and also library endmembers with the lowest SAD of the estimated endmembers. The shaded area shows where noisy bands were removed.

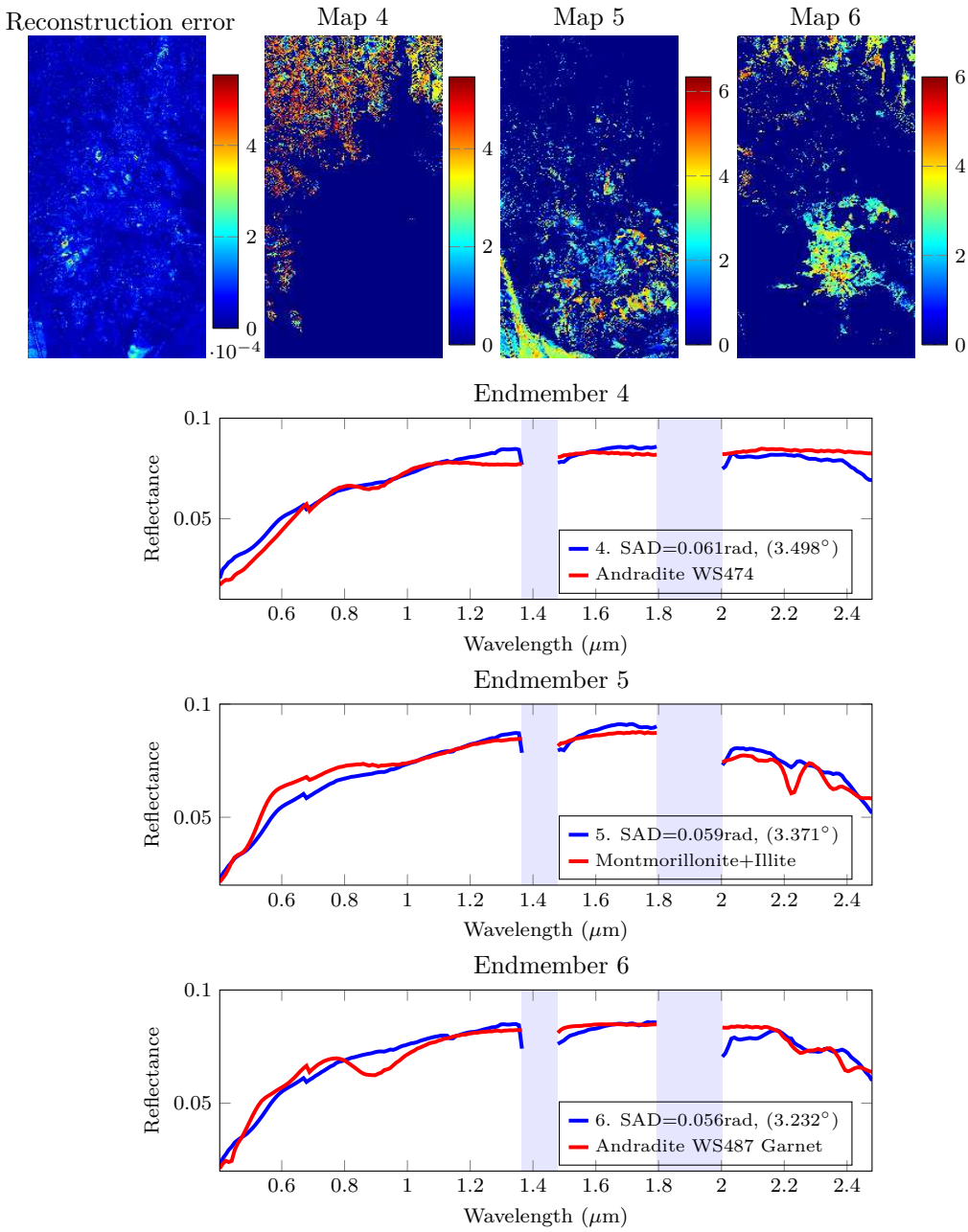


Figure 4.7: Unmixing results using SpTv. The first image in the top row is the reconstruction error. The other three images in the top row are three out of fourteen abundance maps. The plots show the corresponding estimated endmembers and also library endmembers with the lowest SAD of the estimated endmembers. The shaded area shows where noisy bands were removed.

4.7 Classification

The abundances created by the proposed sparse TV method will now be used as features for classification. All tuning parameters (r, h and γ), excluding q, are chosen using cross-validation. Here, $q = 1$ is used which results in ℓ_1 sparse regularization. A Random Forest classifier [162] is used to classify the features. To evaluate the method, two different data sets are used, the first is the Indian Pines and the second is the Pavia University data set. These data sets are detailed in Appendices B.3 and B.4, respectively.

4.7.1 Experimental Results

The classification results using the features obtained from the proposed sparse TV method (SpTv) are compared to other feature extraction methods. The other methods are PCA, unmixing using only TV (Tv) and unmixing using only sparsity (Sp).

The classification results using the features obtained from the proposed sparse TV method (SpTv) are compared to other feature extraction methods. The other methods are Principal Component Analysis (PCA), unmixing using only TV (Tv) and unmixing using only sparsity (Sp). All tuning and model parameters for these methods, except for q, are chosen using 5-fold cross-validation using only training data samples. For these tests, $q = 1$.

A Random Forest classifier using 200 trees is used to classify the features. The classification results are shown in tables 4.4 and 4.5. The training samples for the Indian Pines data set are chosen randomly for each experiment, but the training samples for the Pavia University dataset have been manually selected [163]. The accuracy for each class, the overall accuracy and the average accuracy are all given. The number of training and test samples are also shown, along with the κ coefficient and the number of features selected for each method. The classification maps, ground truth, and training samples, are shown in figures 4.8 and 4.9, respectively.

From the results given in tables 4.4 and 4.5 , it can be seen that using TV plays a key role in increasing the classification accuracy. The accuracy of the Indian Pines data set can be further increased by using sparsity. For the Pavia data set, cross-validation chose $h_q = 0$, so in that case only TV regularization is used.

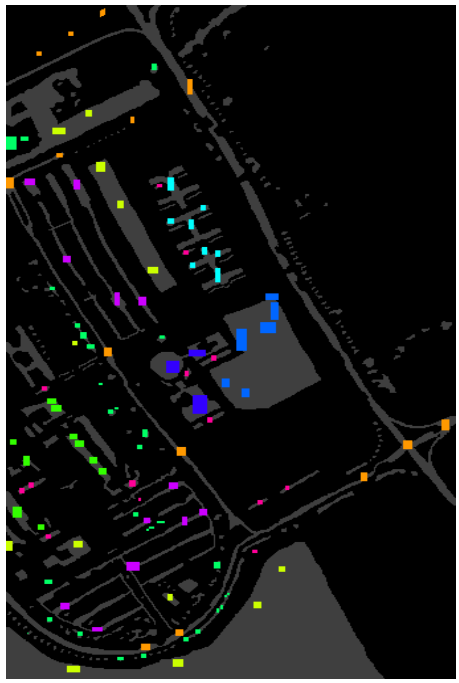
The features found by the SpTv method yield much better classification results than the features of PCA. This is accomplished using relatively few features.

Table 4.4: The classification accuracies for the Pavia University data set using the SpTv method. The bold numbers show the highest classification result.

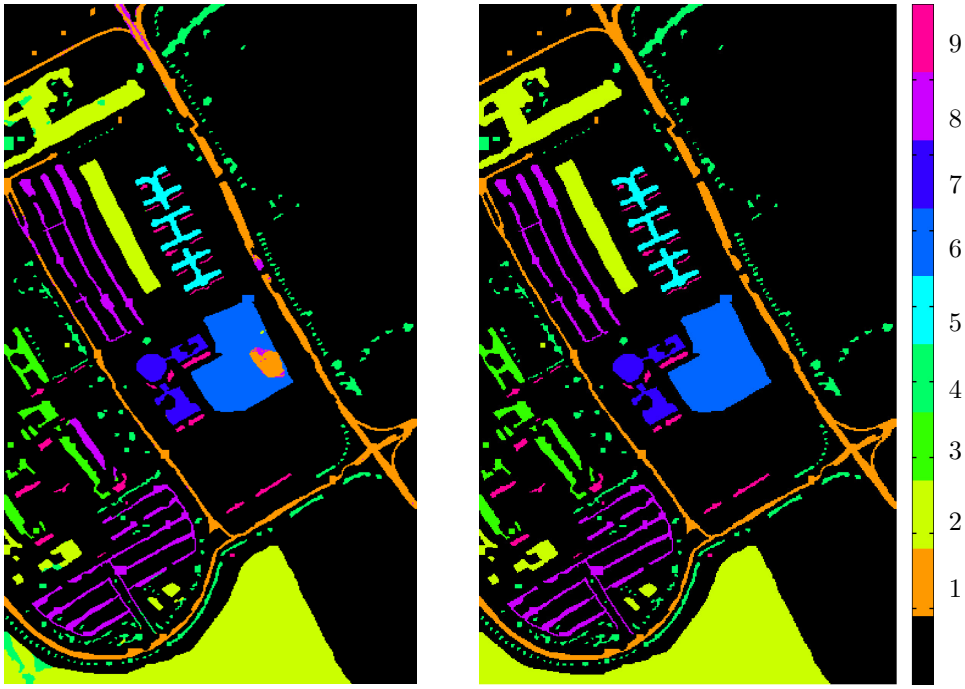
Class name	Test/Train	PCA	SpTv
1. Asphalt	6631/548	0.79	**0.95**
2. Meadows	18649/540	0.60	**0.97**
3. Gravel	2099/392	0.61	**0.84**
4. Trees	3064/524	0.98	**0.99**
5. Painted metal sheets	1345/265	**1.00**	**1.00**
6. Bare Soil	5029/532	**0.94**	0.90
7. Bitumen	1330/375	0.84	**1.00**
8. Self-Blocking Bricks	3682/514	0.88	**0.98**
9. Shadows	947/231	0.97	**1.00**
Average Class Acc.		0.844	**0.959**
Overall Acc.		0.747	**0.957**
κ		0.689	0.943
Number of features		10	13

Table 4.5: The average classification accuracies (and variance) for the Indian Pines data set using the SpTv method. The values are averaged from 20 Random Forest experiments. The bold numbers show the highest classification result.

Class name	Test/Train	PCA / var	Sp / var	Tv / var	SpTv / var
1. Corn-notill	1384/50	0.71 / 0.001	0.551 / 0.001	0.86 / 0.002	**0.90** / 0.001
2. Corn-min	784/50	0.64 / 0.002	0.677 / 0.002	0.92 / 0.001	**0.97** / 0.000
3. Corn	184/50	0.88 / 0.001	0.872 / 0.002	0.99 / 0.000	**1.00** / 0.000
4. Grass/Pasture	447/50	0.90 / 0.001	0.859 / 0.001	**0.96** / 0.001	**0.96** / 0.000
5. Greass/Trees	697/50	0.92 / 0.001	0.903 / 0.001	**0.99** / 0.000	**0.99** / 0.000
6. Hay-windrowed	439/50	0.99 / 0.000	0.971 / 0.000	**1.00** / 0.000	**1.00** / 0.000
7. Soybean-no till	918/50	0.81 / 0.001	0.690 / 0.002	**0.94** / 0.001	**0.93** / 0.001
8. Soybean-min till	2418/50	0.63 / 0.003	0.600 / 0.002	0.89 / 0.001	**0.94** / 0.001
9. Soybean-clean	564/50	0.87 / 0.001	0.734 / 0.001	0.93 / 0.001	**0.95** / 0.000
10. Wheat	162/50	0.99 / 0.000	0.988 / 0.000	**0.99** / 0.000	**1.00** / 0.000
11. Woods	1244/50	0.87 / 0.002	0.847 / 0.002	0.97 / 0.000	**1.00** / 0.000
12. Bldg-Grass-Trees-Drives	330/50	0.72 / 0.003	0.655 / 0.003	**0.99** / 0.000	**0.99** / 0.000
13. Stone-Steel-Towers	45/50	0.98 / 0.000	0.987 / 0.000	0.99 / 0.000	**1.00** / 0.000
14. Alfalfa	39/15	0.72 / 0.007	0.779 / 0.005	**0.98** / 0.002	0.97 / 0.001
15. Grass/pasture-mowed	11/15	0.82 / 0.007	0.864 / 0.008	**0.98** / 0.002	**0.98** / 0.001
16. Oats	5/15	0.98 / 0.004	0.970 / 0.005	**1.00** / 0.000	**1.00** / 0.000
Average Class Acc.		0.839 / 0.015	0.809 / 0.021	0.961 / 0.002	**0.974** / 0.001
Overall Acc.		0.771 / 0.000	0.714 / 0.000	0.929 / 0.000	**0.955** / 0.000
κ		0.741	0.679	0.919	0.949
Number of features		20	11	11	11

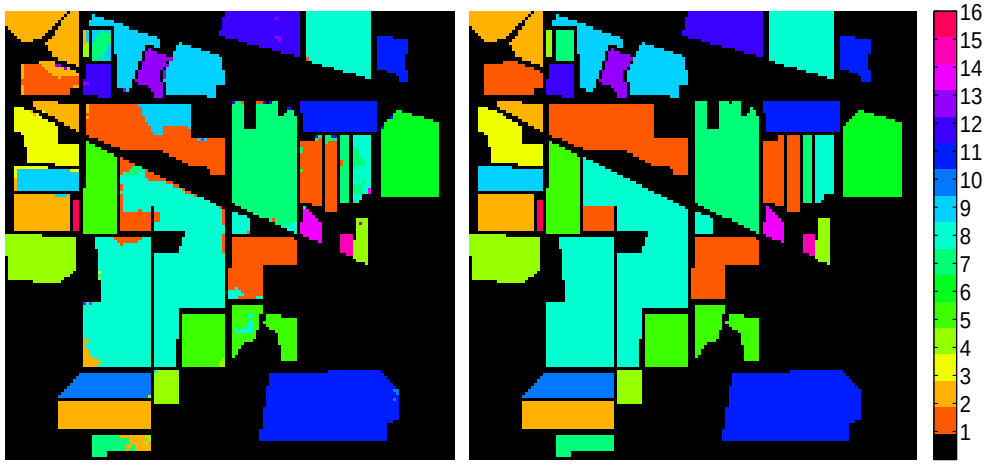


(a) Training samples. The gray area shows all labeled pixels.



(b) Classification results. (c) Ground truth.

Figure 4.8: Pavia University. The training samples, the ground truth and the classification results using the features of SpTv.



(b) Classification results. (c) Ground truth.

Figure 4.9: Indian Pines. The ground truth and the classification results using the features of SpTv.

4.8 Conclustions

A sparse and smooth unmixing method is proposed. The method exploits both the sparsity and the piecewise smoothness of abundance maps by minimizing a novel penalized cost function. The method combines ℓ_q $(0 \leq q \leq 1)$ and total variation regularization to achieve sparse and piecewise smooth abundance maps. Both the sparseness and the smoothness of the method can be varied via tuning parameters.

The dyadic expansion is used to decouple the minimization problem. This simplifies the computations and optimized TV image denoising algorithms can be used, with modifications to incorporate the ℓ_q sparsity regularization. The well known ASC is not used but rather the ENC which constrains the endmembers to have unit norm.

Using simulated data, it is shown that neither sparse nor smooth regularization, respectively, is able to minimize all three metrics that are used to evaluate the unmixing. By combining sparseness and smoothness, a better solution is obtained where all the metrics are close to their global minima. The simulations also show that using the ℓ_1 norm can yield a low reconstruction error, but better results w.r.t. the endmembers and abundances can be achieved by using an ℓ_q norm with $q < 1$.

The method is also tested on two real hyperspectral data sets. The former is the Urban data set and the unmixing solution is shown to improve when the proposed regularization is used, compared to no regularization. The second real data set is the Cuprite data set. The result obtained is compared to another well established unmixing method and the estimated endmembers are compared to a spectral library.

Both methods found similar endmembers, and library endmembers could be identified that closely resembled the estimated endmembers.

The method is also used for classification, where the abundance maps are used as features for a Random Forest classifier. All of the tuning parameters are selected using cross-validation, with the exception that $q = 1$ which means that an ℓ_1 sparsity regularizer is used. The Indian Pines and Pavia University data sets are classified in this manner.

A combination of sparse and TV regularization yields the best classification results for the Indian Pines. For the Pavia data set, cross-validation chose $h_q = 0$, so in that case only TV regularization was used.

The classification accuracies are significantly improved using the proposed method compared to classification accuracies obtained using the features provided by PCA.

Chapter 5

Semi-Supervised Hyperspectral Unmixing

In this chapter we propose unmixing methods that rely on spectral dictionaries. The first method uses hard constraints to force a subset of the endmembers to be identical to known endmembers while the second method does the same using soft constraints. This information about known endmembers can be acquired from a spectral library or extracted from the data. The third and fourth methods make the pure pixel assumption, and the endmembers will be selected from a dictionary constructed from the data matrix. It may be appealing to have the endmembers coincide with certain pixels in the image. By doing this, the dictionary will be physically meaningful and may be interpreted unambiguously from the data set. All of the methods are evaluated using both simulated and real hyperspectral data.

5.1 Semi-Supervised unmixing

In this section, two semi-supervised hyperspectral unmixing method are proposed. Supervised and unsupervised unmixing is combined by using known endmembers while estimating the unknown endmembers. The abundances are jointly estimated for both the known and unknown endmembers. This process is repeated until the methods have converged to their final solution. In the first method, the ASC is enforced using the matrix augmentation given in (2.9).

The abundance sum constraint (ASC) which states that the sum of all the abundances in each pixel should amount to one is often enforced in hyperspectral unmixing. This constraint is enforced in the first method but it will not be enforced in the second method. It has received some criticism in the literature [2,150]. An endmember norm constraint (ENC), e.g. that the endmembers spectra should sum to one will however be enforced in the second method. Adding this constraint will ensure that endmembers are identified solely by their spectral signatures and not by their intensities. This constraint also regularizes the method and avoids the degenerate solution.

If some material is known to be in the image and its spectral signature is known (e.g. in a spectral library), then the unmixing will be improved by incorporating this information into the unmixing process. In some cases it is possible to identity pure pixels in an image. The spectral signatures of these pure pixels may be viewed as noisy endmembers and they can be used to aid the unmixing process.

5.1.1 Prior work

In [91] a sparse semi-supervised unmixing approach was presented. In that case a spectral library was used, and the unmixing amounts to finding a small number of materials in the library that best represent the observed data.

Semi-supervised unmixing similar to the methods proposed here has been used in other fields. In [164], semi-blind independent component analysis (ICA [64]) of functional magnetic resonance imaging (fMRI) data is proposed where the columns of the mixing matrix are constrained to be close to pre-specified time courses. Supervised singular value decomposition (SVD) for ICA of fMRI data is proposed in [165]. The approach taken is to supervise the SVD by restricting it to find the best low rank approximation within a certain subspace.

To distinguish between the two methods we will refer to the first hard constrained method as hard-constrained-unmixing (HCU) and the second method as soft-constrained-unmixing (SCU).

5.1.2 HCU Model

The low rank model is used to represent the hyperspectral data,

$$\underset{M\times P}{\boldsymbol{Y}} = \underset{M\times r_k}{\boldsymbol{X}} \quad \underset{r_k\times P}{\boldsymbol{B}} + \underset{M\times r_u}{\boldsymbol{A}} \quad \underset{r_u\times P}{\boldsymbol{S}} + \underset{M\times P}{\boldsymbol{N}}, \tag{5.1}$$

Each pixel in the image is represented by one column in $\boldsymbol{Y}$. Each column in $\boldsymbol{X}$ represents one known endmember, each column in $\boldsymbol{A}$ represents one unknown endmember. The ith rows in $\boldsymbol{S}$ and $\boldsymbol{B}$ represent the abundance maps associated with the ith columns in $\boldsymbol{X}$ and $\boldsymbol{A}$, respectively. $\boldsymbol{N}$ is additive noise. The number of spectral bands in the data is denoted M, r_k is the number of known endmembers, r_u is the number of unknown endmembers and P is the number of pixels.

In (5.1), $\boldsymbol{X}$ and $\boldsymbol{Y}$ are the only known variables and the task is to estimate $\boldsymbol{B}$, $\boldsymbol{A}$ and $\boldsymbol{S}$. To estimate these unknown variables, the following cost function is minimized,

$$\begin{aligned} J(\boldsymbol{A},\boldsymbol{S}) &= \|\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{B}-\boldsymbol{A}\boldsymbol{S}\|^2 + \delta^2\|\mathbf{1}_P^\mathsf{T} - \mathbf{1}_r^\mathsf{T}\boldsymbol{Z}\|^2 \\ &= \|\boldsymbol{Y}-\boldsymbol{W}\boldsymbol{Z}\|^2 + \delta^2\|\mathbf{1}_P^\mathsf{T} - \mathbf{1}_r^\mathsf{T}\boldsymbol{Z}\|^2 \\ &= \|\boldsymbol{Y}_f-\boldsymbol{W}_f\boldsymbol{Z}\|^2 \\ &\text{s.t. } \boldsymbol{A},\ \boldsymbol{B},\ \boldsymbol{S}\geq 0, \end{aligned} \tag{5.2}$$

where $\boldsymbol{W}=[\boldsymbol{X}\boldsymbol{A}]$, $\boldsymbol{Z}=[\boldsymbol{B}^\mathsf{T}\boldsymbol{S}^\mathsf{T}]^\mathsf{T}$ and $\mathbf{1}_r$ and $\mathbf{1}_P$ are column vectors of ones. The parameter, δ, controls how strongly the ASC should be enforced. The matrices $\boldsymbol{Y}_f$ and $\boldsymbol{W}_f$ in (5.2) are augmented using (2.9).

Estimation

A (block) cyclic descent algorithm [135] is used to minimize (5.2). The algorithm consists of an abundance estimation step and an endmember estimation step. The abundances are estimated while holding the unknown endmembers fixed and then the unknown endmembers are estimated while holding the abundances fixed. This process is repeated until the cost function has converged.

To estimate the abundances, a majorization-minimization (MM) [147] approach is used. The MM approach iteratively minimizes a majorizing function. The objective function is driven downward in each step until a local minimum is found. The majorizing function [148] used is

$$Q = \text{tr}(\boldsymbol{Z} - \boldsymbol{Z}^{(t)})^{\mathsf{T}}(\alpha \boldsymbol{I}_r - \boldsymbol{W}_f^{\mathsf{T}}\boldsymbol{W}_f)(\boldsymbol{Z} - \boldsymbol{Z}^{(t)}), \tag{5.3}$$

where tr() is the trace operator, $\boldsymbol{Z}^{(t)}$ is an estimation of $\boldsymbol{Z}$ in iteration t and α is the maximum eigenvalue of $\boldsymbol{W}_f^{\mathsf{T}}\boldsymbol{W}_f$. The cost function that will be minimize is given by

$$F(\boldsymbol{Z}, \boldsymbol{Z}^{(t)}) = J + Q = \|\boldsymbol{E} - \boldsymbol{Z}\|^2 \tag{5.4}$$

where $\boldsymbol{E} = \boldsymbol{W}_f^{\mathsf{T}}\boldsymbol{Y}_f + \alpha\boldsymbol{Z}^{(t)} - \boldsymbol{W}_f^{\mathsf{T}}\boldsymbol{W}_f\boldsymbol{Z}^{(t)}$. By estimating $\boldsymbol{Z}$, both $\boldsymbol{B}$ and $\boldsymbol{S}$ have been jointly estimated.

To estimate the unknown endmembers, $\boldsymbol{A}$, we minimize (5.2) with respect to $\boldsymbol{A}$. We write the terms of (5.2) that depend on $\boldsymbol{A}$,

$$J_{\boldsymbol{A}} = \|(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})_+ - \boldsymbol{A}\boldsymbol{S}\|^2 \tag{5.5}$$

where $(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{B})_+ = \max(0, \boldsymbol{Y}-\boldsymbol{X}\boldsymbol{B})$. A nonnegative quadratic programming method similar to [137] (2.20) is used to minimize (5.5).

5.1.3 SCU model

The hyperspectral model used here is the low rank model (2.5),

$$\underset{M\times P}{\boldsymbol{Y}} = \underset{M\times r}{\boldsymbol{A}} \quad \underset{r\times P}{\boldsymbol{S}} + \underset{M\times P}{\boldsymbol{N}}, \tag{5.6}$$

where all the variables are the same as defined in Chapter 2.1. The cost function that is optimized to estimate the unknown variables is given by

$$\begin{aligned} &J(\boldsymbol{A}, \boldsymbol{S}) = \|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{S}\|^2 + h\|\boldsymbol{D} - \boldsymbol{A}_s\|_F^2 + g\|\mathbf{1}_r^{\mathsf{T}} - \mathbf{1}_M^{\mathsf{T}}\boldsymbol{A}\|^2 \\ &\text{s.t. } \boldsymbol{A},\ \boldsymbol{S} \geq 0 \end{aligned} \tag{5.7}$$

where $\boldsymbol{D}$ is an $M \times r_k$ matrix consisting of the library endmembers, r_k is the number of library endmembers, $\boldsymbol{A}_s$ are the estimated supervised endmembers, and $\mathbf{1}_r$ and $\mathbf{1}_M$ are column vectors of ones.

The first term in (5.7) is the fidelity term. The second term is controlled by the tuning parameter, h, and dictates how strongly the supervised endmembers $\boldsymbol{A}_s$, should resemble the library endmembers in $\boldsymbol{D}$. If h is very large, the endmembers in $\boldsymbol{A}_s$ will be forced to be identical to the library endmembers. The third term is controlled by g and promotes the columns of $\boldsymbol{A}$ (the endmembers) to sum to one. If g is set very large, all endmembers will sum to exactly one.

Estimation

The MM approach that was used to estimate the abundances for the HCU model is also used here. The majorizing function that is used is

$$Q_{\boldsymbol{S}} = \operatorname{tr}(\boldsymbol{S} - \boldsymbol{S}^{(k)})^{\mathsf{T}}(\alpha \boldsymbol{I}_r - \boldsymbol{A}^{\mathsf{T}}\boldsymbol{A})(\boldsymbol{S} - \boldsymbol{S}^{(k)}), \tag{5.8}$$

where α is now the maximum eigenvalue of $\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}$. The cost function that will be minimized is given by

$$F_{\boldsymbol{S}}(\boldsymbol{S}, \boldsymbol{S}^{(k)}) = J + Q_{\boldsymbol{S}} = \|\boldsymbol{B} - \boldsymbol{S}\|^2, \tag{5.9}$$

where $\boldsymbol{B} = \boldsymbol{A}^{\mathsf{T}}\boldsymbol{Y} + \alpha \boldsymbol{S}^{(k)} - \boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{S}^{(k)}$.

As with the abundances, an MM approach is used to estimate the endmembers. The MM function is

$$Q_{\boldsymbol{A}} = \operatorname{tr}(\boldsymbol{A} - \boldsymbol{A}^{(k)})^{\mathsf{T}}(\alpha \boldsymbol{I}_r - \boldsymbol{S}\boldsymbol{S}^{\mathsf{T}})(\boldsymbol{A} - \boldsymbol{A}^{(k)}), \tag{5.10}$$

where $\boldsymbol{A}^{(k)}$ is an estimation of $\boldsymbol{A}$ at iteration t and α is the maximum eigenvalue of $\boldsymbol{S}\boldsymbol{S}^{\mathsf{T}}$. The cost function that will be minimized is

$$\begin{aligned} F_{\boldsymbol{A}}(\boldsymbol{A}, \boldsymbol{A}^{(k)}) =& J + Q_{\boldsymbol{A}} \\ =& \alpha\|\boldsymbol{B} - \boldsymbol{A}\|^2 + h\|\boldsymbol{D} - \boldsymbol{A}_s\|^2 + g\|\boldsymbol{1}_r^{\mathsf{T}} - \boldsymbol{1}_M^{\mathsf{T}}\boldsymbol{A}\|^2, \end{aligned} \tag{5.11}$$

where $\boldsymbol{B} = 1/\alpha(\boldsymbol{Y}\boldsymbol{S}^{\mathsf{T}} + \alpha \boldsymbol{A}^{(k)} - \boldsymbol{S}\boldsymbol{S}^{\mathsf{T}})$. To minimize (5.11), a nonnegative quadratic programming approach similar to (2.20) is used.

5.1.4 Simulations

Spectral signatures from the USGS spectral library[1] are used to generate the simulated data. The number of pixels is set $P = 300$ and the abundance maps are generated following a Dirichlet probability density function. One pure pixel for each endmember is included in the data. Spectral variability is introduced into the data so the sum of the columns of $\boldsymbol{S}$ will range from 0.7 to 1.3.

[1] http://speclab.cr.usgs.gov/spectral.lib06

HCU Simulations

In these simulations, eight endmembers are used and Gaussian white noise is added to all bands uniformly so the signal to noise ratio is 40dB.

The method is evaluated using the spectral angle distance (SAD) given in (A.2), and the normalized mean square error (nMSE) defined as

$$\text{nMSE} = \frac{\|\mu - \widehat{\mu}\|_F^2}{\|\mu\|_F^2}, \tag{5.12}$$

where F denotes the Frobenius norm, $\mu = [\boldsymbol{X}\boldsymbol{A}][\boldsymbol{B}^\mathsf{T}\boldsymbol{S}^\mathsf{T}]^\mathsf{T}$ and $\widehat{\mu}$ is an estimate of μ.

In this simulation, the SAD ans nMSE are calculated while the number or endmembers (r_u) are varied. When $r_u = 8$, there are eight unknown endmembers in the image and when $r_u = 1$ there is one unknown endmember and seven known endmembers. The columns of $\boldsymbol{X}$ are populated with the spectral signatures from the pure pixels in the data. The matrix $\boldsymbol{A}$ is initialized with selected pixels from the data set (not the pure pixels). The ASC is very mildly enforced by setting $\delta = 0.3$. In Figure 5.1, the mean SAD from 100 simulations (along with one standard deviation) is plotted when $1 \leq r_u \leq 8$. A new abundance map is generated in each simulation. The mean SAD is defined as

$$\overline{\text{SAD}} = \frac{1}{L}\sum_{l=1}^{L} \text{SAD}(\widehat{\boldsymbol{W}}_l, \boldsymbol{W}), \tag{5.13}$$

where $\widehat{\boldsymbol{W}}_l$ are the estimated endmembers (known and unknown) for simulation l, and $L = 100$. The mean nMSE and mean running time are similarly defined. The mean nMSE is shown in Figure 5.2. The mean running time of the algorithm for $1 \leq r_u \leq 8$ is shown in Figure 5.3. Both the SAD and the time needed for the method to converge decrease when r_u is decreased.

The nMSE does however increase when r_u is decreased, note that the endmembers used to populate $\boldsymbol{X}$ are endmembers corrupted with noise.

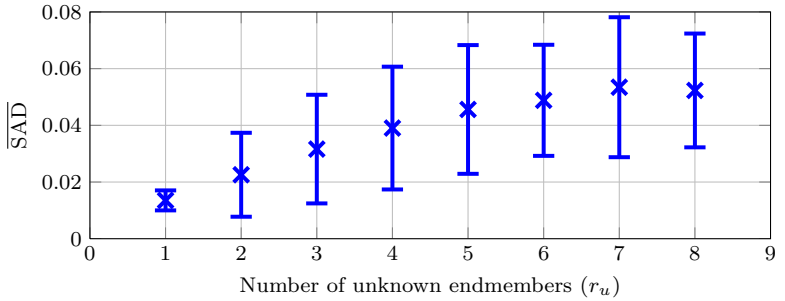


Figure 5.1: Semi-supervised HCU. The mean spectral angle distance, along with one standard deviation, for $1 \leq r_u \leq 8$.

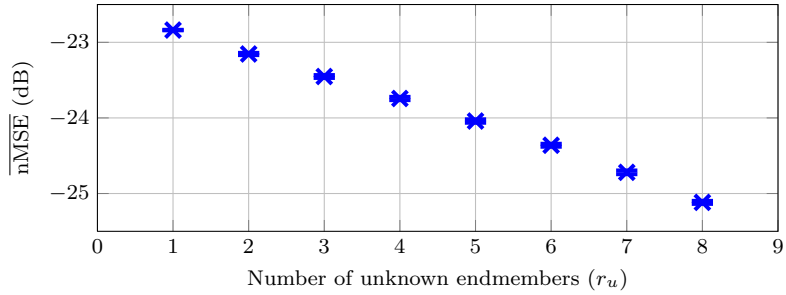


Figure 5.2: Semi-supervised HCU. The mean nMSE, along with one standard deviation, for $1 \leq r_u \leq 8$.

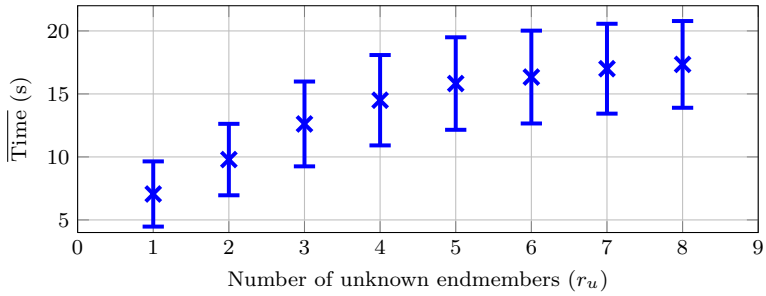


Figure 5.3: Semi-supervised HCU. The mean computational time, along with one standard deviation,for $1 \leq r_u \leq 8$.

SCU Simulations

In these simulations, six endmembers are used and Gaussian white noise is added to all bands uniformly so the signal to noise ratio is 35dB.

To evaluate the method, the spectral angle distance (SAD), and normalized mean square error ($\text{nMSE}_{\boldsymbol{AS}}$), defined in Appendix A, are used.

The simulations are similar to the HCU case. The SAD and $\text{nMSE}_{\boldsymbol{AS}}$ are varied while changing the number of unknown endmembers (r_u) and the tuning parameter, h. The matrix $\boldsymbol{D}$ is initialized in the same way as in the HCU case. The tuning parameter g is set to 1 which means the sum of each estimated endmember will be very close to 1. Figure 5.4 gives the mean SAD for $1 \leq r_u \leq 6$ and $10^{-2} \leq h \leq 10^{7}$ from 100 simulations. In each simulation, new abundance maps and noise components are generated.

In Figure 5.4, it can be seen that when r_u is decreased, the mean SAD can be lowered and the minima is not always found when h is chosen very large. This means that a

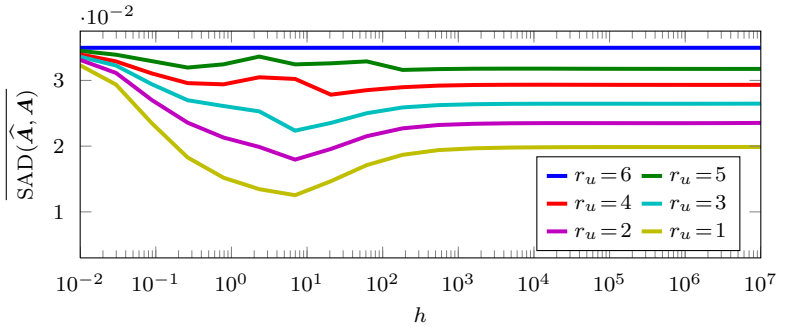


Figure 5.4: Semi-supervised SCU. The mean SAD when $1 \leq r_u \leq 6$ and $10^{-2} \leq h \leq 10^{7}$.

soft constraint can result in better unmixing, if h is chosen correctly. The mean SAD is at a minima for $r_u \leq 3$ when $h = 6.95$. In Figure 5.5, the mean SAD along with one standard deviation is plotted for $1 \leq r_u \leq 6$ and $h = 6.95$.

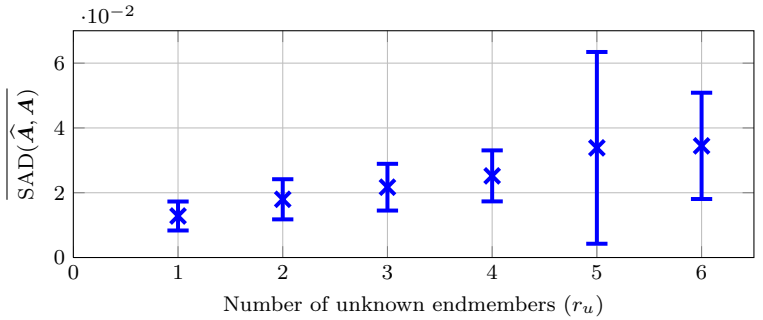


Figure 5.5: Semi-supervised SCU. The mean SAD, along with one standard deviation, when $1 \leq r_u \leq 6$ and $h = 6.95$.

In Figure 5.6, the mean $\mathrm{nMSE}_{\boldsymbol{AS}}$ along with one standard deviation is plotted for $1 \leq r_u \leq 6$ and $h = 6.95$. The $\mathrm{nMSE}_{\boldsymbol{AS}}$ does increase when r_u decreases, but bear in mind that the library, $\boldsymbol{D}$ was not populated with true endmembers but with noisy pure pixels from the data.

5.1.5 Real Data

The real data set is the Urban hyperspectral image, detailed in Appendix B.1. A generated RGB image of the data set is shown in Figure 5.7.

At least eight different macroscopic materials can be seen in the RGB image: asphalt, dirt, three different colored rooftops, and two types of vegetation and shadow areas. The number of endmembers is set to eight and the algorithm is initialized with pixels at located at the regions marked with x's and labeled 1-8 in the RGB image.

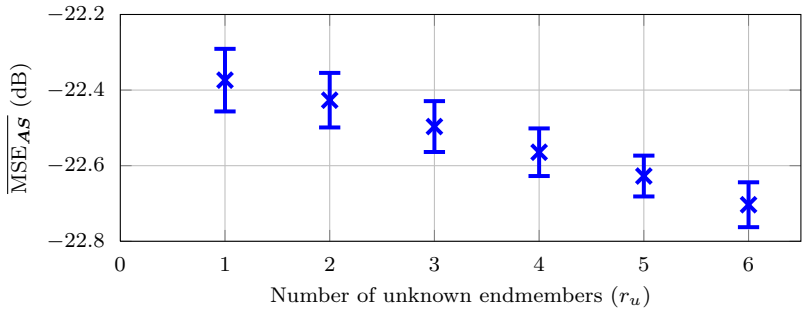


Figure 5.6: Semi-supervised SCU. The mean $\text{nMSE}_{\boldsymbol{AS}}$, along with one standard deviation, when $1 \leq r_u \leq 6$ and $h = 6.95$.

In the RGB image, there are also 3 regions of interest marked with yellow borders. The first region in the top right corner of the RGB image shows an asphalt road, grass and a parking lot. The second region in the lower left corner and shows a residential area (rooftops, trees, and grass). The third area is in the lower right corner and shows grass, trees and a road.

The same procedure as was used for the simulated data is applied here. The two unmixing methods are used for $1 \leq r_u \leq 8$. When $r_u = 8$ there are no known endmembers. When $r_u = 7$ the first endmember assumed to be known. This is the endmember associated with asphalt roads. A spectral signature is generated by averaging a small area around the 'x' marked 1 in the image. This signature is assumed to be an endmember and to represent asphalt. The signature is used to populate the $\boldsymbol{X}$ matrix (or $\boldsymbol{D}$ for SCU) . When $r_k = 6$ we populate $\boldsymbol{X}$ with two signatures, associated with asphalt and rooftops (marked 1 and 2 in the RGB image). This procedure is then repeated for $3 \leq r_k \leq 8$.

Three abundance maps for $r_u = \{8, 7, 5, 3, 1\}$ are given in Figures 5.8 and 5.9. The top rows show the maps associated with asphalt (region 1), the middle rows show the maps associated with rooftops (region 2) and the bottom rows show the maps associated with grass (region 3).

It can be seen that the abundance maps improve when r_k is increased. When $r_k > 0$, the quality of the abundance maps improves and they better represent the material they are associated with.

Figure 5.7: Semi-supervised unmixing. The generated RGB image of the Urban data set. The red x's labeled 1-8 show the locations of the spectra used to initialize the algorithm. The yellow rectangles show the locations of the regions of interest for the first three endmembers.

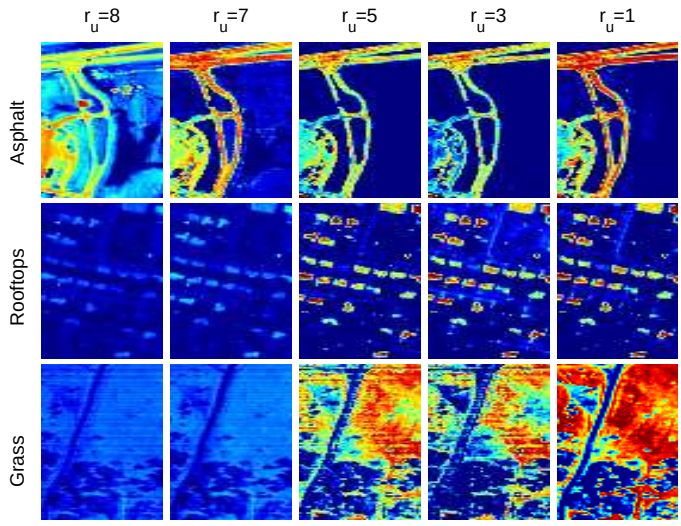


Figure 5.8: Semi-supervised HCU. The abundance maps for the regions of interest when $r_u = \{8, 7, 5, 3, 1\}$.

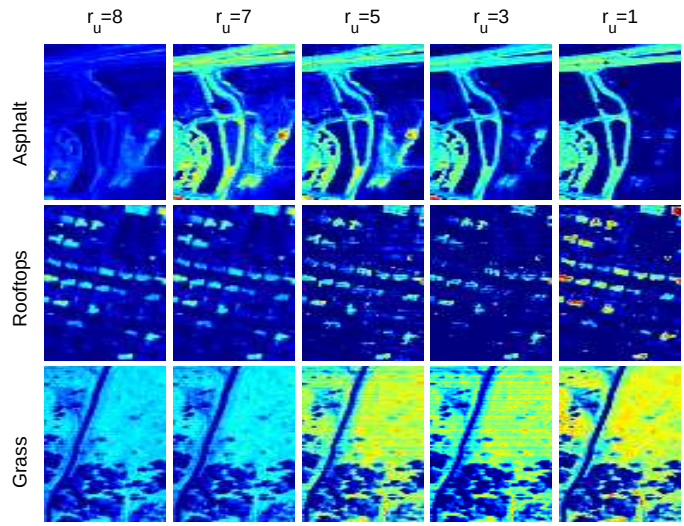


Figure 5.9: Semi-supervised SCU. The abundance maps of the regions of interest when $r_u = \{8, 7, 5, 3, 1\}$.

5.2 Matrix Decomposition Using the Pure Pixel Assumption

Finding a low rank approximation of a matrix that lies in the span of few columns of the matrix is a hard combinatorial problem. In the context of hyperspectral imaging this amounts to finding endmembers that are already present in the data. The assumption is made that there are pure pixels within the data. This assumption is a strong requisite and may not hold for all data sets [2].

This methodology has also been called CUR [32] decomposition. The original CUR [32] takes a randomized algorithmic approach. However, [31] proposed a convex optimization version of CUR which methods here are loosely based on. The method proposed in [77] can also be considered a relaxation of CUR. Related work that also estimate endmembers using sparse regression is [95], where the sparsity regularization is applied in a different manner than is done here.

The methods proposed here are useful for finding low rank approximations for hyperspectral data based on endmembers that appear in the data cube. The methods can also be used for initializing other methods that further refine the solution. The methods automatically select the rank of the data set, which usually is a very hard problem. However, a sparseness tuning parameter which affects the rank of the data set must be chosen. One novelty of the proposed methods is the use of the vector ℓ_0 penalty combined with the nonnegativity of the abundance maps.

5.2.1 The Cost Function

Using the definitions given in Section 2.1, the cost function considered is given by

$$J(\boldsymbol{S}) = \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{Y}_m\boldsymbol{S}\|^2 + \frac{\lambda}{2}\sum_{i=1}^{P} R_q(\boldsymbol{s}_i), \quad \text{s.t. } \boldsymbol{S} \geq 0, \tag{5.14}$$

where $\boldsymbol{Y}_m$ be a subset of m columns from $\boldsymbol{Y}$. This subset will make up a dictionary, or candidate endmembers. $R_q(\boldsymbol{s}_i)$ is the regularization term and the tuning parameter, λ, is nonnegative and controls the sparseness of the solution, that is, the number of nonzero rows in $\boldsymbol{S}$.

Two sparsity inducing regularization terms are considered. The first is the standard group lasso regularization, i.e. the ℓ_2 penalty ($q = 2$) defined as

$$R_2(\boldsymbol{s}_i) = \|\boldsymbol{s}_i\|_2 = \Big(\sum_{p=1}^{P} s_{ip}^2\Big)^{1/2}. \tag{5.15}$$

The second regularization term is the vector ℓ_0 ($q = 0$) penalty defined as

$$R_0(\boldsymbol{s}_i) = \|\boldsymbol{s}_i\|_0 = I(\|\boldsymbol{s}_i\| \neq 0) \tag{5.16}$$

where $I(\cdot)$ is the indicator function.

Estimation

To estimate the abundance matrix, $\boldsymbol{S}$, a majorization-minimization [147] (MM) approach is used. The majorizing function [148] that will be minimized is (ignoring constants)

$$\begin{aligned} Q(\boldsymbol{S}, \boldsymbol{S}^{(k)}) =& J(\boldsymbol{S}) + \frac{1}{2}\mathrm{tr}(\boldsymbol{S} - \boldsymbol{S}^{(k)})^\mathsf{T}(\alpha\boldsymbol{I}_r - \boldsymbol{Y}_m^\mathsf{T}\boldsymbol{Y}_m)(\boldsymbol{S} - \boldsymbol{S}^{(k)}), \\ =& \frac{1}{2}\mathrm{tr}(\alpha\boldsymbol{S}^\mathsf{T}\boldsymbol{S} - 2\boldsymbol{B}^\mathsf{T}\boldsymbol{S}) + \frac{\lambda}{2}\sum_{i=1}^{P} R_q(\boldsymbol{s}_i) \\ & \text{s.t. } \boldsymbol{S} \geq 0, \end{aligned} \tag{5.17}$$

where $\boldsymbol{S}^{(k)}$ is an estimate of $\boldsymbol{S}$,

$$\boldsymbol{B} = \boldsymbol{Y}_m^\mathsf{T}\boldsymbol{Y} + (\alpha\boldsymbol{I}_r - \boldsymbol{Y}_m^\mathsf{T}\boldsymbol{Y}_m)\boldsymbol{S}^{(k)},$$

and α is the maximum eigenvalue of $\boldsymbol{Y}_m^\mathsf{T}\boldsymbol{Y}_m$. This choice of α ensures that $(\alpha\boldsymbol{I}_r - \boldsymbol{Y}_m^\mathsf{T}\boldsymbol{Y}_m)$ is nonnegative definite.

The solution to the minimization problem with ℓ_0 regularization is

$$\boldsymbol{s}_i{}^{(k+1)} = \begin{cases} \frac{1}{\alpha}\boldsymbol{b}_{i+} & \text{if } \frac{1}{\alpha}\boldsymbol{b}_{i+}\boldsymbol{b}_{i+}^\mathsf{T} \geq \lambda \\ 0 & \text{otherwise,} \end{cases} \tag{5.18}$$

where $\boldsymbol{b}_i$ is row i in $\boldsymbol{B}$ and $(\cdot)_+ = \max(0, \cdot)$. The solution to the minimization problem with ℓ_2 regularization is

$$\boldsymbol{s}_i^{(k+1)} = \frac{1}{\alpha}(\boldsymbol{b}_{i+}\boldsymbol{b}_{i+}^\mathsf{T} - \lambda)_+ \left(\frac{\boldsymbol{b}_{i+}}{\|\boldsymbol{b}_{i+}\|_2}\right). \tag{5.19}$$

Two iterative algorithms are implemented that solve these minimization problems, respectively. Algorithm 7 uses (5.18) to solve the minimization problem with ℓ_0 regularization and Algorithm 8 uses (5.19) for the ℓ_2 case. Throughout the rest of this chapter, Algorithm 7 will referred to as the ℓ_0 *method* and Algorithm 8 will referred to as the ℓ_2 *method*.

Algorithm 7: An iterative algorithm for estimating $\boldsymbol{S}$ using ℓ_0 regularization.

Initialize: $\boldsymbol{Y}_m$, $\widehat{\boldsymbol{S}}^{(0)}$ and set $k = 0$
while $\boldsymbol{S}^{(k+1)}$ *has not converged* **do**
  $\boldsymbol{S}^{(k+1)} = \boldsymbol{S}^{(k)}$
  while $\boldsymbol{S}^{(k+1)}$ *has not converged* **do**
    $\boldsymbol{B}_+ = \max(0, \boldsymbol{Y}_m^\mathsf{T}\boldsymbol{Y} + (\alpha\boldsymbol{I}_r - \boldsymbol{Y}_m^\mathsf{T}\boldsymbol{Y}_m)\boldsymbol{S}^{(k+1)})$
    $\boldsymbol{S}^{(k+1)} = \frac{1}{\alpha}\boldsymbol{B}_+$
  Calculate the thresholds, $th = \frac{1}{\alpha}\boldsymbol{B}^\mathsf{T}\boldsymbol{B}$
  If $\min(th_i) < \lambda$ then $\boldsymbol{s}_i{}^{(k+1)} = 0$.
  $k = k + 1$

Algorithm 8: An iterative algorithm for estimating $\boldsymbol{S}$ using ℓ_2 regularization.

Initialize: $\boldsymbol{Y}_m$, $\widehat{\boldsymbol{S}}^{(0)}$ and set $k = 0$
while $\boldsymbol{S}^{(k+1)}$ *has not converged* **do**
  $\boldsymbol{B}_+ = \max(0, \boldsymbol{Y}_m^\mathsf{T}\boldsymbol{Y} + (\alpha\boldsymbol{I}_r - \boldsymbol{Y}_m^\mathsf{T}\boldsymbol{Y}_m)\boldsymbol{S}^{(k+1)})$
  $\boldsymbol{s}_i^{(k+1)} = \frac{1}{\alpha}(\boldsymbol{b}_{i+}\boldsymbol{b}_{i+}^\mathsf{T} - \lambda)_+ \left(\frac{\boldsymbol{b}_{i+}}{\|\boldsymbol{b}_{i+}\|_2}\right)$
  $k = k + 1$

5.2.2 Experimental Results

The methods are evaluated using both simulated and real data. The simulated data set is created using 6 spectral signatures as endmembers and the abundance maps are generated following a Dirichlet distribution. The real data set is Urban data set described in Appendix B.1.

Simulated data

The simulated data is generated according to the model,

$$\underset{M\times P}{\boldsymbol{Y}} = \underset{M\times r}{\boldsymbol{A}} \quad \underset{r\times P}{\boldsymbol{S}} + \underset{M\times P}{\boldsymbol{N}} . \tag{5.20}$$

The number of endmembers is $r = 6$, the spectral bands are $M = 160$ and the number of pixels is $P = 1000$ and N is Gaussian noise.

Spectral signatures from the USGS digital spectral library[2] make up the columns of the $\boldsymbol{A}$ matrix. The spectral signatures are shown in Figure 5.10. Each column in the $\boldsymbol{S}$ matrix is generated following a Dirichlet probability density function, with parameter equal to one.

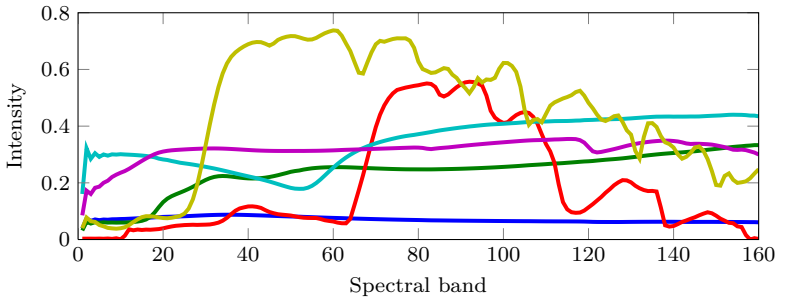


Figure 5.10: The six spectral signatures that are used in the PP simulations.

To evaluate the method, the SAD, given in (A.2) in Appendix A is used.

The SNR in the data is varied from 20dB to 40dB. $\boldsymbol{Y}_m$ is initialized by finding a subset of 50 distinct spectral signatures in $\boldsymbol{Y}$. Each signature in $\boldsymbol{Y}_m$ is constrained to have a SAD of at least 0.075 from all other signals in $\boldsymbol{Y}_m$. In Figure 5.11 the SAD can be seen, calculated for SNR values ranging from 20dB to 40dB. For each specified SNR, the SAD is calculated for 30 runs and the average value is shown. In each run, a new noise matrix is added to the original noiseless data. The sparseness parameter, λ, is chosen so the number of estimated endmembers is the same as the number of true endmembers used to create the data set.

The ℓ_0 method gives a lower SAD for all SNR values and is thus able to find endmembers that are closer to the true endmembers that were used to generate the data set.

[2] http://speclab.cr.usgs.gov/spectral.lib06

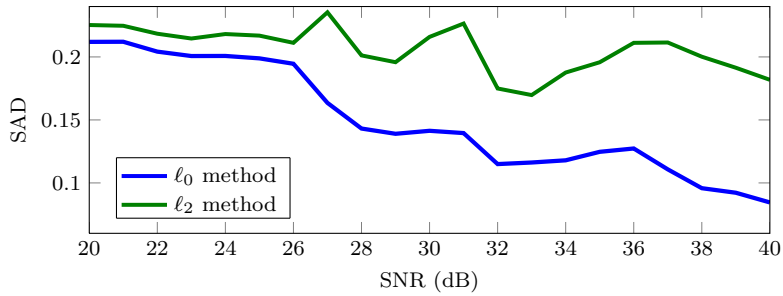


Figure 5.11: Using the PP assumption. The SAD between the estimated endmembers and the true endmembers for the simulation data.

Real data

The ℓ_0 method gave lower SAD values than the ℓ_2 method, so the focus will be on the ℓ_0 method. This method will be evaluated using the Urban data set. Bands (89, 90, 103-109, 130-152, 204-210) are removed and the remaining 171 bands are used. The same method that was used in the simulated data experiments is used here to initialize $\boldsymbol{Y}_m$, and it is composed of 38 signatures that are present in the data set.

In Figure 5.12, the error, $ERR = \|\boldsymbol{Y} - \boldsymbol{Y}_m\boldsymbol{S}\|^2$, is plotted w.r.t. the number of non-zero rows (r) in the $\boldsymbol{S}$ (the number of endmembers in the solution). The error decreases considerably when the number of endmembers increase from 1 to 5 and then decreases slightly when $5 < r < 8$. Increasing the number of endmembers beyond 8 does not yield a lower error.

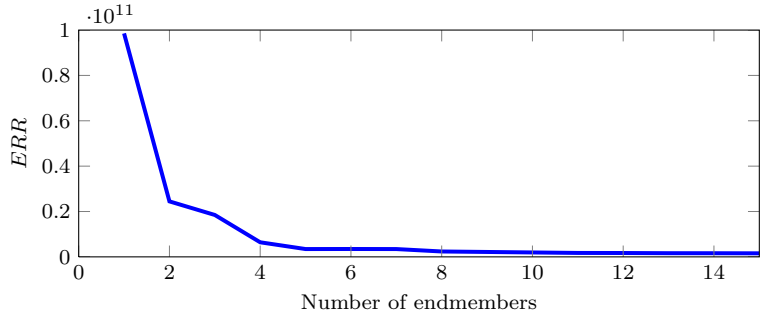


Figure 5.12: Using the PP assumption. $ERR = \|\boldsymbol{Y} - \boldsymbol{Y}_m\boldsymbol{S}\|$ plotted w.r.t. the number of endmembers.

The sparseness parameter, λ, is chosen so the algorithm converges having 7 endmembers. An RGB image of the scene, the endmembers and their abundance maps are shown in Figure 5.13. The abundance maps are clearly associated with macroscopic material visible in the RGB image. Some of the endmembers are however not pure in the sense that they are not associated with only one material. The abundance matrix,

$\boldsymbol{S}$ estimated by the algorithm is very sparse, approximately 60% of all entries in the matrix are zero.

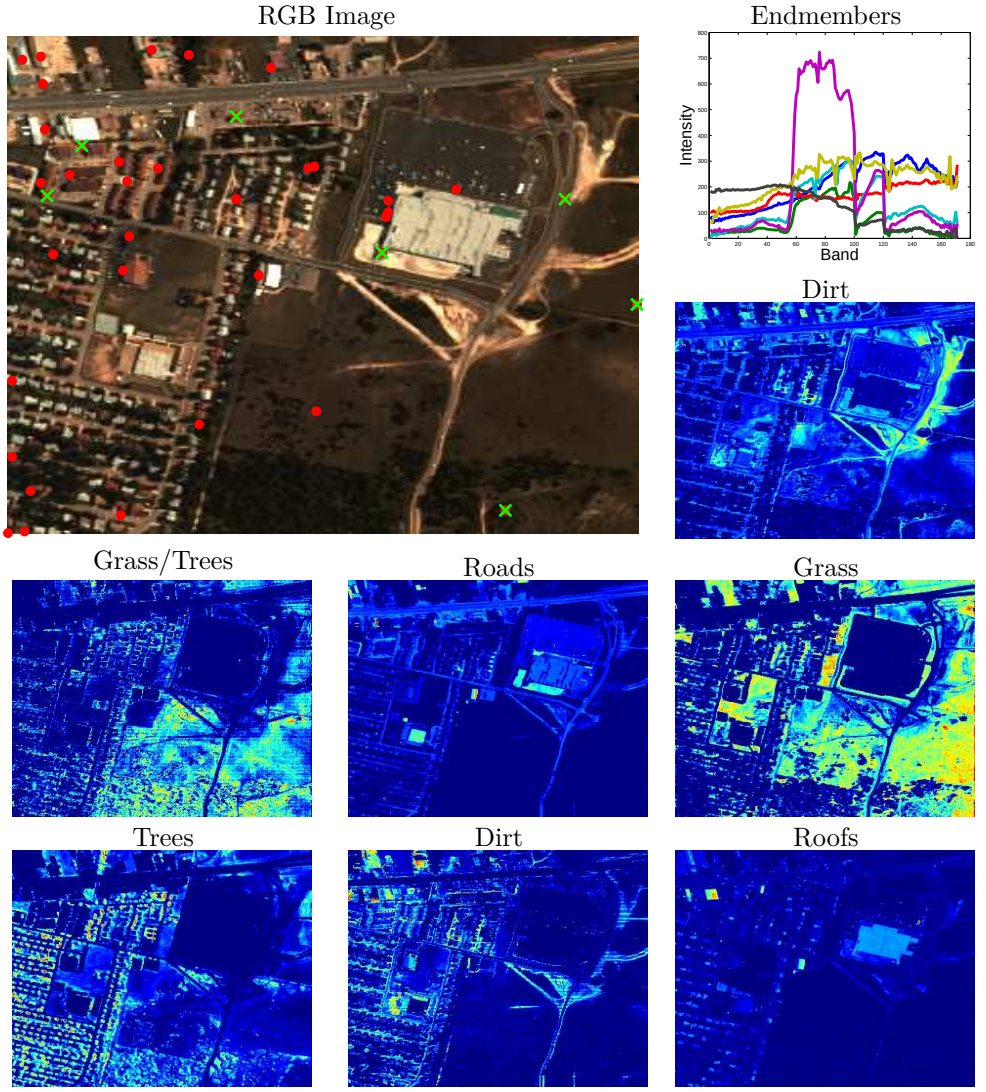


Figure 5.13: Using the PP assumption. The RGB image, the endmembers and the abundance maps estimated from the Urban data set. The red dots on the RGB image show the locations of the candidate endmembers that make up $\boldsymbol{Y}_m$. The green crosses show the location of the 7 endmembers found by the algorithm.

5.3 Conclusions

Four hyperspectral unmixing methods that rely in part on endmembers dictionaries have been proposed.

Two first methods combine supervised and unsupervised unmixing. *A priori* information about known endmembers is added to the unmixing model and we estimate endmembers that are not known. Using *A priori* information improves the unmixing results and also reduces the computations. The first method employs the ASC and uses hard constraints to force a subset of the endmembers to exactly match library endmembers. The second method used the ENC and softly constrains the endmembers to be similar to library endmembers.

The known endmembers can be identified in the data as pure pixels or be supplied from a spectral library. The method was tested on both simulated data and a real hyperspectral image. The methodology used here can easily be incorporated into other NMF [72] type unmixing or more sophisticated cyclic-descent unmixing methods such as sparsity inducing unmixing methods [122, 123].

The latter two methods make the pure pixel assumption. The endmembers are selected from a dictionary, constructed from the original data matrix. This allows us to unambiguously interpret the endmembers from the data set. Both methods use sparsity regularization to control the number of endmembers in the solution. In the first method, ℓ_2 regularization is used, which is implemented using an iterative soft thresholding algorithm. The second method uses ℓ_0 regularization, which results in a hard thresholding algorithm.

These two methods were compared using simulated data by calculating the SAD between the estimated endmembers and the true endmembers. The method using ℓ_0 regularization performed better than the ℓ_2 method in the sense that the SAD was lower for all simulations.

The ℓ_0 method was also tested using the Urban data set. The abundance maps were clearly associated with distinct macroscopic material visible in the RGB image, although some abundance maps were not associated with only one material.

Chapter 6

Conclusions

In this chapter, the conclusions and main contributions are reviewed. Future research topics are also discussed.

6.1 Main Contributions

The goal of this thesis is to present hyperspectral unmixing methods using the low rank model. The methods are designed to exploit different characteristics of the hyperspectral data. These characteristics are incorporated into objective functions that are then optimized. The main contributions of this thesis are listed here.

6.1.1 Sparse Hyperspectral Unmixing Using ℓ_q Regularization

An iterative cyclic descent unmixing algorithm, using an MM technique, is presented. An ℓ_q sparsity penalty, where $0 \leq q \leq 1$, is used to promote sparseness in the abundances and a first order roughness penalty is used to promote smooth endmembers. Both the degree of sparseness and the roughness can be varied with tuning parameters. The roughness penalty can also be canceled at known discontinuities in the endmember spectra. The ASC can also be softly enforced.

The algorithm was tested using simulated data and the optimal parameters that minimized the reconstruction errors, and the spectral angle distance were found. The simulations showed that choosing $0 < q < 1$ could outperform the ℓ_1 norm, both in terms of reconstruction errors and spectral angle distance.

In the simulations, it was observed that low values of q resulted in slightly better unmixing results when the sparsity was increased. It was also noted that the value of q that gave the best unmixing results decreased when the SNR of that data decreased. Simulations also showed that in low SNR scenarios, the first order roughness penalty could improve the unmixing results.

The algorithm was also tested using both the Urban and Cuprite data sets. The Urban data set was unmixed, both when the ASC was enforced and when it was not enforced. These two results were compared and it was concluded that not enforcing the ASC yielded better unmixing of this data set. The endmembers and their abundance maps in both solutions were easily associated with clearly visible distinct material in the RGB image. The Cuprite data set was unmixed and the most prominent endmembers found were compared to a spectral library.

6.1.2 Hyperspectral Unmixing Using Total Variation and ℓ_q Regularization

A piecewise smooth and sparse unmixing method is proposed. The method minimizes a novel cost function that exploits the sparsity and piecewise smoothness of abundance maps. The method combines ℓ_q $(0 \leq q \leq 1)$ and total variation regularization to achieve sparse and piecewise smooth abundance maps. Both the sparseness and the smoothness of the method can be varied via tuning parameters.

The dyadic expansion is used to decouple the minimization problem. A TV image denoising algorithm is modified to incorporate the ℓ_q sparsity regularization. This modified algorithm is then used by the method. The well known ASC is not used but rather the ENC which constrains the endmembers to have unit norm.

Simulations showed that neither sparse nor smooth regularization, respectively, was able to minimize all three metrics that were used to evaluate the unmixing. A combination of sparseness and smoothness, gave a better solution, where all the metrics were close to their global minima. The simulations also showed that using the ℓ_1 norm can yield a low reconstruction error, but better results w.r.t. the endmembers and abundances could be achieved by using an ℓ_q norm with $q < 1$.

The method was also tested using the Urban and Cuprite data sets. The solution found, when unmixing the Urban data was improved using the proposed regularization. The result obtained using the Cuprite data set was compared to another well established unmixing method and the estimated endmembers were compared to a spectral library. Both methods estimated similar endmembers and these endmembers closely resembled endmembers found in a spectral library.

The method was also used for classification. The abundance maps were used as features for a Random Forest classifier. All of the tuning parameters were selected using cross-validation, with the exception that $q = 1$, which means that an ℓ_1 sparsity regularizer was used. The Indian Pines and Pavia University data sets were classified in this manner.

A combination of sparse and TV regularization gave the best classification results for the Indian Pines. For the Pavia data set, cross-validation chose $h_q = 0$, so in that case only TV regularization was used. The classification accuracies are significantly improved using the proposed method compared to classification accuracies obtained using the features provided by PCA.

6.1.3 Semi-Supervised Hyperspectral Unmixing

Two methods for semi-supervised hyperspectral unmixing are proposed. Supervised and unsupervised unmixing is combined in a way that allows certain known endmembers to be specified, while others are estimated from the data. The abundance maps

are estimated for all the endmembers. These known endmembers can be acquired from a spectral library or extracted from the data. *A priori* information about known endmembers is added to the unmixing model and we estimate endmembers that are not known. Using *A priori* information improves the unmixing results and also reduces the computations.

The first of these methods employs hard regularization, forcing the known endmembers to be exactly identified beforehand. This method also employs the ASC.

The second method improves on the first method by using soft regularization to constrain the endmembers to be similar to the library endmembers. There may be calibration mismatches between the real image spectra and the spectra available from a library. If the library endmembers are identified as pure pixels in the data, they are not noise-free and are thus not pure endmembers. To account for these problems, the estimated endmembers are allowed to deviate from the library endmembers to some degree. This is achieved with soft regularization.

Simulations show that the unmixing can be improved and that softly constraining the supervised endmembers to be similar to specific pure pixels in the data could yield better results than fixing specific pure pixels as the endmembers.

The methods were used to unmix the Urban data set. The results showed that the abundance maps could be improved when using the proposed methods.

6.1.4 Matrix Decomposition Using the Pure Pixel Assumption

It may be appealing to have the endmembers coincide with certain pixels in the image. This is the pure pixel assumption. This guarantees that the endmembers can be fully explained by material seen in the image. Two hyperspectral unmixing methods are proposed that make the pure pixel assumption. The endmembers are selected from a dictionary, constructed from the original data matrix. Both methods use sparsity regularization to control the number of endmembers in the solution. In the first method, ℓ_2 regularization is used, which is implemented using an iterative soft thresholding algorithm. The second method uses vector-ℓ_0 regularization, which results in a hard thresholding algorithm.

These two methods are compared using simulated data, and the method using vector-ℓ_0 regularization is found to perform better than the ℓ_2 method.

The ℓ_0 method was also tested using the Urban data set. The abundance maps were clearly associated with distinct macroscopic material visible in the RGB image, although some abundance maps were not associated with only one material.

6.2 Further work

The work presented here can be extended or continued in a number of ways. Some possible future research topics are listed below.

- Many unmixing methods are highly dependent on both tuning and model parameters. However, parameter estimation in hyperspectral unmixing is a daunting task which has received some attention but there are many unexplored avenues in this field.
- Some hyperspectral images are huge and in the future, they will increase in size and resolution. This will strain the hardware used, and put a constraint on the complexity of the unmixing methods. One possible way to tackle this issue is to use distributed or parallel methods. This is however not a trivial problem. How to split the data, and combine the results, is a interesting research topic that has received some well deserved attention recently. Adapting the methods presented here to work in a parallel or distributed manner is a worthy topic.
- The semi-constrained unmixing method could be further developed to simultaneously attempt to identify the correct endmembers in a spectral library while also identifying the unknown endmembers.
- The methods presented here are applied on remote sensing hyperspectral data. These methods can be applied on other different data, such as medical data, e.g., magnetic resonance images (MRI) and functional MRI (fMRI) signals. It can be interesting to adapt the methods presented here to work with different data.

Appendix

Evaluation Metrics

A

Here are detailed three key metrics that are used to evaluate unmixing results.

To evaluate the quality of hyperspectral unmixing methods, three metrics will be used: the spectral angle distance (SAD), the normalized abundance mean-square error ($\text{nMSE}_{\boldsymbol{S}}$) and the normalized signal reconstruction mean-square error ($\text{nMSE}_{\boldsymbol{AS}}$).

Given an endmember matrix, $\boldsymbol{A}$, where each column represents one endmember, the spectral angle distance (SAD_i) between an estimated endmember, $\widehat{\boldsymbol{a}}_{(i)}$ and a true endmember, $\boldsymbol{a}_{(i)}$ is defined as

$$\text{SAD}_i(\boldsymbol{A}, \widehat{\boldsymbol{A}}) = \arccos\left(\frac{\boldsymbol{a}_{(i)}^{\mathsf{T}}\widehat{\boldsymbol{a}}_{(i)}}{\|\boldsymbol{a}_{(i)}\|\|\widehat{\boldsymbol{a}}_{(i)}\|}\right). \tag{A.1}$$

This metric compares the similarity of an estimated endmember to the true endmember. An SAD_i of zero means that the two endmembers signatures being compared are essentially the same (SAD_i ignores scale and sign). We define the SAD as the average SAD_i of all of the endmembers, that is,

$$\text{SAD} = \frac{1}{r}\sum_{i=1}^{r}\text{SAD}_i(\boldsymbol{A}, \widehat{\boldsymbol{A}}). \tag{A.2}$$

Two mean-square errors are also used, the former being normalized abundance mean-square error, $\text{nMSE}_{\boldsymbol{S}}$, defined as

$$\text{nMSE}_{\boldsymbol{S}} = \frac{\|\boldsymbol{S} - \widehat{\boldsymbol{S}}\|^2}{\|\boldsymbol{S}\|_F^2}. \tag{A.3}$$

This metric compares the abundances estimate abundances, $\widehat{\boldsymbol{S}}$, to the true abundances, $\boldsymbol{S}$. The latter mean-square error is the normalized reconstruction error, measuring how accurately the real noise free data is estimated by the product of the estimated endmembers and abundances,

$$\text{nMSE}_{\boldsymbol{AS}} = \frac{\|\boldsymbol{AS} - \widehat{\boldsymbol{A}}\widehat{\boldsymbol{S}}\|^2}{\|\boldsymbol{AS}\|_F^2}. \tag{A.4}$$

We will refer to the SAD, $\text{nMSE}_{\boldsymbol{S}}$ and $\text{nMSE}_{\boldsymbol{AS}}$ as *the metrics*. These metrics can only be used when the true endmembers and abundances are known, as is the case with simulated data.

Appendix B

Hyperspectral Data

This appendix describes the hyperspectral data sets used in this thesis.

B.1 Urban

The first real data set is a 307 × 307 pixel HYDICE (Hyperspectral Digital Imagery Collection Experiment) hyperspectral image. The area is an urban landscape, showing trees, grass fields, houses, and roads. A ground truth is not available for this data set. The data is composed of 210 spectral bands, has a spectral resolution of 10nm and covers the $400 - 2400$nm spectral range.

Figure B.1: An RGB image of the Urban data set.

B.2 Cuprite

The Cuprite data set is an AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) hyperspectral image of the Cuprite mining district in Nevada. This data is composed of 224 spectral bands, has a spectral resolution of 10nm, and covers the $400 - 2500$nm spectral range. Here we use a 350×350 pixel subset of the whole image. This site has received much attention by researchers and several well documented minerals are exposed in the landscape.

Figure B.2: An RGB image of a 350×350 subset of the Cuprite data set.

B.3 Indian Pines

The Indian pines data set is composed of 145×145 pixels and 220 spectral bands, captured by the AVIRIS sensor. The spatial resolution is 20m and the spectral resolution is 10nm, ranging from 400nm to 2500nm. There are sixteen classes of interest in this data set.

Figure B.3: An RGB image of Indian Pines and the classification ground truth.

B.4 Pavia University

The Pavia University data set was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS-03) sensor. There are 115 spectral band and cover the 400nm to 900nm range, and the spatial resolution is 1.3m. The size of the image is 610×340 pixels. There are nine classes of interest in this data set.

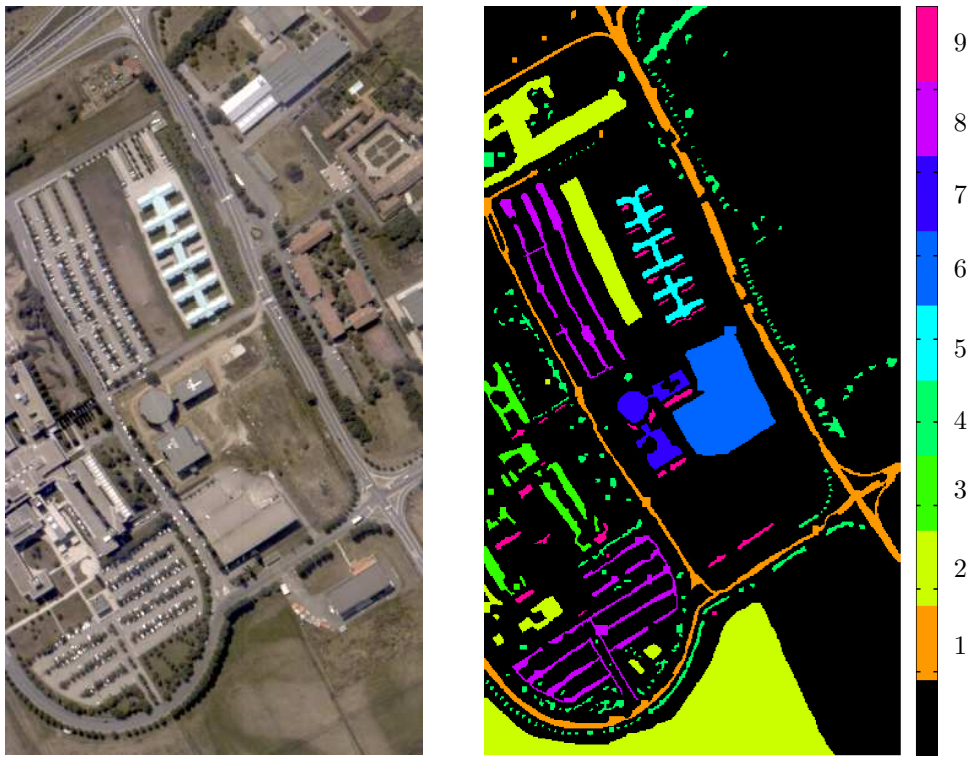


Figure B.4: An RGB image of the Pavia University and the classification ground truth.

Appendix C

Multiplicative Update Rules For NMF

In this appendix the multiplicative update rules for NMF as proposed in [71] are derived.

To formulate the NMF rules, the cost function that is used is defined as

$$J = \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{S}\|^2 = \sum_{p=1}^{P} \frac{1}{2}\|\boldsymbol{y}_{(p)} - \boldsymbol{A}\boldsymbol{s}_{(p)}\|^2. \tag{C.1}$$

C.1 Update Rule For $\boldsymbol{S}$

First, (C.1) is minimized with respect to the rows of $\boldsymbol{S}$. For one row, the minimization problem is

$$\min_{\boldsymbol{s}\geq 0} J(\boldsymbol{s}), \text{ where } J(\boldsymbol{s}) = \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{s}\|^2, \tag{C.2}$$

where the (p) subscript has been dropped for brevity. Using a majorization-minimization [147] (MM) approach, akin to the one used in Section 3.5, the minimization problem becomes

$$\boldsymbol{s}^{(k+1)} = \arg\min_{\boldsymbol{s}\geq 0} F(\boldsymbol{s}) = \arg\min_{\boldsymbol{s}\geq 0} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{A}^{(k)}\boldsymbol{s}\| + \frac{1}{2}(\boldsymbol{s} - \boldsymbol{s}^{(k)})^{\mathsf{T}}\boldsymbol{H}(\boldsymbol{s} - \boldsymbol{s}^{(k)}), \tag{C.3}$$

where $\boldsymbol{s}^{(k)}$ and $\boldsymbol{A}^{(k)}$ are current estimation of $\boldsymbol{s}$ and $\boldsymbol{A}$, respectively,

$$\begin{aligned} \boldsymbol{H} &= \operatorname{diag}(\boldsymbol{x}) - (\boldsymbol{A}^{(k)})^{\mathsf{T}}\boldsymbol{A}^{(k)}, \\ \boldsymbol{x} &= ((\boldsymbol{A}^{(k)})^{\mathsf{T}}\boldsymbol{A}^{(k)}\boldsymbol{s}^{(k)}) \oslash \boldsymbol{s}^{(k)}. \end{aligned}$$

Setting the derivative of $F(\boldsymbol{s})$ to zero, gives

$$((\boldsymbol{A}^{(k)})^{\mathsf{T}}\boldsymbol{A}^{(k)} + \boldsymbol{H})\boldsymbol{s} = (\boldsymbol{A}^{(k)})^{\mathsf{T}}\boldsymbol{y} - \boldsymbol{H}\boldsymbol{s}^{(k)}. \tag{C.4}$$

By using that $\boldsymbol{H}\boldsymbol{s}^{(k)} = 0$ and $((\boldsymbol{A}^{(k)})^{\mathsf{T}}\boldsymbol{A}^{(k)} + \boldsymbol{H}) = \operatorname{diag}(\boldsymbol{x})$, (C.4) can be written,

$$\operatorname{diag}(((\boldsymbol{A}^{(k)})^{\mathsf{T}}\boldsymbol{A}^{(k)}\boldsymbol{s}^{(k)}) \oslash \boldsymbol{s}^{(k)}))\boldsymbol{s} = (\boldsymbol{A}^{(k)})^{\mathsf{T}}\boldsymbol{y} \tag{C.5}$$

$$\Rightarrow \boldsymbol{s} = (\boldsymbol{A}^{(k)})^{\mathsf{T}}\boldsymbol{y} \odot \boldsymbol{s}^{(k)} \oslash ((\boldsymbol{A}^{(k)})^{\mathsf{T}}\boldsymbol{A}^{(k)}\boldsymbol{s}^{(k)}). \tag{C.6}$$

This rule can be used for all rows of $\boldsymbol{S}$, giving

$$\boldsymbol{S}^{(k+1)} = \boldsymbol{S}^{(k)} \odot (\boldsymbol{A}^{(k)})^{\mathsf{T}}\boldsymbol{Y}) \oslash ((\boldsymbol{A}^{(k)})^{\mathsf{T}}\boldsymbol{A}^{(k)}\boldsymbol{S}^{(k)}). \tag{C.7}$$

C.2 Update Rule For $\boldsymbol{A}$

The the cost function, (C.1), can be written as

$$J = \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{S}\|^2 = \frac{1}{2}\|\boldsymbol{Y}^\mathsf{T} - \boldsymbol{S}^\mathsf{T}\boldsymbol{A}^\mathsf{T}\|^2, \tag{C.8}$$

An update rule for each row in $\boldsymbol{A}$ (e.g., column in $\boldsymbol{A}^\mathsf{T}$) can be derived using the same MM approach as was done for $\boldsymbol{S}$. By substituting $\boldsymbol{Y} \to \boldsymbol{Y}^\mathsf{T}$, $\boldsymbol{A}^{(k)} \to (\boldsymbol{S}^{(k)})^\mathsf{T}$, and $\boldsymbol{S}^{(k)} \to (\boldsymbol{A}^{(k)})^\mathsf{T}$ in equation (C.7), the update rule for $\boldsymbol{A}^\mathsf{T}$ is

$$(\boldsymbol{A}^{(k)})^\mathsf{T} = (\boldsymbol{A}^{(k)})^\mathsf{T} \odot \boldsymbol{S}^{(k)}\boldsymbol{Y}^\mathsf{T} \oslash (\boldsymbol{S}^{(k)}(\boldsymbol{S}^{(k)})^\mathsf{T}(\boldsymbol{A}^{(k)})^\mathsf{T}), \tag{C.9}$$

and by transposing, the update rule for $\boldsymbol{A}$ is found to be

$$\boldsymbol{A}^{(k+1)} = \boldsymbol{A}^{(k)} \odot (\boldsymbol{Y}(\boldsymbol{S}^{(k)})^\mathsf{T}) \oslash (\boldsymbol{A}^{(k)}\boldsymbol{S}^{(k)}(\boldsymbol{S}^{(k)})^\mathsf{T}). \tag{C.10}$$

Bibliography

[1] R. Green, M. Eastwood, C. Sarture, T. Chrien, M. Aronsson, B. Chippendale, J. Faust, B. Pavri, C. Chovit, M. Solis, and et al., "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sensing of Environment*, vol. 65, no. 3, pp. 227–248, 1998. [Cited on page 1]

[2] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 354 –379, April 2012. [Cited on pages 1, 4, 6, 21, 48, 71, and 80]

[3] A. Plaza, J. Benediktsson, J. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sensing of Environment*, vol. 113, pp. 110–122, 2009. [Cited on page 2]

[4] B.-C. Gao C.O. Davis and A.F.H. Goetz, "A review of atmospheric correction techniques for hyperspectral remote sensing of land surfaces and ocean color," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2006, pp. 1979–1981. [Cited on page 3]

[5] B. Datt, T.R. McVicar, T.G. Van Niel, D.L.B. Jupp, and J.S. Pearlman, "Preprocessing eo-1 hyperion hyperspectral data to support the application of agricultural indexes," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 6, pp. 1246–1259, June 2003. [Cited on page 3]

[6] S. Liang and H. Fang, "An improved atmospheric correction algorithm for hyperspectral remotely sensed imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 1, no. 2, pp. 112–117, April 2004. [Cited on page 3]

[7] J.P. Kerekes and J.E. Baum, "Hyperspectral imaging system modeling," *Lincoln Laboratory Journal, Special Issue on Spectral Imaging*, vol. 1, no. 14, pp. 117–130, 2003. [Cited on page 3]

[8] D.A. Landgrebe and E. Malaret, "Noise in remote-sensing systems: The effect on classification error," *IEEE Transactions on Geoscience and Remote Sensing*, vol. GE-24, no. 2, pp. 294–300, March 1986. [Cited on page 3]

[9] N. Acito, M. Diani, and G. Corsini, "Signal-dependent noise modeling and model parameter estimation in hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 8, pp. 2957–2971, Aug 2011. [Cited on page 3]

[10] Y. Qian and M. Ye, "Hyperspectral imagery restoration using nonlocal spectral-spatial structured sparse representation with noise estimation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 2, pp. 499–515, April 2013. [Cited on page 3]

[11] L. Gómez-Chova, L. Alonso, L. Guanter, G. Camps-Valls, J. Calpe, and J. Moreno, "Correction of systematic spatial noise in push-broom hyperspectral sensors: application to CHRIS/PROBA images," *Applied Optics*, vol. 47, no. 28, pp. F46–F60, Oct 2008. [Cited on page 3]

[12] D.J. Brady, *Optical Imaging and Spectroscopy*, Wiley, Hoboken, NJ, first edition, 2009. [Cited on page 3]

[13] N. Keshava and J.F. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 44 –57, Jan. 2002. [Cited on pages 3 and 4]

[14] D.C. Heinz and Chein-I Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 3, pp. 529–545, Mar 2001. [Cited on pages 4 and 15]

[15] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 3, pp. 765–777, March 2007. [Cited on pages 4, 6, 7, and 15]

[16] A. Huck, M. Guillaume, and J. Blanc-Talon, "Minimum dispersion constrained nonnegative matrix factorization to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 6, pp. 2590–2602, June 2010. [Cited on pages 4, 7, and 15]

[17] X. Liu, W. Xia, B. Wang, and L. Zhang, "An approach based on constrained nonnegative matrix factorization to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 2, pp. 757–772, Feb 2011. [Cited on pages 4, 7, and 15]

[18] Z. Yang, G. Zhou, S. Xie, S. Ding, J.-M. Yang, and J. Zhang, "Blind spectral unmixing based on sparse nonnegative matrix factorization," *IEEE Transactions on Image Processing*, vol. 20, no. 4, April 2011. [Cited on pages 4, 7, and 15]

[19] N. Dobigeon, J.-Y. Tourneret, C. Richard, J.C.M. Bermudez, S. McLaughlin, and A.O. Hero, "Nonlinear unmixing of hyperspectral images: Models and algorithms," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 82–94, Jan 2014. [Cited on page 5]

[20] R. Heylen, M. Parente, and P. Gader, "A review of nonlinear hyperspectral unmixing methods," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 1844–1868, June 2014. [Cited on page 5]

[21] I. Meganem, Y. Deville, S. Hosseini, P. Deliot, and X. Briottet, "Linear-quadratic blind source separation using nmf to unmix urban hyperspectral images," *IEEE Transactions on Signal Processing*, vol. 62, no. 7, pp. 1822–1833, April 2014. [Cited on page 5]

[22] I. Meganem, P. Deliot, X. Briottet, Y. Deville, and S. Hosseini, "Linear-quadratic mixing model for reflectances in urban environments," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 1, pp. 544–558, Jan 2014. [Cited on page 5]

[23] C. Bateson and B. Curtiss, "A tool for manual endmember selection and spectral unmixing," *Remote Sensing of Environment*, pp. 3–6, 1993. [Cited on page 5]

[24] C. Bateson and B. Curtiss, "A method for manual endmember selection and spectral unmixing," *Remote Sensing of Environment*, vol. 55, no. 3, pp. 229–243, 1996. [Cited on page 5]

[25] C. Bateson, G.P. Asner, and C.A. Wessman, "Endmember bundles: a new approach to incorporating endmember variability into spectral mixture analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 2, pp. 1083–1094, Mar 2000. [Cited on page 5]

[26] B. Somers, M. Zortea, A. Plaza, and G.P. Asner, "Automated extraction of image-based endmember bundles for improved spectral unmixing," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 5, no. 2, pp. 396–408, April 2012. [Cited on page 5]

[27] B. Somers, G.Asner, . Tits, and P. Coppin, "Endmember variability in spectral mixture analysis: A review," *Remote Sensing of Environment*, vol. 115, no. 7, pp. 1603–1616, 2011. [Cited on page 5]

[28] J. Settle, "On the effect of variable endmember spectra in the linear mixture model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 2, pp. 389–396, Feb 2006. [Cited on page 5]

[29] L. Zhuang, B. Zhang, L. Gao, J. Li, and A. Plaza, "Normal endmember spectral unmixing method for hyperspectral imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. PP, no. 99, pp. 1–9, 2014. [Cited on page 5]

[30] A. Zare and K.C. Ho, "Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing," *Signal Processing Magazine, IEEE*, vol. 31, no. 1, pp. 95–104, Jan 2014. [Cited on page 5]

[31] J. Bien, Y. Xu, and M. W. Mahoney, "CUR from a Sparse Optimization Viewpoint," *Advances in Neural Information Processing Systems 23*, pp. 217–225, 2010. [Cited on pages 6 and 80]

[32] Michael W Mahoney and Petros Drineas, "CUR matrix decompositions for improved data analysis," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 3, pp. 697–702, 2009. [Cited on pages 6 and 80]

[33] J.W. Boardman, "Geometric mixture analysis of imaging spectrometery data," *Proc. Int. Geoscience and Remote Sensing Symp.*, vol. 4, pp. 2369–2371, 1994. [Cited on page 6]

[34] M.E. Winter, "N-findr: An algorithm for fast autonomous spectral endmember determination in hyperspectral data," *Proc. SPIE Image Spectrometry V*, vol. 3753, pp. 266–277, 1999. [Cited on page 6]

[35] Q. Du, H. Ren, and C.-I. Chang, "A comparative study for orthogonal subspace projection and constrained energy minimization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 6, pp. 1525–1529, June 2003. [Cited on page 6]

[36] John H. Gruninger, Anthony J. Ratkowski, and Michael L. Hoke, "The sequential maximum angle convex cone (smacc) endmember model," *Proc. SPIE*, vol. 5425, pp. 1–14, 2004. [Cited on page 6]

[37] J.M.P. Nascimento and J.M. Bioucas Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, April 2005. [Cited on pages 6, 29, 44, 53, and 61]

[38] A. Plaza and C.-I. Chang, "Impact of initialization on design of endmember extraction algorithms," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 11, pp. 3397–3407, Nov 2006. [Cited on page 6]

[39] C.-I Chang and A. Plaza, "A fast iterative algorithm for implementation of pixel purity index," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 1, pp. 63–67, Jan 2006. [Cited on page 6]

[40] C.-I. Chang, C-C. Wu, W-M. Liu, and Y-C. Ouyang, "A new growing method for simplex-based endmember extraction algorithm," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 10, pp. 2804–2819, Oct 2006. [Cited on page 6]

[41] A. Zare and P. Gader, "Hyperspectral band selection and endmember detection using sparsity promoting priors," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 2, pp. 256–260, April 2008. [Cited on page 6]

[42] M. Zortea and A. Plaza, "A quantitative and comparative analysis of different implementations of n-findr: A fast endmember extraction algorithm," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 4, pp. 787–791, Oct 2009. [Cited on page 6]

[43] C.-I Chang, C.-C. Wu, C.-S. Lo, and M.-L. Chang, "Real-time simplex growing algorithms for hyperspectral endmember extraction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 4, pp. 1834–1850, April 2010. [Cited on page 6]

[44] F. Schmidt, A. Schmidt, E. Tréguier, M. Guiheneuf, S. Moussaoui, and N. Dobigeon, "Implementation strategies for hyperspectral unmixing using bayesian source separation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 11, pp. 4003–4013, Nov 2010. [Cited on page 6]

[45] W. Xiong, C.-I Chang, C.-C Wu, K. Kalpakis, and H.M. Chen, "Fast algorithms to implement n-findr for hyperspectral endmember extraction," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 4, no. 3, pp. 545–564, Sept 2011. [Cited on page 6]

[46] T.-H. Chan, Ma W, K, A. Ambikapathi, and C.-Y. Chi, "A simplex volume maximization framework for hyperspectral endmember extraction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4177–4193, Nov 2011. [Cited on page 6]

[47] S. Dowler and M. Andrews, "On the convergence of n-findr and related algorithms: To iterate or not to iterate?," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 1, pp. 4–8, Jan 2011. [Cited on page 6]

[48] O. Duran and M. Petrou, "Robust endmember extraction in the presence of anomalies," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, pp. 1986–1996, June 2011. [Cited on page 6]

[49] J. Cui, X. Li, and L. Zhao, "Linear mixture analysis for hyperspectral imagery in the presence of less prevalent materials," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 7, July 2013. [Cited on page 6]

[50] M.C Castan, J. Chanussot, V. Achard, X. Briottet, and M. Shimoni, "A physics-based unmixing method to estimate subpixel temperatures on mixed pixels," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 4, pp. 1894–1906, April 2015. [Cited on page 6]

[51] X. Wu, B. Huang, A. Plaza, Y. Li, and C. Wu, "Real-time implementation of the pixel purity index algorithm for endmember identification on gpus," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 5, pp. 955–959, May 2014. [Cited on page 6]

[52] N. Gillis, Da Kuang, and Haesun Park, "Hierarchical clustering of hyperspectral images using rank-two nonnegative matrix factorization," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 53, no. 4, pp. 2066–2078, April 2015. [Cited on pages 6 and 7]

[53] J. Li, I. Dopido, P. Gamba, and A. Plaza, "Complementarity of discriminative classifiers and spectral unmixing techniques for the interpretation of hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2899–2912, May 2015. [Cited on page 6]

[54] A.A. Green, M. Berman, P. Switzer, and M.D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 26, no. 1, pp. 65–74, Jan 1988. [Cited on page 6]

[55] A. Ifarraguerri and C.-I. Chang, "Multispectral and hyperspectral image analysis with convex cones," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 2, pp. 756–770, Mar 1999. [Cited on page 6]

[56] M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Dunne, and J.F. Huntington, "Ice: a statistical approach to identifying endmembers in hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 10, pp. 2085–2095, Oct 2004. [Cited on page 6]

[57] A. Zare and P. Gader, "Sparsity promoting iterated constrained endmember detection in hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 3, pp. 446–450, July 2007. [Cited on page 6]

[58] J. Li and J.M. Bioucas-Dias, "Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data," in *IEEE International Geoscience and Remote Sensing Symposium*, July 2008, vol. 3, pp. III – 250–III – 253. [Cited on page 6]

[59] J.M. Bioucas-Dias, "A variable splitting augmented lagrangian approach to linear spectral unmixing," in *Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (Whispers)*, Aug 2009, pp. 1–4. [Cited on page 6]

[60] T.-H. Chan, C.-Y. Chi, Y.-M. Huang, and W.-K. Ma, "A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, Nov 2009. [Cited on page 6]

[61] A. Ambikapathi, T.-H. Chan, W.-K. Ma, and C.-Y. Chi, "Chance-constrained robust minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4194–4209, Nov 2011. [Cited on page 6]

[62] M. Arngren, M. N. Schmidt, and J. Larsen, "Unmixing of hyperspectral images using bayesian non-negative matrix factorization with volume prior," *Journal of Signal Processing Systems*, vol. 65, no. 3, pp. 479–496, 2011. [Cited on page 6]

[63] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, no. 3, pp. 287–314, Apr. 1994. [Cited on page 7]

[64] A. Hyvarinen, J. Karhunen, and Erkki O., *Independent Component Analysis*, Wiley-Interscience, 1st edition, May 2001. [Cited on pages 7 and 72]

[65] L. Parra, K.-R. Mueller, C. Spence, A. Ziehe, and P. Sajda, "Unmixing hyperspectral data," *Adv. Neural Inf. Process. Systems*, vol. 12, pp. 942–948, 2000. [Cited on page 7]

[66] J. Bayliss, J. A. Gualtieri, and R. Cromp, "Analysing hyperspectral data with independent component analysis," *Proc. SPIE*, vol. 3240, pp. 133–143, 1997. [Cited on page 7]

[67] C. Chen and X. Zhang, "Independent component analysis for remote sensing study," *Proc. SPIE*, vol. 3871, pp. 150–158, 1999. [Cited on page 7]

[68] T. M. Tu, "Unsupervised signature extraction and separation in hyperspectral images: A noise-adjusted fast independent component analysis approach," *Opt. Eng.*, vol. 39, no. 4, pp. 897–906, 2000. [Cited on page 7]

[69] N. Kosaka and Y. Kosugi, "ICA aided linear spectral mixture analysis of agricultural remote sensing images," *Proc. 4th Int. Symp. Independent Component Analysis and Blind Signal Separation*, pp. 221–226, 2003. [Cited on page 7]

[70] J.M.P. Nascimento and J.M. Bioucas Dias, "Does independent component analysis play a role in unmixing hyperspectral data?," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 43, no. 1, pp. 175–187, Jan. 2005. [Cited on page 7]

[71] Daniel D. Lee and H. Sebastian Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct 1999. [Cited on pages 7, 15, and 97]

[72] D.D. Lee and S.H. Seung, "Algorithms for Non-negative Matrix Factorization," *NIPS*, pp. 556–562, 2000. [Cited on pages 7, 15, 16, and 86]

[73] H Zhou, B Wang, and LM Zhang, "A new scheme for blind decomposition of mixed pixels based on non-negative matrix factorization," in *ADVANCES IN NEURAL NETWORKS - ISNN*, 2005, vol. 3497, pp. 651–658. [Cited on page 7]

[74] S.A. Robila and L. Maciack, "New approaches for feature extraction in hyperspectral imagery," in *Systems, Applications and Technology Conference, LISAT*, May 2006, pp. 1–7. [Cited on page 7]

[75] X. Tao, B. Wang, L. Zhang, and J.Q. Zhang, "A new scheme for decomposition of mixed pixels based on nonnegative matrix factorization," in *IEEE International Geoscience and Remote Sensing Symposium, IGARSS*, July 2007, pp. 1759–1762. [Cited on page 7]

[76] Y. Qian, S. Jia, J. Zhou, and A. Robles-Kelly, "Hyperspectral unmixing via $L_{1/2}$ sparsity-constrained nonnegative matrix factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4282 –4297, Nov. 2011. [Cited on pages 7, 15, 19, 21, 22, and 49]

[77] E. Esser, M. Möller, S. Osher, G. Sapiro, and J. Xin, "A convex model for non-negative matrix factorization and dimensionality reduction on physical space," *Arxiv preprint*, vol. stat.ML, pp. 14, 2011. [Cited on pages 7 and 80]

[78] R. Rajabi and H. Ghassemian, "Spectral unmixing of hyperspectral imagery using multilayer nmf," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 1, pp. 38–42, Jan 2015. [Cited on pages 7 and 15]

[79] A. Zymnis, S.J. Kim, J. Skaf, M. Parente, and S. Boyd, "Hyperspectral image unmixing via alternating projected subgradients," in *Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers, 2007. ACSSC 2007*, 2007, pp. 1164–1168. [Cited on pages 7 and 22]

[80] D. Stein, "Application of the normal compositional model to the analysis of hyperspectral imagery," in *IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, Oct 2003, pp. 44–51. [Cited on page 7]

[81] O. Eches, N. Dobigeon, C. Mailhes, and J.-Y. Tourneret, "Bayesian estimation of linear mixtures using the normal compositional model. application to hyperspectral imagery," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1403–1413, June 2010. [Cited on page 7]

[82] A. Zare, P. Gader, and G. Casella, "Sampling piecewise convex unmixing and endmember extraction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 3, pp. 1655–1665, March 2013. [Cited on page 7]

[83] Xiaoxiao Du, A. Zare, P. Gader, and D. Dranishnikov, "Spatial and spectral unmixing using the beta compositional model," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 1994–2003, June 2014. [Cited on page 7]

[84] P. Bosdogianni, M. Petrou, and J. Kittler, "Mixture models with higher order moments," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 2, pp. 341–353, Mar 1997. [Cited on page 7]

[85] M.-D. Iordache, J.M. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Journal of Geoscience and Remote Sensing*, vol. 50, no. 11, pp. 4484–4502, Nov 2012. [Cited on page 8]

[86] Y.C. Pati, R. Rezaiifar, and P.S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *IEEE Asil. Conf. on Sig., Sys. and Comp. (ASSC)*, Nov 1993, vol. 1, pp. 40–44. [Cited on page 8]

[87] S. Chen, D. Donoho, , and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev*, vol. 43, no. 4, pp. 129–159, 2001. [Cited on page 8]

[88] M.-D. Iordache, J.M. Bioucas-Dias, A. Plaza, and B. Somers, "Music-csr: Hyperspectral unmixing via multiple signal classification and collaborative sparse regression," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 7, pp. 4364–4382, July 2014. [Cited on page 8]

[89] M.-D. Iordache, J.M. Bioucas-Dias, and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 1, pp. 341–354, Jan 2014. [Cited on page 8]

[90] J.M. Bioucas-Dias and M.A.T. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, June 2010, pp. 1–4. [Cited on page 8]

[91] M.-D. Iordache, J.M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, pp. 2014–2039, 2011. [Cited on pages 8, 22, and 72]

[92] M.-D. Iordache, L. Tits, J.M. Bioucas-Dias, A. Plaza, and B. Somers, "A dynamic unmixing framework for plant production system monitoring," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2016–2034, June 2014. [Cited on page 8]

[93] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Unmixing sparse hyperspectral mixtures," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2009, vol. 4, pp. IV–85–IV–88. [Cited on page 8]

[94] M.-D. Iordache, J.M. Bioucas-Dias, and A. Plaza, "Total variation regulatization in sparse hyperspectral unmixing," in *Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, June 2011, pp. 1–4. [Cited on page 8]

[95] M.-D. Iordache, J.M. Bioucas-Dias, and A. Plaza, "Hyperspectral unmixing with sparse group lasso," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2011, pp. 3586–3589. [Cited on pages 8 and 80]

[96] M.-D. Iordache, L. Tits, B. Somers, and A. Plaza, "Improved signal unmixing of vegetation using sparse group selection," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2013, pp. 3018–3021. [Cited on page 8]

[97] M.-D. Iordache, J.M. Bioucas-Dias, and A. Plaza, "Collaborative sparse unmixing of hyperspectral data," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2012, pp. 7488–7491. [Cited on page 8]

[98] J. Eckstein and D. Bertsekas, "On the douglas-rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Progr.*, vol. 5, pp. 293–318, 1992. [Cited on page 8]

[99] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1-4, pp. 259–268, Nov. 1992. [Cited on pages 8 and 47]

[100] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical Imaging and Vision*, vol. 20, pp. 89–97, 2004. [Cited on pages 8 and 47]

[101] J.M. Bioucas-Dias and J.M.P. Nascimento, "Hyperspectral subspace identification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 8, pp. 2435 –2445, Aug. 2008. [Cited on pages 8, 9, 40, and 44]

[102] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, Jan. 2003. [Cited on page 9]

[103] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, March 1978. [Cited on page 9]

[104] C. M. Stein, "Estimation of the Mean of a Multivariate Normal Distribution," *The Annals of Statistics*, vol. 9, no. 6, 1981. [Cited on page 9]

[105] V. Solo, "A sure-fired way to choose smoothing parameters in ill-conditioned inverse problems," *International Conference on Image Processing*, vol. 3, pp. 89–92, Sep 1996. [Cited on page 9]

[106] C.-I Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 3, pp. 608–619, March 2004. [Cited on pages 9 and 61]

[107] M.O. Ulfarsson and V. Solo, "Vector ℓ_0 sparse variable pca," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 1949–1958, May 2011. [Cited on page 9]

[108] M.O. Ulfarsson and V. Solo, "Dimension estimation in noisy pca with sure and random matrix theory," *IEEE Transactions on Signal Processing*, vol. 56, no. 12, pp. 5804–5816, Dec 2008. [Cited on page 9]

[109] M.O. Ulfarsson and V. Solo, "Rank selection in noisy pca with sure and random matrix theory," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, March 2008, pp. 3317–3320. [Cited on page 9]

[110] M.O. Ulfarsson and V. Solo, "Tuning parameter selection for underdetermined reduced-rank regression," *IEEE Signal Processing Letters*, vol. 20, no. 9, pp. 881–884, Sept 2013. [Cited on page 9]

[111] F. Palsson, M.O. Ulfarsson, and J.R. Sveinsson, "Sparse gaussian noisy independent component analysis," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 4224–4228. [Cited on page 9]

[112] I.T. Jolliffe, *Principal Component Analysis*, Springer, New York, NY, 2nd edition, October 2002. [Cited on page 9]

[113] N. Keshava, "Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 42, no. 7, pp. 1552–1565, July 2004. [Cited on page 9]

[114] C.-I Chang and S. Wang, "Constrained band selection for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 6, pp. 1575–1585, June 2006. [Cited on page 9]

[115] G.H. Golub and C. Reinsch, "Singular value decomposition and least squares solutions," *Numerische Mathematik*, vol. 14, no. 5, pp. 403–420, 1970. [Cited on page 9]

[116] J.B. Lee, A.S. Woodyatt, and M. Berman, "Enhancement of high spectral resolution remote-sensing data by a noise-adjusted principal components transform," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 28, no. 3, pp. 295–304, May 1990. [Cited on page 9]

[117] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 40–51, Jan 2007. [Cited on page 9]

[118] N. Falco, J.A. Benediktsson, and L. Bruzzone, "A study on the effectiveness of different independent component analysis algorithms for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2183–2199, June 2014. [Cited on page 9]

[119] C. Lee and D.A. Landgrebe, "Decision boundary feature extraction for neural networks," *IEEE Transactions on Neural Networks*, vol. 8, no. 1, pp. 75–83, Jan 1997. [Cited on page 10]

[120] C. Lee and D.A. Landgrebe, "Feature extraction based on decision boundaries," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 4, pp. 388–400, Apr 1993. [Cited on page 10]

[121] B.-C. Kuo and D.A. Landgrebe, "Nonparametric weighted feature extraction for classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 5, pp. 1096–1105, May 2004. [Cited on page 10]

[122] J. Sigurdsson, M.O. Ulfarsson, and J.R. Sveinsson, "Smooth and sparse hyperspectral unmixing using an l_0 penalty," *Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, June 2013. [Cited on pages 10, 11, 21, 22, 49, and 86]

[123] J. Sigurdsson, M.O. Ulfarsson, and J.R. Sveinsson, "Hyperspectral unmixing with ℓ_q regularization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 11, pp. 6793–6806, Nov 2014. [Cited on pages 10, 11, 21, 49, and 86]

[124] J. Sigurdsson, M.O. Ulfarsson, J.R. Sveinsson, and J.A. Benediktsson, "A smooth hyperspectral unmixing method using cyclic descent," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2012, pp. 3082–3085. [Cited on pages 10, 11, 22, and 49]

[125] J. Sigurdsson, M.O. Ulfarsson, and J.R. Sveinsson, "Hyperspectral image unmixing using dyadic cyclic descent, total variation and ℓ_q sparse regularization," submitted. [Cited on pages 10 and 11]

[126] J. Sigurdsson, M.O. Ulfarsson, and J.R. Sveinsson, "Total variation and ℓ_q based hyperspectral unmixing for feature extraction and classification," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2015. [Cited on page 11]

[127] J. Sigurdsson, M.O. Ulfarsson, and J.R. Sveinsson, "Semi-supervised hyperspectral unmixing," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2014, pp. 3458–3461. [Cited on pages 11 and 12]

[128] J. Sigurdsson, M.O. Ulfarsson, and J.R. Sveinsson, "Endmember constrained semi-supervised hyperspectral unmixing," *Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, June 2014. [Cited on pages 11 and 12]

[129] J. Sigurdsson, M.O. Ulfarsson, J.R. Sveinsson, and J.A. Benediktsson, "Sparse representation of hyperspectral data using CUR matrix decomposition," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2013, pp. 433–436. [Cited on pages 11 and 12]

[130] J. Sigurdsson, M.O. Ulfarsson, J.R. Sveinsson, and J.A. Benediktsson, "Smooth spectral unmixing using total variation regularization and a first order roughness penalty," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2013, pp. 2160–2163. [Cited on pages 12 and 49]

[131] D. Heinz, C.-I. Chang, and M.L.G. Althouse, "Fully constrained least-squares based linear unmixing [hyperspectral image classification]," in *IEEE International Geoscience and Remote Sensing Symposium.*, 1999, vol. 2, pp. 1401–1403 vol.2. [Cited on page 14]

[132] R.J. Hanson K.H. Haskell, *An algorithm for linear least squares problems with equality and nonnegativity constraints*, Sandia Laboratories, Albuquerque, N.Mex, 1978. [Cited on page 14]

[133] Richard J. Hanson and Karen H. Haskell, "Algorithm 587: Two algorithms for the linearly constrained least squares problem," *ACM Transactions on Mathematical Software*, vol. 8, no. 3, pp. 323–333, Sep 1982. [Cited on page 14]

[134] D. Donoho and V. Stodden, "When does non-negative matrix factorization give a correct decomposition into parts?," in *Advances in neural information processing systems*, 2003. [Cited on page 16]

[135] D.P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, MA, 1999. [Cited on pages 17, 24, and 73]

[136] C.-J. Lin, "Projected gradient methods for nonnegative matrix factorization," *Neural Comput.*, vol. 19, no. 10, pp. 2756–2779, Oct. 2007. [Cited on page 17]

[137] F. Sha, L.K. Saul, and D.D. Lee, "Multiplicative updates for nonnegative quadratic programming in support vector machines," *Advances in Neural Information Processing Systems*, vol. 15, 2002. [Cited on pages 17, 26, and 73]

[138] P.O. Hoyer, "Non-negative sparse coding," in *Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing*, 2002, pp. 557–565. [Cited on page 18]

[139] J. Eggert and E. Korner, "Sparse coding and nmf," in *Proceedings of the IEEE International Joint Conference on Neural Networks*, July 2004, vol. 4, pp. 2529–2533. [Cited on page 18]

[140] S. Alliney and S.A Ruzinsky, "An algorithm for the minimization of mixed l1 and l2 norms with application to bayesian estimation," *IEEE Transactions on Signal Processing*, vol. 42, no. 3, pp. 618–627, Mar 1994. [Cited on page 21]

[141] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society (Series B)*, vol. 58, pp. 267–288, 1996. [Cited on page 21]

[142] D. Gu, A.R. Gillespie, A.B. Kahle, and F.D. Palluconi, "Autonomous atmospheric compensation (aac) of high resolution hyperspectral thermal infrared remote-sensing imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 6, pp. 2557 – 2570, Nov. 2000. [Cited on page 21]

[143] G. Marjanovic and V. Solo, "On l_q optimization and matrix completion," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5714–5724, 2012. [Cited on pages 22, 23, 25, 36, 50, and 51]

[144] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Processing Letters*, vol. 14, no. 10, pp. 707–710, 2007. [Cited on page 22]

[145] J. Sigurdsson and M.O. Ulfarsson, "Smooth noisy pca using a 1st order roughness penalty," in *Machine Learning for Signal Processing (MLSP), 2010 IEEE International Workshop on*, 2010, pp. 325–330. [Cited on page 22]

[146] P. J. Green and B. W. Silverman, *Nonparametric Regression and Generalized Linear Models. A Roughness Penalty Approach*, Chapman and Hall, New York, NY, first edition, 1994. [Cited on page 24]

[147] D.R. Hunter and K. Lange, "A tutorial on MM algorithms," *The American Statistician,*, pp. 30–37, 2004. [Cited on pages 24, 73, 81, and 97]

[148] I.W. Selesnick, "Sparse signal restoration," http://cnx.org/content/m32168, Sep. 2009. [Cited on pages 24, 73, and 81]

[149] J.O. Ramsay and B.W. Silverman, *Functional Data Analysis*, Springer Series in Statistics. Springer, 2nd edition, June 2005. [Cited on page 27]

[150] M.-D. Iordache, J.M. Bioucas-Dias, and A Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 11, Nov 2012. [Cited on pages 44 and 71]

[151] R.N. Clark and G.A. Swayze, "Evolution in imaging spectroscopy analysis and sensor signal-to-noise: An examination of how far we have come," *Summaries of the Sixth Annual JPL Airborne Earth Science Workshop*, March 1996. [Cited on pages 44 and 61]

[152] A. Zare, O. Bchir, H. Frigui, and P. Gader, "A comparison of deterministic and probabilistic approaches to endmember representation," in *2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2010, pp. 1–4. [Cited on page 44]

[153] J.M. Bioucas-Dias and M.A.T. Figueiredo, "A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration," *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 2992–3004, Dec 2007. [Cited on page 47]

[154] J.-L. Starck, M. Elad, and D.L. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE Transactions on Image Processing*, vol. 14, no. 10, pp. 1570–1582, Oct 2005. [Cited on page 47]

[155] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2419–2434, Nov 2009. [Cited on page 47]

[156] M.V. Afonso, J.M. Bioucas-Dias, and M.A.T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2345–2356, Sept 2010. [Cited on page 47]

[157] M.V. Afonso, J.M. Bioucas-Dias, and M.A.T. Figueiredo, "An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 681–695, March 2011. [Cited on page 47]

[158] T. Goldstein and S. Osher, "The split bregman method for l1-regularized problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 323–343, 2009. [Cited on pages 47, 50, and 52]

[159] J.M. Bioucas-Dias, M.A.T. Figueiredo, and J.P. Oliveira, "Total variation-based image deconvolution: a majorization-minimization approach," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2006, vol. 2. [Cited on pages 47 and 49]

[160] M.O. Ulfarsson and V. Solo, "Sparse component analysis via dyadic cyclic descent," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Jul 2014. [Cited on page 49]

[161] Y. Wang, W. Yin, and Y. Zhang, "A fast algorithm for image deblurring with total variation regularization," *CAAM Technical Reports*, Jul 2007. [Cited on page 51]

[162] B. Leo, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001. [Cited on page 65]

[163] P. Ghamisi, J.A. Benediktsson, G. Cavallaro, and A. Plaza, "Automatic framework for spectral-spatial classification based on supervised feature extraction and morphological attribute profiles," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2147–2160, June 2014. [Cited on page 65]

[164] V. D. Calhoun, T. Adali, M. C. Stevens, K. A. Kiehl, and J. J. Pekar, "Semi-blind ica of fMRI: A method for utilizing hypothesis-derived time courses in a spatial ica analysis," *Neuroimage 25(2)*, pp. 527–538, Apr. 2005. [Cited on page 72]

[165] P. Bai, H.P. Shen, X.M. Huang, and Y. Truong, "A supervised singular value decomposition for independent component analysis of fMRI," *Statistica Sinica 18*, pp. 1233–1253, 2008. [Cited on page 72]

Index