LINEAR HYPERSONTICAL UNMIXING USING
\( \ell_0 \)-NORM APPROXIMATIONS AND
NONNEGATIVE MATRIX FACTORIZATION

by

YASER ESMAEILI SALEHANI

A thesis submitted to the
Department of Electrical and Computer Engineering
in conformity with the requirements for
the degree of Doctor of Philosophy

Queen’s University
Kingston, Ontario, Canada
October 2016

Copyright © Yaser Esmaeili Salehani, 2016
Abstract

Spectral unmixing (SU) is a technique to characterize mixed pixels of the hyperspectral images measured by remote sensors. Most of the existing spectral unmixing algorithms are developed using the linear mixing models. Since the number of endmembers/materials present at each mixed pixel is normally scanty compared with the number of total endmembers (the dimension of spectral library), the problem becomes sparse. This thesis introduces sparse hyperspectral unmixing methods for the linear mixing model through two different scenarios. In the first scenario, the library of spectral signatures is assumed to be known and the main problem is to find the minimum number of endmembers under a reasonable small approximation error. Mathematically, the corresponding problem is called the $\ell_0$-norm problem which is NP-hard problem. Our main study for the first part of thesis is to find more accurate and reliable approximations of $\ell_0$-norm term and propose sparse unmixing methods via such approximations. The resulting methods are shown considerable improvements to reconstruct the fractional abundances of endmembers in comparison with state-of-the-art methods such as having lower reconstruction errors. In the second part of the thesis, the first scenario (i.e., dictionary-aided semiblind unmixing scheme) will be generalized as the blind unmixing scenario that the library of spectral signatures is also estimated. We apply the nonnegative matrix factorization (NMF)
method for proposing new unmixing methods due to its noticeable supports such as considering the nonnegativity constraints of two decomposed matrices. Furthermore, we introduce new cost functions through some statistical and physical features of spectral signatures of materials (SSoM) and hyperspectral pixels such as the collaborative property of hyperspectral pixels and the mathematical representation of the concentrated energy of SSoM for the first few subbands. Finally, we introduce sparse unmixing methods for the blind scenario and evaluate the efficiency of the proposed methods via simulations over synthetic and real hyperspectral data sets. The results illustrate considerable enhancements to estimate the spectral library of materials and their fractional abundances such as smaller values of spectral angle distance (SAD) and abundance angle distance (AAD) as well.
Co-Authorship

List of publications as a result of the contributions of this thesis:


• **Y. Esmaeili Salehani**, S. Gazor, “Sparse Data Reconstruction via Adaptive
\( \ell_p \)-norm and Multilayer NMF”, accepted for publication in the 7th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEEE IEMCON 2016), Vancouver, Canada, October 2016.


- **Y. Esmaeili Salehani**, S. Gazor, “Sparse Hyperspectral Unmixing via Varying \( \ell_p \)-norm Approximation of \( \ell_0 \)-norm”, *under 2nd revision*. 
Acknowledgments

I would like to appreciate everyone who made this dissertation possible.

First, I would like to thank my supervisors and specifically Professor Saeed Gazor for his guidance and support in all academical stages paved the way for me. I thank him for his patience, knowledge and time to have wonderful discussion and ideas throughout this work. I am also grateful to my PhD committee members, Professor Soosan Beheshti, Professor Xiang Li, Professor Aboelmagd Noureldin, Professor Thomas Dean and Professor Tucker Carrington to read my thesis, to attend my defence and to give excellence comments and observations. Also, I would like to thank my course instructors, Professor Fady Alajaji for the Information theory course and Professor Ali Ghrayeb for both Coding theory and MIMO communication courses, valuable and great discussions in the classes and afterwards.

I would like to say a special thank you to my wife for her patience and unconditional supports for all the time. She helped me to concentrate on completing this dissertation and supported me faithfully during my endeavors. Nothing I can say can do justice to how I feel about your support as always, Mona.

I also wish to express my appreciation for my parents and Shahnaz. Without their belief in my goals, their kindness and affection during my life, I would not be
able to enter the field of scientific research. I would like to thank my mother-in-
law and father-in-law for their support and kindness throughout my PhD study as well. Moreover, I would like to thank my sister, brothers and brother-in-law for their all supports for this journey. Finally, I wish to thank my colleagues and friends at Queen’s University.
To:

Danesh and V-D. I
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAD</td>
<td>Abundance Angle Distance</td>
</tr>
<tr>
<td>ADMM</td>
<td>Alternating Direction Method of Multipliers</td>
</tr>
<tr>
<td>ANC</td>
<td>Abundance Nonnegativity Constraint</td>
</tr>
<tr>
<td>ASC</td>
<td>Abundance Sum-to-one Constraint</td>
</tr>
<tr>
<td>AWGN</td>
<td>Additive White Gaussian Noise</td>
</tr>
<tr>
<td>AVIRIS</td>
<td>Airbone Visible/Infrared Imaging Spectrometer</td>
</tr>
<tr>
<td>BCG</td>
<td>Basic Conjugate Gradient</td>
</tr>
<tr>
<td>BP</td>
<td>Basis Pursuit</td>
</tr>
<tr>
<td>BPDN</td>
<td>Basis Pursuit Denoising</td>
</tr>
<tr>
<td>CBP</td>
<td>Constrained Basis Pursuit</td>
</tr>
<tr>
<td>CBPDN</td>
<td>Constrained Basis Pursuit Denoising</td>
</tr>
<tr>
<td>CDF</td>
<td>Cumulative Distribution Function</td>
</tr>
<tr>
<td>CHL</td>
<td>Collaborative Hierarchical LASSO</td>
</tr>
<tr>
<td>CHSR</td>
<td>Collaborative Hierarchical Sparse Regression</td>
</tr>
<tr>
<td>CLS</td>
<td>Constrained Least Squares</td>
</tr>
<tr>
<td>CSR</td>
<td>Constrained Sparse Regression</td>
</tr>
<tr>
<td>DCT</td>
<td>Discrete Cosine Transform</td>
</tr>
<tr>
<td>EEA</td>
<td>Endmember Extraction Algorithm</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>EM</td>
<td>Expectation-Maximization</td>
</tr>
<tr>
<td>GLNMF</td>
<td>Graph Regularized $\ell_{1/2}$-Nonnegative Matrix Factorization</td>
</tr>
<tr>
<td>GSUnSAL</td>
<td>Group Sparse Unmixing via variable Splitting Augmented Lagrangian</td>
</tr>
<tr>
<td>HSI</td>
<td>Hyperspectral Image</td>
</tr>
<tr>
<td>IRLS</td>
<td>Iteratively Reweighted Least Squares</td>
</tr>
<tr>
<td>JPL</td>
<td>Jet Propulsion Laboratory</td>
</tr>
<tr>
<td>KKT</td>
<td>Karush-Kuhn-Tucker</td>
</tr>
<tr>
<td>LARS</td>
<td>Least Angle Regression</td>
</tr>
<tr>
<td>LASSO</td>
<td>Least Absolute Selection and Shrinkage Operator</td>
</tr>
<tr>
<td>LMM</td>
<td>Linear Mixing Model</td>
</tr>
<tr>
<td>MC</td>
<td>Mutual Coherence</td>
</tr>
<tr>
<td>MLNMF</td>
<td>MultiLayer Nonnegative Matrix Factorization</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean Square Error</td>
</tr>
<tr>
<td>NCCHL</td>
<td>Non-negative Constrained Collaborative Hierarchical LASSO</td>
</tr>
<tr>
<td>NCHL</td>
<td>Non-negative Constrained Hierarchical LASSO</td>
</tr>
<tr>
<td>NCLS</td>
<td>Nonnegative Constrained Least Squares</td>
</tr>
<tr>
<td>NMF</td>
<td>Nonnegative Matrix Factorization</td>
</tr>
<tr>
<td>OMP</td>
<td>Orthogonal Matching Pursuit</td>
</tr>
<tr>
<td>PNMM</td>
<td>Postnonlinear Mixing Model</td>
</tr>
<tr>
<td>PoS</td>
<td>Probability of Success</td>
</tr>
<tr>
<td>PPNM</td>
<td>Polynomial Postnonlinear Model</td>
</tr>
<tr>
<td>PPI</td>
<td>Pixel Purity Index</td>
</tr>
<tr>
<td>RSNR</td>
<td>Reconstruction Signal to Noise Ratio</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>SAD</td>
<td>Spectral Angle Distance</td>
</tr>
<tr>
<td>SDSNMF</td>
<td>Substance Dependence Constrained Sparse Nonnegative Matrix Factorization</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal to Noise Ratio</td>
</tr>
<tr>
<td>SSoM</td>
<td>Spectral Signatures of Materials</td>
</tr>
<tr>
<td>SU</td>
<td>Spectral Unmixing</td>
</tr>
<tr>
<td>SUnSAL</td>
<td>Sparse Unmixing via variable Splitting Augmented Lagrangian</td>
</tr>
<tr>
<td>USGS</td>
<td>United States Geological Survey</td>
</tr>
<tr>
<td>VCA</td>
<td>Vertex Component Analysis</td>
</tr>
<tr>
<td>VNA</td>
<td>Varying $\ell_p$-Norm Approximation</td>
</tr>
</tbody>
</table>
# Contents

Abstract i  
Co-Authorship iii  
Acknowledgments v  
Contents vii  
List of Tables xiv  
List of Figures xvi  

## Chapter 1: Introduction 1  
1.1 Motivation ................................................. 2  
1.2 Thesis Organization and Contributions .................. 6  

## Chapter 2: Background 9  
2.1 Hyperspectral Imaging: A general background ............ 9  
2.2 Hyperspectral Unmixing: An overview .................... 12  
2.3 System Models and Linear Sparse Hyperspectral Unmixing ..... 16  
  2.3.1 Dictionary-aided semiblind sparse unmixing methods .... 18  
  2.3.2 Non-dictionary-aided blind sparse unmixing methods ..... 21  

## Chapter 3: Semiblind Sparse Hyperspectral Unmixing Via $\ell_p$-norm 24  
3.1 Adaptive LASSO Hyperspectral Unmixing Using ADMM .......... 24  
  3.1.1 The problem formulation .................................. 25  
  3.1.2 Our proposed unmixing algorithms ..................... 26  
  3.1.3 Simulation results ....................................... 29  
  3.1.4 Conclusion ............................................... 32
### Chapter 3: Sparse Hyperspectral Unmixing via Varying $\ell_p$-norm Approximation

- 3.2 Sparse Hyperspectral Unmixing via Varying $\ell_p$-norm Approximation of $\ell_0$-norm
  - 3.2.1 $\ell_p$ regularized problem with variable $p$  
  - 3.2.2 Updating rules for pair $(p(t), \epsilon(t))$
  - 3.2.3 Proposed methods for solving (3.15) and (3.16)
  - 3.2.4 The convergence analysis
  - 3.2.5 Experimental results and analysis
  - 3.2.6 Conclusion

### Chapter 4: Semiblind Sparse Hyperspectral Unmixing Through $\ell_0$-norm Approximation

- 4.1 Sparse Hyperspectral Unmixing via arctan Approximation of $\ell_0$-norm
  - 4.1.1 The proposed method
  - 4.1.2 Simulation results
  - 4.1.3 Conclusion
- 4.2 $\ell_0$-Norm Sparse Hyperspectral Unmixing Using Arctan Smoothing
  - 4.2.1 Our proposed unmixing method: Arctan approximation of the $\ell_1$- and $\ell_0$-norms
  - 4.2.2 Updating the regularized parameter $\lambda$
  - 4.2.3 Experimental results and analysis
  - 4.2.4 Conclusion

### Chapter 5: Blind Sparse Hyperspectral Unmixing Through Non-negative Matrix Factorization

- 5.1 Collaborative Unmixng Hyperspectral Imagery via NMF
  - 5.1.1 Proposed collaborative NMF-based unmixing method
  - 5.1.2 Experimental results
  - 5.1.3 Conclusion
- 5.2 Multilayer NMF Sparse Hyperspectral Unmixing via Adaptive $\ell_p$-norm
  - 5.2.1 Problem formulation and proposed methods
  - 5.2.2 Experiment over real data
  - 5.2.3 Conclusion
- 5.3 Smooth and Sparse Regularization for NMF Hyperspectral Unmixing
  - 5.3.1 Arctan-NMF: our proposed method for unmixing
  - 5.3.2 Experimental results
  - 5.3.3 Computational complexity
  - 5.3.4 Conclusion

### Chapter 6: Conclusions

- 6.1 Summary and Concluding Remarks
- 6.2 Future Work

xii
Bibliography  

Appendix A: Proof for Chapter 3  
A.1 Proof of Theorem 1  

Appendix B: Proof for Chapter 4  
B.1 Proof of Theorem 2  

Appendix C: Proof for Chapter 5  
C.1 Proof of Proposition 2
List of Tables

3.1 The comparison results of RSNR values for ALwADMM1, ALwADMM2 and the proposed method of [1] ......................... 32
3.2 Parameter setting for our proposed methods (VNA\textsubscript{i}; i = 1, \ldots, 6) and SUnSAL tuned via hit and trial for different sparsity levels at SNR = 25dB. ........................................ 50
3.3 Parameter setting for SUnSAL algorithm, CZ method and our proposed methods (VNA\textsubscript{i}(i = 1, \ldots, 6)) for different SNRs of white Gaussian noise and colored noise (sparsity level = 5). .................... 55
3.4 Comparison of the measured processing time per pixel (in Seconds) for different unmixing algorithms over three different libraries on an Intel Core i7-2600 (at 3.5 GHz) and 8 GB of RAM Memory (SNR = 40 dB, Sparsity level = 5). .................. 56
4.1 The comparison results of RSNR values for different methods. .......... 64
4.2 The processing time of different algorithms per pixel (in seconds) using 4 different libraries and an i7-2600-3.5-GHz Intel Core processor with 8 GB of RAM memory. ............................... 86
5.1 Comparison of the SAD results of the various methods on the Samson hyperspectral data set ................................. 105
5.2 Comparison of the SAD results of the various methods on the Samson hyperspectral data set. . . . . . . . . . . . . . . . . . . . . . . . . . . 126

5.3 Comparison of the SAD results of the various unmixing methods on the AVIRIS cuprite data . . . . . . . . . . . . . . . . . . . . . . . . 129

5.4 Number of operations of the Arctan-NMF in terms of required flops in each iteration \((2N < L < P)\). . . . . . . . . . . . . . . . 131
List of Figures

2.1 The JPL’s AVIRIS hyperspectral data cube provided on a NASA ER-2 plane over Moffett Field [2]. .................................................. 10

2.2 Hyperspectral imaging concept: the modified version of the JPL’s AVIRIS [2]. ................................................................. 11

2.3 A schematic of the linear mixing model. ................................. 13

2.4 A schematic of the non-linear mixing model: intimate media (right side) and bilinear model (left side). ................................. 15

3.1 The results of reconstruction fractions of endmembers for ALwADM- M1, ALwADMM2, adaptive lasso using LARS (alasso) [1] and lasso via SUnSAL [3], the index of nonzero elements are shown. ........ 30

3.2 The mean square errors of the estimated fractional abundances versus SNRs for ALwADMM1, ALwADMM2, least square method [4] and alasso [1]. ................................................................. 31

3.3 The variation of the cost function through different steps and iterations of the proposed methods, VNA1 to VNA6 for AWGN at SNR = 30dB, sparsity level = 5: (a) The cost function (3.25) calculated for 5 steps of Algorithm 1 in each iteration, (b) The cost function (3.27) calculated for 3 steps of Algorithm 2 in each iteration. ..................... 45
3.4 Estimated entries of \( \mathbf{x} \) with 3 non-zero elements using proposed methods in AWGN at SNR = 30dB. ......................... 46
3.5 The comparison of RNSR values between our proposed algorithms and the other state-of-the-art methods with respect to different sparsity levels \( \| \mathbf{x} \|_0 \) for different libraries \(( \mathbf{M}_1, \cdots, \mathbf{M}_6 )\). ......................... 51
3.6 The comparison of PoS values between our proposed algorithms and the other state-of-the-art methods with respect to different sparsity levels \( \| \mathbf{x} \|_0 \) for different libraries \(( \mathbf{M}_1, \cdots, \mathbf{M}_6 )\). ......................... 52
3.7 The RSNR values obtained by different sparse unmixing methods versus SNRs for the simulated data with white- and correlated- noise over the USGS spectral library \( \mathbf{M}_5 \) and \( \mathbf{M}_6 \) for the sparsity level of 5. . 54
3.8 USGS map showing the location of different minerals in the Cuprite mining district in Nevada. The map is available online at: http://speclab.cr.usgs.gov/cuprite95.tgiff.2.2um_map.gif. ...... 57
3.9 Bands 5 (a) and 40 (b) of the subimage of AVIRIS cuprite Nevada dataset. .............................. 58
3.10 Fractional abundances estimated from different algorithms for five end-members appeared in the scene of AVIRIS Cuprite hyperspectral data. 59
4.1 The original data pixels and the corresponding estimated maps by different methods for three SNRs. ............................................. 65
4.2 Comparison of arctan(\( \sigma \mathbf{x} \))/arctan(\( \sigma \)) with \( x^p \) for different values of \( \sigma \) and \( p \). ............................. 70
4.3 The comparison of RSNR values and their corresponding probability of success (PoS) between our proposed methods and the other state-of-the-art methods with respect to different sparsity levels using $M_{\text{Prune}}$. (a) SNR = 15 dB; (b) SNR = 30 dB; (c) SNR = 50 dB.

4.4 The RSNR values obtained by different sparse unmixing methods versus SNRs for the simulated data. (a) Sparsity level = 1; (b) sparsity level = 3; (c) sparsity level = 5.

4.5 The impact of library properties and colored noise over (a) RSNR and (b) PoS values obtained by different sparse unmixing methods when $||x||_0 = 4$ and SNR = 30dB.

4.6 Estimated abundance fraction maps for the subimage of AVIRIS cuprite using different unmixing methods.

4.7 The estimated CDF of the fractional abundances of different methods over AVIRIS cuprite mining in Nevada.

5.1 Comparison of the estimated spectral signatures of four sample materials with the corresponding ground-truth values (SNR = 25 dB) (a) Almandine (b) Bronzite (c) Biotite (d) Chlorite

5.2 Simulation results of unmixing methods of (a) SAD (b) AAD as functions of SNRs.

5.3 A visualization of the abundances estimated from different methods for three components of Samson hyperspectral data set: #1 Soil, #2 Tree #3 Water.

5.4 Signatures of 5 sample materials from USGS library (splib06): Almandine, Beryl, Carnallite, Meionite, and Strontianite.
5.5 The squared of the DCT of the spectral signatures of the mentioned materials in Figure 5.4. ................................. 110
5.6 The squared of the DCT of all materials signatures averaged over USGS library (splib06). ........................................ 110
5.7 The metric values as a function of $\omega$ for different data sets and SNRs. 117
5.8 The metric values as a function of iteration numbers for different data sets and SNRs. ........................................ 118
5.9 The metric values as a function of $\beta$ for different data sets and SNRs. 119
5.10 The metric values as a function of maximum number of endmembers and degree of purities contributed at each pixel at SNR = 20 dB. . . 120
5.11 The metric values as a function of purities for $E_m = 5$ at SNR = 20 dB for different unmixing methods. ................................. 121
5.12 The metric values as a function of the maximum number of endmembers contributed at each mixed pixel for $n_T = 0.8$ at SNR = 20 dB for different unmixing methods. ................................. 122
5.13 The metric values as a function of the noise power (SNR in dB) for different unmixing methods over two different data sets. .............. 123
5.14 The metric values as a function of the number of pixels in the scene for $mE = 6$ and $aT = 0.7$ at SNR = 20 dB. ................................. 124
5.15 The impact of the initialization of the unmixing method over different metrics for two data sets DS1 and DS2. ................................. 125
5.16 A visualization of fractional abundance maps of endmembers estimated by different unmixing methods over Samson hyperspectral image: a) Arctan-NMF ($\xi = 4, \delta = 0.1$) b) $\ell_2$-NMF [5] ($\delta = 15$) c) $\ell_1$-NMF [6] ($\delta = 15$) d) CNMF [7] ($\delta = 15$) e) MLNMF [8] ($\delta = 25$) f) VCA-FCLS ($\delta = 15$).

5.17 Cuprite AVIRIS results: Comparison of the (black and solid lines) original spectral signatures with the (red and dotted lines) endmember signatures extracted by our Arctan-NMF and their corresponding abundance maps of endmembers.
Chapter 1

Introduction

Hyperspectral image (HSI) data enable us to identify materials through their unique spectral signatures as *fingerprints*. These characteristics have been developed by scientists in the laboratory as *imaging spectroscopy*. The concept of hyperspectral remote sensing, known as the imaging spectroscopy, has been initiated in the 1980’s and its technology has been used for various applications later on. Theses applications may include to find new oil fields in geology, to detect defective pills in pharmaceutical, to exploit minerals in mineralogy, to identify invisible defects in food quality inspection, to detect and analyze tumors in medical imaging and to monitor the growth and health of crops in agriculture [9, 10]. Hyperspectral remote sensors involve sensing the objects in hundreds of spectral bands of relatively narrow bandwidths (5-10 nm). These sensors capture the electromagnetic energy emitted from materials and collect HSIs as a data cube with two-dimensional spatial information over many contiguous bands of high spectral resolution [9, 10]. Due to insufficient spatial resolution of sensors and mixing of the materials in a microscopic scale, the mixture of spectral signatures of materials (SSoM) often occurs in practice [9, 10, 11]. Thus, spectral
unmixing (SU) is necessary to determine which materials are present in the measured pixels. SU methods have been an active research area in recent years since they meet various challenges such as modeling types and sparse problems [9, 10, 12], and references therein.

1.1 Motivation

Since the number of endmembers appeared in each mixed pixel is scanty in comparison with the dimension of available spectral libraries in practice, the spectral unmixing can be considered as a sparse regression problem. Indeed, as the library size increases beyond the number of existing materials in each mixed pixel, a number of elements of $\mathbf{x}$ becomes zero and the problem is referred to as sparse unmixing, e.g. see [13, 14, 15, 16, 17, 18, 19, 20, 21, 22]. Hence, the measured hyperspectral vectors can be approximated with a linear combination of a small number of spectral signatures in the given library as regressors. In this case, we assume that each mixed pixel contains significantly smaller number of materials than the number of the selected endmembers, i.e., the number of non-zero elements of $\mathbf{x}$ is much smaller than its length. This problem is referred as the $\ell_0$-norm problem which is NP-hard problem [23, 24, 25].

Although different unmixing techniques have been proposed through the $\ell_1$-norm approach, e.g. [1, 14, 20, 17, 3], several researchers have investigated different replacements for the $\ell_0$-norm term appeared in the objective function in order to improve the sparsity. The iterative reweighted methods such as iteratively reweighted least squares (IRLS) algorithm [26] and reweighted $\ell_1$ minimization problem [27], the sparse representation through the greedy algorithms [28, 29], the $\ell_p$-norm problem as
1.1. MOTIVATION

a weighted $\ell_2$-norm based on IRLS [30] as well as smoothes $\ell_0$-norm minimization [31, 32, 18, 33, 34, 22, 35] are such problems to solve the optimization problems dealing with the hyperspectral unmixing procedure.

Generally, the approximation of the $\ell_0$-norm by the $\ell_p$-norm term for fixed values of $p$ has been studied for various problems with different constraints (e.g. [36, 31, 37, 38]). Furthermore, different algorithms have been proposed to solve the corresponding $\ell_p$-norm problem in order to give more accurate solutions for the desired $\ell_0$-norm problem. For instance, the sparse $\ell_p - \ell_2$ minimization problem based on the relaxing the $\ell_0$-norm to the $\ell_p$-norm for constant values of norm $p$ is proposed in [30]. Moreover, some IRLS algorithms that are based on $\ell_p$-norm minimization [39, 40, 38] and the smoothed approximate $\ell_p$-norm employing a basic conjugate gradient (BCG) method [41] are more examples to approximate the $\ell_0$-norm problem. However, the value of norm $p$ in these proposed methods except [41] (i.e. it was proposed recently for compressive sensing and we show that their method are not suitable for the SU in this paper by simulation results) is fixed during the process of minimization problem. Thus, we desire to propose an algorithm for the SU by minimizing the number of nonzero elements of the solution through $\ell_p$-norm while $p$ is iteratively decreasing. Indeed by starting $p$ at 1, the problem is convex and a unique solution is possible under some conditions [38, 42]. This is motivated by our theorem that the set of minima for the $\ell_p$-norm problem is continuous in terms of $p$, which implies that smooth iterative reduction of $p$ results in an enhanced solution. To introduce our proposed method, a weighted $\ell_1$-norm approximation of the $\ell_p$-norm problem is studied that a parameter $\epsilon$ is involved to deal with the fact that the $\ell_p$-norm problem is not Lipschitz continuous for $p < 1$. 
Although the methods using the $\ell_p$-norm ($0 < p < 1$) may lead to the sparser solutions, they have been suffered from numerical problems since for $p < 1$, the $\ell_p$-norm is not a Lipschitz continuous function. In fact, these methods must introduce an extra parameter to make it Lipschitz continuous, which leads to more approximations. Thus, using smooth functions such as logarithm-based approximation function [18], gaussian function [31, 32] and arctan function [33] for approximating the $\ell_0$-norm are recent interests to apply them for the sparse unmixing schemes. The smoothness of the approximated problems leads to converge to acceptable solutions in addition to other advantages such as ones mentioned in [18, 31, 32, 33, 22, 34, 35]. Hence, we desire to introduce an iterative method through smooth functions which starts as an $\ell_1$-norm optimization which is convex and has a unique solution, and converges fast and iteratively tends to be $\ell_0$-norm problem.

In all of the above-mentioned methods, the main goal is to find the fractional abundances of the materials for a given spectral library. Indeed, the endmembers may be extracted using algorithms such as the vertex component analysis (V-CA) method [43], automatic morphological endmember extraction (AMEE) algorithm [44], the N-FINDR algorithm [45], the pixel purity index (PPI) [46] or their variants [47, 48, 49]. In spite of such two stages-based unmixing methods that employ both the endmember extraction methods (e.g., N-FINDR [45] and VCA [43]) and mixed pixels decomposition methods (i.e., $\ell_2$-norm, $\ell_1$-norm, and $\ell_0$-norm approximation, e.g., [33, 50, 22]), one-step methods such as nonnegative matrix factorization (NMF) [51, 52] based approach is highly interested because of its noticeable supports. First of all, the nonnegativity constraints for both spectral signatures and their fractional abundances are automatically included in the NMF-based methods. Furthermore,
1.1. MOTIVATION

it can make decomposition matrices to be more intractable because of a part-based representation of the data, see [53, 5] and references therein. Furthermore, various algorithms have recently been proposed to exploit the sparse property of abundances and to enhance the classical NMF method for a blind unmixing scenario. By introducing $\ell_{1/2}$-norm in [5] significant enhancement in sparsity is obtained. Manifold regularization into the sparsity constraint [54], multilayer NMF decomposition [8] and substance dependence constrained sparse NMF [55] are some other unmixing methods that investigated $\ell_{1/2}$-NMF problem and more properties of the hyperspectral images. The graph-regularized $\ell_{1/2}$-NMF (GLNMF) method in [54] uses the local structural information of hyperspectral images with a nearest neighbor graph as the internal manifold structure. The multilayer NMF method (MLNMF) [8] applied the sparseness constraints for both endmember and fractional abundances matrices for each layer. The substance dependence constrained sparse NMF (SDSNMF) method [55] uses the spectral information as well as the substance dependence of hyperspectral data through the $k$-nearest neighbor of the pixels. Indeed, the SDSNMF method utilizes the sparsity property of pixels by representing a weighting matrix and considers the similarities of the abundances for the scene. Although all of these methods obtained satisfactory results, they prone to high level of noise and may exhibit numerical instability due to non-differentiability of the $\ell_{1/2}$-norm term. Inspired by the above motivation, we will study some intrinsic properties of the sparsity of the observed hyperspectral pixels such as simultaneous (collaborative) sparse regression model and the smooth characteristic of the spectral signatures of materials in order to investigate the impacts of such features over SU through NMF problem.
1.2 Thesis Organization and Contributions

This thesis comprises three main chapters. In Chapter 3, we propose two linear sparse SU schemes through a dictionary-aided approach. We introduce a weighted $\ell_p$-norm method to approximate the original $\ell_0$-norm problem defined by equation (2.6). Our first scheme solves the problem (2.7) using the Alternative Direction Method of Multipliers (ADMM) algorithm. Indeed, we consider the unmixing problem as an adaptive version of Least Absolute Selection and Shrinkage Operator (LASSO) problem. In the second part of Chapter 3, we generalize our initial scheme and propose to use the $\ell_p$-norm for $p < 1$ as a better approximation for the $\ell_0$-norm. In this case, our proposed method is to iteratively reduce $p$ from 1 to 0 in order to take advantage of the unique optimal solution for $p = 1$ and then track the optimal solution for $p < 1$ as $p$ is reducing. This scheme is supported by our theorem that the set of minima for the weighted $\ell_p$-norm problem is continuous in terms of $p$, which implies that smooth iterative reduction of $p$ results in an enhanced solution. Thus, smaller values of $p$ result in better approximation. However, they result in an increase in the number of local optima which either trap the algorithms in a suboptimal solution or translate into increased computational complexity.

Then, we investigate arctan function as a smooth function in Chapter 4, which is Lipschitz continuous function, to approximate the $\ell_0$-norm term to tackle these kinds of numerical problems. In the first part of Chapter 4, we utilize the arctan function with a constant parameter, i.e. $\sigma$, to adjust the sparsity term. We use the ADMM approach to solve the acquired minimization problem and propose our unmixing method. In the next part, we improve our previous main idea of the adaptive weighted $\ell_p$-norm to introduce new unmixing scheme using arctan function with an
adjustable parameter called $\sigma$. We show that the set of local optima of the objective function is a continuous set versus $\sigma$. Thus, it confirms that the gradual increase of $\sigma$ along with an iterative minimization of the objective function leads to the optimal solution. Also, we study the importance of the Lagrangian multiplier on the performance of our proposed methods. We evaluate the efficiency of the proposed methods through two common updating rules for the Lagrangian parameter in addition to the fixed values of the Lagrangian parameter selected by hit and trial. We evaluate our proposed methods through different experiments over both synthetic and real hyperspectral data and the results reveal the superiorities and shortcomings of the proposed methods in comparison with several state-of-the-art methods.

In Chapter 5, we study a blind scenario of unmixing methods for LMM where the library of SSoM must be estimated. First, we consider the collaborative property of the hyperspectral pixels through the NMF problem. We use the fact that all pixels share the same set of SSoM lying into a lower dimensional subspace and the fractional abundances of endmembers can be effected by imposing sparsity among the endmembers collaboratively. Hence, we formulate the NMF problem through the $\ell_{2,q}$-norm term of the fractional abundances of all pixels. We employ the multiplicative updating rules to solve the acquired minimization problem for the $\ell_{2,q}$-norm based NMF approach and show the efficiency of the proposed methods through simulations. Afterwards, we introduce a new method of unmixing algorithm using adaptive $\ell_p$-norm term and multilayer NMF method [56]. We propose two cost functions where the first one only considers the Frobenius norm of error and the $\ell_p$-norm of fractional abundances and the second one considers an additional term of $\|A\|_F$ to bound the errors of the estimated dictionary matrix. Then, we introduce our proposed methods
based on solving two mentioned cost functions and examine the efficiency of both algorithms through real hyperspectral data set. In the last part of Chapter 5, we study a new method of blind sparse hyperspectral unmixing by approximating of $\ell_0$-norm through arctan function and adding a weighted $\ell_2$-norm of the spectral library represent the smooth property of SSoM [57]. We exemplify that the first few subbands contain most of the energy of SSoM which result in smooth spectral signatures. This intrinsic property of spectral signatures leads to enhance the results and we show it through the simulations over both synthetic and real hyperspectral data.

Finally, Chapter 6 concludes the thesis and outlines future works.
Chapter 2

Background

2.1 Hyperspectral Imaging: A general background

Remotely sensed data contain four possible types of resolution depending on spectral sensors: panchromatic, multispectral, hyperspectral and ultraspectral [58, 59]. The panchromatic sensor contains only one spectral band and has the lowest resolution. Multispectral sensors sample the electromagnetic spectrum in tens of spectral bands and give the higher spectral resolution compared with the panchromatic sensors. However, hyperspectral and ultraspectral sensors enable to record the SSoM with higher resolution of spectra than the first two types of sensors [59]. Hyperspectral remote sensing technology concerns to capture images with a range of spectra from the visible sections of 0.4-0.7 $\mu$m to near-infrared of 2.4 $\mu$m in several hundreds of narrow bands in the range of 10 nm wide [9, 10]. There are two kinds of hyperspectral remote sensors: spaceborne (e.g., Hyperion) and airborne (e.g., Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) and Compact Airborne Spectrographic Imager (CASI)) [58].

The hyperspectral imaging sensors provide a three dimensional data structure
2.1. HYPERSPECTRAL IMAGING: A GENERAL BACKGROUND

Figure 2.1: The JPL’s AVIRIS hyperspectral data cube provided on a NASA ER-2 plane over Moffett Field [2].

called data cube with the spatial dimensions in the face and the spectral dimension in the depth, respectively. Figure 2.1 shows the image of cube from the Jet Propulsion Laboratory (JPL’s) AVIRIS over Moffett Field, California [2]. The false-color image on the top of the cube in Figure 2.1 has been used to show a detailed structure in the water and evaporation ponds on the right. The Moffett Field airport is visible on the top of the cube as well.

The basic principle of a typical HSI is shown in Figure 2.2. Hyperspectral camera sensors measure the spectrum of the light reflected at each pixel. Then, the sampled spectrum is used to identify the material/materials present in the pixel by their reflectance.

HSI process concerns with the analysis of information from the electromagnetic energy emitted from materials. There are various hyperspectral data analysis methods for the hyperspectral imaging. The following lists briefly describe the main and active area of research data analysis for HSI processing [9, 59, 13, 60, 61]:
2.1. HYPERSPECTRAL IMAGING: A GENERAL BACKGROUND

Figure 2.2: Hyperspectral imaging concept: the modified version of the JPL's AVIRIS [2].

- **Target detection** is to search the pixels of a HSI for the presence of a specific spectral signature as a target [59, 13, 10].

- **Classification** is to assign all pixels of a HSI into a specific class as a label [9].

- **Change detection** is the process of identifying what has changed on the ground by comparing the results of HSI data with each other, either spatially or non-spatially [58, 13].

- **Dimensionality reduction** is the process of reducing the number of dimensions of a HSI to assist the progress of succeeding tasks [60].

- **Feature extraction** is a form of dimensionality reduction to transform a high
2.2. HYPERSPECTRAL UNMIXING: AN OVERVIEW

resolution HSI into a set of features [60, 61].

• *Segmentation* is a method of partitioning a HSI into disjoint regions such that they correspond to objects in the HSI [61].

• *Spectral unmixing* is a process of estimating the number of reference materials called spectral signatures as well as their fractional abundances [13, 60].

2.2 Hyperspectral Unmixing: An overview

The observed pixels of HSIs collected by the hyperspectral sensors are labeled as either *pure* or *mixed* pixels. A pure pixel contains only a single component originating from the constituent substance called endmember. However, in a mixed pixel, there is a combination of endmembers with different proportions/fractions called abundances. In a mixed pixel, the estimated SSoM does not correspond to a known unique feature of material. Thus, finding which materials are present in the mixed pixels is challenging and hyperspectral unmixing methods are necessary to analyze such pixels of HSIs. Keshava introduced three major phases for the typical end-to-end spectral unmixing [60] as follows

• *Dimensionality reduction* is the process of reducing the dimensionality of the measured hyperspectral data cube in order to facilitate the computational complexity of the processing issue.

• *Endmember determination* consists of estimating the constituent spectra regards to the mixed pixels in the scene.

• *Inversion phase* estimates the fractional abundances of the endmembers contributed to each mixed pixel.
Following the mixed pixels structure, we consider two types of mixing models: linear mixing model (LMM) and non-linear mixing model (NLMM). The LMM is widely used for the macroscopic mixing scale when only one constituent material interferes the incident light. Hence, linear spectral unmixing (LSU) methods are the most common technique in remotely sensed hyperspectral data interpretation. In a typical LMM, the measured spectral vector at each pixel is a linear combination of the SSoM and their fractional abundances. Figure 2.3 shows a schematic of a typical LMM.

LMM is not fitted for some scenarios such as the one that the source radiation is assumed to be scattered several times before being recorded at the sensors [62, 63]. Thus, another model is needed to describe the mixed spectra for such multiple scattering emitted from different materials. NLMM is applicable in urban scenes [64],
vegetated areas [65] and those containing specific spectral signatures such as soil, sand and trees [63, 66]. Figure 2.4 shows two types of NLMM. The intimate mixture was initially introduced by Hapke as the bidirectional reflectance spectroscopy [62]. It considers the situation that the photons are interacting with all materials of the object simultaneously while the pure spectral components are not be able to be distinguished spatially (Figure 2.4, left). Recently, more kinds of nonlinear models have been considered such as radial basis function networks [67, 68], kernel-based models [69, 70, 71, 72, 73], post-nonlinear mixing model (PNMM) [74] and bilinear mixing model [65, 74, 75, 76, 77]. The latter model has been proposed for the specific kind of nonlinearity that multiple lights are scattered in a two-layer media. It considers multiple scattering interactions by adding the interaction terms into LMM as shown in Figure 2.4 (right side).

SU methods can be categorized as fully supervised, semi-supervised and non-supervised schemes depending on the available knowledge of the parameters involving for the unmixing model. In a general form of supervised SU method, the knowledge of parameters contributing to the mixing model is given a priori. This may contain the knowledge of the endmember’s spectral signatures of the measured pixels. For instance, the method of supervised nonlinear spectral unmixing [74], the endmembers contributed to the data cube are estimated by an endmember extraction algorithm firstly, and then, the abundances and the nonlinearity coefficients for all pixels of the image will be estimated. In an unsupervised SU, the endmembers’ spectra are assumed to be unknown and must be estimated directly from the hyperspectral data e.g., [43, 45]. For example, the proposed unsupervised SU method for NLMM [63], the endmember spectra, the kernel parameters and the fractional abundances shall
be estimated from the measured HSI data, while only the number of endmembers is supposed to be known. Finally in a semi-supervised SU method, some information are assumed to be given a priori (e.g., the library of spectral signatures of materials) [78]. For instance, in the proposed method of semi-supervised unmixing for LMM [1], the spectral library is supposed to be known and the main objective is to find the endmembers (number and kind) contributing to the hyperspectral scene as well as their corresponding abundances.
2.3. SYSTEM MODELS AND LINEAR SPARSE HYPERSPECTRAL UNMIXING

2.3 System Models and Linear Sparse Hyperspectral Unmixing

Although the above-mentioned three classifications in Section 2.2 are considered in some literature, they are not easy to be distinguished in some cases because of the type of available knowledge as well as the mixing model used for such scenarios. Moreover, in LMM which is the fundamental mixing model to propose our unmixing methods in the thesis, the major availability of knowledge is whether the spectral library is given a priori or not. Hence, we consider another categorization for LMM which is more appropriate to classify the proposed methods as follows:

- **Dictionary-aided semiblind unmixing methods** that the spectral library is assumed to be known and the number of endmembers and the fractional abundances shall be estimated.

- **Blind unmixing methods** that the number of endmembers is assumed to be known and the goal is to find the library of spectral signatures and the fractional abundances.

As mentioned earlier, LMM is applicable for many scenarios that the incident light interacts with only one material as shown in Figure 2.3 and it is a very common model in literature as well, e.g. [9, 13, 60, 61]. In this model, the measured spectra of pixels for a HSI is considered as a linear combination of the spectral signatures scattered from the materials and their fractional abundances. The measured reflectance values can be presented by the $L$-dimensional vector $\mathbf{y} = (y_1, \ldots, y_i, \ldots, y_L)^T$ where $L$ is the number of spectral bands and $(^T)$ is the transpose of a matrix. The observed
spectral signatures of each pixel at the $i$-th spectral band, $y_i$, is given by

$$y_i = \sum_{j=1}^{N} m_{ij} x_j + e_i.$$  \hspace{1cm} (2.1)

where $m_{ij}$ is the reflectance value of the $j$th endmember, $x_j$ is the corresponding abundance fractions, $e_i$ is the additive noise term for the spectral band $i$ and $N$ is the number of endmembers. The compact matrix form of the measured spectrum of each pixel of data cube is expressed as follows

$$y = Mx + e.$$ \hspace{1cm} (2.2)

where, $y$ is an $L \times 1$ column vector, $M = [m_1, m_2, ..., m_N] \in \mathbb{R}^{L \times N}$ is the mixing matrix containing $N$ pure spectral signatures, $x$ is a $N \times 1$ vector of the corresponding fractional abundances of materials and $e$ is a $L \times 1$ additive noise vector. There are two physically meaningful constraints for the fractional abundances as follows

- abundance nonnegativity constraint (ANC)

$$0 \leq x_i \leq 1, \hspace{1cm} i = 1, 2, ..., N.$$ \hspace{1cm} (2.3)

- abundance sum-to-one constraint (ASC)

$$1^T x = \sum_{i=1}^{N} x_i = 1.$$ \hspace{1cm} (2.4)

where $1^T$ is a $N$-dimensional vector of ones.
2.3. SYSTEM MODELS AND LINEAR SPARSE HYPERSPECTRAL UNMIXING

By representing the mentioned constraints to the \( N - 1 \) standard simplex in the \( N \)-dimensional space of \( S \) in geometrical terms, e.g. [1], the abundance vector \( \mathbf{x} \) can be written as

\[
\mathbf{x} \in S, \quad S = \{ x \in \mathbb{R}^n | \sum_{i=1}^{n} x_i = 1, x_i \geq 0, \forall i \}. \tag{2.5}
\]

2.3.1 Dictionary-aided semiblind sparse unmixing methods

For a dictionary-aided semiblind sparse unmixing method, the fractional abundances can be obtained by solving the following minimization problem

\[
\min_{\mathbf{x} \in S} ||\mathbf{x}||_0 \quad \text{subject to} \quad ||\mathbf{y} - \mathbf{Mx}||_2^2 \leq \varepsilon, \tag{2.6}
\]

where \( ||\mathbf{x}||_0 \) is the number of nonzero \( x_i \)'s, \( \varepsilon \) is a small known positive value represents the quadratic error and \( S \) is the feasible polytope set in (2.5) defined by nonnegativity (ANC) and sum-to-one (ASC) constraints.

Finding the optimal solution of equation (2.6) is a NP-hard problem [23, 24, 25], i.e., one may minimize \( ||\mathbf{x}||_0 \) by calculating the minimum of \( ||\mathbf{y} - \mathbf{Mx}||_2^2 \) over \( \mathbf{x} \) for all \( 2^N \) possible combinations of constraints for different elements of \( \mathbf{x} \) to be either zero or free and then choosing the solution for which \( ||\mathbf{x}||_0 \) is minimum while \( ||\mathbf{y} - \mathbf{Mx}||_2^2 \leq \varepsilon \) is satisfied. To avoid such an exhaustive search optimization and to reduce the complexity, several approaches are proposed by approximating the optimization problem. As a remedy, several efficient linear sparse techniques are proposed for the unmixing process [9, 12, 14, 16, 3]. Minimizing the \( \ell_1 \)-norm as approximation instead of the \( \ell_0 \)-norm is one of the earliest methods proposed to avoid such exhaustive search for equation (2.6) (e.g., see [79, 80] and the references therein; see also [14, 20, 15, 3, 50, 1, 81]
for unmixing techniques), as follows

$$\min_{x \in S} \|Wx\|_1 \text{ subject to } \|y - Mx\|_2^2 \leq \varepsilon,$$  \hspace{1cm} (2.7)

where $\|Wx\|_1 = \sum_{i=1}^{N} w_i|x_i|$ is a weighted $\ell_1$-norm of $x$, $W$ is a diagonal matrix and $w_i$'s are its diagonal entries. In [14, 20, 15, 3], the above problem is considered using $W = I$ which results in the following pure (i.e., all weights equal to 1) $\ell_1$-norm problem

$$\min_{x \geq 0} \|x\|_1 \text{ subject to } \|y - Mx\|_2^2 \leq \varepsilon.$$ \hspace{1cm} (2.8)

where $\|x\|_1 = \sum_{i=1}^{n} |x_i|$. The equivalent Lagrangian minimization of (2.8) can be considered as follows

$$\min_{x \geq 0} \frac{1}{2}\|y - Mx\|_2^2 + \lambda\|x\|_1,$$ \hspace{1cm} (2.9)

where $\lambda$ is the Lagrange multiplier. The *sparse unmixing algorithm via variable splitting and augmented Lagrangian* (SUnSAL) algorithm has been proposed in [3] to solve such $\ell_1$-norm problem for the unmixing purpose as well as its variants [14, 13].

Although there are different unmixing techniques mainly based on the $\ell_1$-norm penalization [1, 14, 20, 17, 3], several researchers have investigated different replacements for the $\ell_0$-norm term appeared in the objective function in order to improve the sparsity. The iterative reweighted methods such as *iteratively reweighted least squares* (IRLS) algorithm [26] as well as reweighted $\ell_1$ minimization problem [27], the sparse representation through the greedy algorithms [28, 29] and the $\ell_p$-norm problem as
a weighted $\ell_2$-norm based on IRLS [30] are some methods to solve the optimization problems dealing with the hyperspectral unmixing. The existing methods using the $\ell_p$-norm ($0 < p < 1$) have been suffered from numerical problems since for $p < 1$, the $\ell_p$-norm is not a Lipschitz continuous function. In fact, these methods must introduce an extra parameter to make it Lipschitz continuous, which leads to more approximations. Thus, using smooth functions such as logarithm-based approximation function [18], gaussian function [31, 32] and arctan function [33] for approximating the $\ell_0$-norm are recent interests to apply them for the sparse unmixing schemes. The smoothness of the approximated problems converges to acceptable solutions as well as other advantages such as ones mentioned in [18, 31, 32, 33, 22, 34, 35].

**The ADMM Algorithm: An overview**

The basic idea of the ADMM approach is to decompose a harder problem into a set of simpler ones. ADMM algorithm can be used to solve different versions of *constrained sparse regression* (CSR) and lasso problems, e.g. [82, 3]. This method enables to solve a set of problems as the following form

$$\min f_1(x) + f_2(z) \text{ subject to } Ax + Bz = c. \quad (2.10)$$

where $x \in \mathbb{R}^N$, $z \in \mathbb{R}^m$, $A \in \mathbb{R}^{L \times N}$, $B \in \mathbb{R}^{L \times m}$ and $c \in \mathbb{R}^L$. The functions $f_1$ and $f_2$ are assumed to be convex. The ADMM algorithm is summarized as the following steps [82]:

The augmented Lagrangian polynomial in the above algorithm is based on the
2.3. SYSTEM MODELS AND LINEAR SPARSE HYPERSPECTRAL UNMIXING

Algorithm 1 ADMM [82]

1: Set $j=0$, choose $\mu > 0$, $z^0$ and $u^0$.
2: Repeat:
3: $x^{j+1} \in \arg \min_x L_\mu(x, z^j, u^j)$
4: $z^{j+1} \in \arg \min_z L_\mu(x^{j+1}, z, u^j)$
5: $u^{j+1} \leftarrow u^j + \mu(Ax^{j+1} + Bz^{j+1} - c)$
6: $j \leftarrow j + 1$
7: stopping criterion is satisfied.

method of multipliers and it is defined as

$$L_\mu(x, z, y) = f_1(x) + f_2(z) + u^T(Ax + Bz - c) + \frac{\mu}{2} ||Ax + Bz - c||_2^2,$$  \hspace{1cm} (2.11)

where $\mu$ is the penalty parameter.

2.3.2 Non-dictionary-aided blind sparse unmixing methods

In all of the above-mentioned methods, the main goal is to find the fractional abundances of the materials for a given spectral library. Indeed, the endmembers are extracted using algorithms such as the vertex component analysis (VCA) method [43], automatic morphological endmember extraction (AMEE) algorithm [44], the N-FINDR algorithm [45], the pixel purity index (PPI) [46] or their variants [47, 48, 49].

In spite of such two stages-based unmixing methods that employ both the endmember extraction methods (e.g., N-FINDR [45] and VCA [43]) and mixed pixels decomposition methods (i.e., $\ell_2$-norm, $\ell_1$-norm, and $\ell_0$-norm approximation, [33, 50, 22]), one-step methods such as nonnegative matrix factorization (NMF) [51, 52] based approach is highly interested because of its noticeable supports. First of all, the non-negativity constraints for both spectral signatures and their fractional abundances
are automatically included in the NMF-based methods. Furthermore, it can make decomposition matrices to be more intractable because of a part-based representation of the data, see [53, 5] and references therein.

To illustrate the basic idea of NMF method, we rewrite the linear model of equation (2.2) for \( P \) pixels as follows

\[
Y = MX + E, \tag{2.12}
\]

where, \( Y \in \mathbb{R}^{L \times P} \), is a \( L \times P \) matrix, \( X = [x_1, x_2, ..., x_P] \in \mathbb{R}^{N \times P} \), \( E \in \mathbb{R}^{L \times P} \) and \( E \) is the mixing matrix defined earlier. The goal of NMF is to decompose the observed matrix \( Y \) as product of two nonnegative matrices as follows

\[
Y \approx AX, \tag{2.13}
\]

where \( A \in \mathbb{R}^{L \times N} \) and \( X \in \mathbb{R}^{N \times P} \) with \( N < \min(L, P) \). The classic NMF method can be used for the hyperspectral unmixing through the following minimization problem

\[
\min_{A, X} ||Y - AX||_F^2 \quad \text{subject to} \quad A \geq 0, X \geq 0 \tag{2.14}
\]

where \( ||.||_F \) denotes the Frobenius norm.

Since the objective function in equation (2.14) is nonconvex with respect to both \( A \) and \( X \), finding the global minimum of (2.14) is difficult and it is NP-hard optimization problem [83] beyond the very large feasible set of solution without any further constraints over that. Hence, using iterative methods such as multiplicative

\footnote{It should be noted that we keep the matrix \( M \) as the true (ground-truth) spectral library for the semiblind unmixing scenarios that it is given a priori. We use the matrix \( A \) for the blind unmixing scenarios that we have to estimate it in a proposed unmixing method.}
Updating rule [7] are considered as one of the efficient tools to solve such NMF-based problems. Furthermore, the regularization methods are of interest in order to add the sparsity constraint terms for several recent proposed unmixing methods. These regularizes can be added to the objective function in (2.14) through different terms such as the $\ell_2$-norm regularizer [84] and the $\ell_1$-norm regularizer [85, 6].

Although adding the $\ell_1$-norm term into the problem of (2.14) leads to the sparser solution than the $\ell_2$-norm, imposing other regularizer terms such as $\ell_p(0 \leq p < 1)$ regularization methods are of recent interests for unmixing purpose [5] as the following Lagrangian form

$$\min_{A \geq 0, X \geq 0} \frac{1}{2} \|Y - AX\|^2_F + \lambda \|X\|_1$$

(2.15)

where $\|X\|_1 = \sum_{i,j=1}^{N \times P} x_{ij}^{\frac{1}{2}}$, $x_{ij}$ is the abundance fractions for the $i$-th endmember at the $j$-th pixel.

Due to the nonconvexity of the corresponding cost function, the algorithm prone to noise corruption and computationally demanding [5]. Although there are various proposed methods based on the NMF approach for unmixing purpose such as $\ell_{1/2}$-norm sparsity constrained [5], substance dependence constrained [55] and manifold regularization into the sparsity constraint [54], they have their own drawbacks and the researches are still working to introduce better (sparser) constrained terms into the cost function in order to improve the current methods. Generally, the unmixing methods using equation (2.15) suffer from numerical problems because of non-differentiability of the $\ell_{1/2}$-norm term at 0’s. Thus, substituting the $\ell_{1/2}$-norm term with smooth functions is one of the recent approach for the scenario of blind unmixing method.
Chapter 3

Semiblind Sparse Hyperspectral Unmixing Via

\( \ell_p \)-norm

3.1 Adaptive LASSO Hyperspectral Unmixing Using ADMM

In this section, we employ the ADMM approach [86, 82] to solve an adaptive version [87] of least absolute shrinkage and selection operator (lasso) [88] problem for the unmixing purpose. Since the contribution of endmembers appearing in each pixel in the HSI application is very small, the optimization problem using the sparsity of the \( \ell_1 \)-norm term is object of interest. Although the least angle regression (LARS) algorithm proposed in [89] is used to solve the adaptive lasso [1], finding another approach in order to achieve the better performance in the sense of reconstruction error with a reasonable computational cost is our main goal. Hence, utilizing the ADMM method of which applied for the various constrained sparse problems in hyperspectral imaging [3] motivates us to apply it for our adaptive lasso optimization problem. By introducing new objective function for the corresponding optimization problem, we propose our methods to solve the adaptive lasso in conjunction with the hyperspectral
unmixing. We show that our approach help us to reconstruct the fractional abundances with the lower mean square error (MSE) and the higher reconstruction signal to noise ratio (RSNR).

3.1.1 The problem formulation

By having the measured hyperspectral data in the form of (2.2), one can consider the following $\ell_0$ norm minimization problem

$$
\min_x ||x||_0 \quad \text{subject to } ||Mx - y||_2^2 \leq \epsilon,
$$

where $||x||_0$ is the number of nonzero $x_i$'s and $\epsilon$ is a small positive parameter. The $\ell_0$ norm problem is an NP-hard problem since it needs a brute force combinatorial search [23]. However, this problem can be replaced by the $\ell_1$ norm problem as follows [88]

$$
\min_x ||x||_1 \quad \text{subject to } ||Mx - y||_2^2 \leq \epsilon,
$$

where $||x||_1 = \sum_{i=1}^N |x_i|$. The corresponding regularization technique for both estimation and variable selection for the $\ell_1$ norm problem was proposed by Tibshirani [88] as the following lasso estimate problem

$$
\hat{x}_l = \arg \min_x ||Mx - y||_2^2 + \lambda ||x||_1,
$$

where $\lambda$ is a nonnegative regularization parameter.

Similar to other various lasso problems regards to HSI process (i.e. [90, 91]),
the lasso problem used an equally penalized of the coefficients for the \( \ell_1 \) norm term involved in its corresponding objective function. Zou [87] introduced an adaptive lasso to mitigate an inconsistent selection of the endmembers as follows

\[
\hat{x}_{al} = \arg\min_x ||Mx - y||_2^2 + \lambda \sum_{i=1}^N w_i |x_i|,
\]  

(3.4)

where \( w = [w_1, w_2, ..., w_N]^T \) is a weighting vector.

An initial solution of the weighting components can be chosen by the least square estimator of the fractional abundances as follows

\[
w_i = \frac{1}{|x_{i,ls}|^\gamma} \quad i = 1, 2, ..., N.
\]  

(3.5)

for \( \gamma \geq 0 \) and \( x_{i,ls} \) is the \( i \)-th element of the least square estimator. Themelis et al used the modified version of LARS algorithm applied by Zou [87] for an augmented optimization problem in which the ASC condition is imposed to the adaptive lasso [1]. Now, we use the ADMM approach to solve the adaptive lasso by choosing the appropriate objective functions and constraints present in a kind of generalized LASSO form [82].

3.1.2 Our proposed unmixing algorithms

First, we reformulate the optimization adaptive lasso problem (3.4) with the ANC and ASC conditions as follows

\[
\min_x \frac{1}{2} ||Mx - y||_2^2 + \lambda ||z||_1 + \nu_1(1^T x) + \nu_{R_N}(x) \quad \text{subject to} \quad Wx - z = 0,
\]  

(3.6)
3.1. ADAPTIVE LASSO HYPERSPECTRAL UNMIXING USING ADMM

where $W = \text{diag}(w)$ is the diagonal matrix of the weighting vector and $\mathcal{I}_Q(x)$ is the indicator function either 0 or $\infty$ if $x \in Q$ or $x \notin Q$, respectively.

Then, we derive the required steps of the update rules by applying the augmented Lagrangian polynomial term (2.11) appearing in the ADMM approach. The acquired steps to solve the minimization problem in (3.4) are given as the following algorithm and we call it the *adaptive lasso with ADMM* (ALwADMM1).

**Algorithm 2** Adaptive LASSO with ADMM (ALwADMM1).

1. Initialize $j = 0$, and choose $W, z^0, u^0, \mu > 0$ and $\gamma > 0$.
2. Repeat:
   3. $B \leftarrow M^T y + \mu W^T (z^j - u^j)$
   4. $A \leftarrow M^T M + \mu W^T W$
   5. $x^{j+1} \leftarrow A^{-1} B - A^{-1} 1 (1^T A^{-1} 1)^{-1} (1^T A^{-1} B - 1)$
   6. $z^{j+1} \leftarrow \max\{0, \text{soft}(W x^{j+1} + u^j, \lambda/\mu)\}$
   7. $u^{j+1} \leftarrow u^j + W x^{j+1} - z^{j+1}$
   8. $W \leftarrow \text{diag}((|x^{j+1}|^\gamma + \delta)^{-1})$
   9. $j \leftarrow j + 1$
10. stopping criterion is satisfied.

The soft function is defined similar to the soft threshold function introduced in [92] as follows

$$\text{soft}(x, \alpha) = \text{sign}(x)(|x| - \alpha)_+, \quad (3.7)$$

where $(x)_+ = \max(0, x)$. In our steps of the proposed algorithm (ALwADMM1), the soft threshold is applied on the vector $x$ in component-wise manner and update the components of $W$ similar to (3.5) in which $x$ is replaced by the value of its previous iteration. Since some elements of vector $x$ will be equal to zero during the iterations, a small positive constant value of $\delta$ is added to avoid inconsistent situation for the updated weights.
3.1. ADAPTIVE LASSO HYPERSPECTRAL UNMIXING USING ADMM

In addition to the first proposed algorithm, we apply the simplified version of ADMM approach used in [3] to solve the problem (3.4) under both conditions of ANC and ASC. This version of ADMM approach have been used for other types of $\ell_1$ norm minimization problems to solve the CSR problem [3] and the sparse group lasso problem [13]). Now, we employ it to solve the problem (3.6). The acquired algorithm is called ALwADMM2.

**Algorithm 3** Adaptive LASSO with ADMM (ALwADMM2).

1: Initialize $j = 0$, and choose $w, z^0, u^0, \mu > 0$ and $\gamma > 0$.
2: Repeat:
3: $A \leftarrow M^T M + \mu I$
4: $B \leftarrow A^{-1}1(1^T A^{-1}1)^{-1}$
5: $v \leftarrow M^T y + \mu(z^j + u^j)$
6: $x^{j+1} \leftarrow A^{-1}v - B(1^T A^{-1}v - 1)$
7: $z^{j+1} \leftarrow \max\{0, \text{soft}(x^{j+1} - u^j, \lambda/\mu), \lambda/\mu w\}$
8: $u^{j+1} \leftarrow u^j - x^{j+1} + z^{j+1}$
9: $w \leftarrow (|x^{j+1}|^\gamma + \delta)^{-1}$
10: $j \leftarrow j + 1$
11: stopping criterion is satisfied.

Finally, both algorithms end up when $||x||_1$ exceeds 1. Adjusting the regularization parameter $\lambda$ and $\mu$ in the proposed algorithms are critical and are computed as explained in [82]. We initialize $\gamma$ (0.2 in our simulations) and propose to update it adaptively as follows:

$$\gamma \left\{ \begin{array}{ll}
2\gamma, & \text{if } ||x^j - z^j||_2 > \mu ||z^j - z^j||_2 \\
\gamma/2, & \text{otherwise}
\end{array} \right. \quad (3.8)$$

where $\ell_j = j - \text{mod}(j, 10)$. The term $||z||_1$ in (3.6) is an approximation of $\ell_1 - \gamma$ - norm of $x$ around the latest solution. Thus, this algorithm allows to reach sparser solutions, since for larger values of $\gamma$ the solution tends to the solution using $||x||_0$. 

3.1. ADAPTIVE LASSO HYPERSPECTRAL UNMIXING USING ADMM

3.1.3 Simulation results

We examine our proposed algorithms using simulated hyperspectral data from the U.S. Geological Survey (USGS) library [93]. We consider $L = 224$ spectral bands and $N = 20$ endmembers as the basis. Then, we generate pixels using $k$ out of $N$ endmembers where $k$ is the number of nonzero elements of the vector of abundances. The fractional abundances are generated randomly with the Dirichlet distribution and noise is generated randomly as an additive white gaussian (AWGN) with zero mean. First, we reconstruct the abundances at different noise levels in dBs. Figure 3.1 illustrates the estimated fractions of endmembers for two proposed algorithms (ALwADMM1 and ALwADMM2), the adaptive lasso employing LARS [1] and the spectral unmixing by variable splitting and augmented lagrangian (SUnSAL) [3]. In this experiment, the number of nonzero elements is 4 where only two values are shown in Figure 3.1. We observe that our proposed algorithms have the better estimation of true values compared with the other methods proposed in [1] and [3] at different SNRs.

In the second experiment, we evaluate our proposed algorithms in terms of the MSE in different SNR values. Figure 3.2 compares the results of our proposed methods with the proposed method of [1] and the least square method [4]. We observe that the proposed algorithms outperform the other mentioned algorithms in terms of MSE for different SNRs.

To evaluate the reconstruction quality of fractional abundances, we use the reconstruction SNR (RSNR) defined in [94, 3] as follows

$$\text{RSNR (dB)} = 10 \log_{10} \left( \frac{E[||x||^2]}{E[||x - \hat{x}||^2]} \right),$$

(3.9)
3.1. ADAPTIVE LASSO HYPERSPECTRAL UNMIXING USING ADMM

Figure 3.1: The results of reconstruction fractions of endmembers for ALwADMM1, ALwADMM2, adaptive lasso using LARS (alasso) [1] and lasso via SUnSAL [3], the index of nonzero elements are shown.
3.1. ADAPTIVE LASSO HYPERSPECTRAL UNMIXING USING ADMM

Figure 3.2: The mean square errors of the estimated fractional abundances versus SNRs for ALwADMM1, ALwADMM2, least square method [4] and alasso [1].

where \( \hat{x} \) is the reconstructed fractional abundance vector obtained from different algorithms. Table 3.1 illustrates the RSNR values for the proposed algorithms and the proposed unmixing method of [1]. The proposed algorithms exhibit the highest RSNR in dBs in comparison with the proposed method of [1]. In particular, more improvements are achieved at higher SNRs. We conclude that significant enhancement is gained instead of using LARS to solve the adaptive lasso problem.
3.1. ADAPTIVE LASSO HYPERSPECTRAL UNMIXING USING ADMM

Table 3.1: The comparison results of RSNR values for ALwADMM1, ALwADMM2 and the proposed method of [1].

<table>
<thead>
<tr>
<th>SNR (dB)</th>
<th>RSNR (dB) (alasso [1])</th>
<th>20</th>
<th>40</th>
<th>60</th>
<th>80</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RSNR (dB) (ALwADMM1)</td>
<td>12.8</td>
<td>22.6</td>
<td>25.5</td>
<td>25.7</td>
</tr>
<tr>
<td></td>
<td>RSNR (dB) (ALwADMM2)</td>
<td>12.8</td>
<td>22.8</td>
<td>27.3</td>
<td>27.3</td>
</tr>
</tbody>
</table>

3.1.4 Conclusion

We proposed a new approach to solve the adaptive lasso problem for unmixing hyperspectral imaging with the alternating direction method of multipliers. After introducing our objective of optimization problems, two algorithms were proposed. These methods are able to solve the weighted lasso problem and gave enhanced performance in HSIs unmixing process compared with the previous state-of-the-art LARS algorithm in [1]. To evaluate the performance, the mean square error and reconstruction SNR were computed for our proposed algorithms and the other proposed unmixing methods over the USGS library. The simulation results showed that the proposed methods outperform other methods. Further experimentations are required to conduct with more real hyperspectral data sets in order to criticize the proposed algorithms.
3.2 Sparse Hyperspectral Unmixing via Varying $\ell_p$-norm Approximation of $\ell_0$-norm

In this section, we propose an algorithm for the SU by minimizing the number of nonzero elements of the solution under the nonnegativity and the sum-to-one constraints through $\ell_p$-norm while $p$ is iteratively decreasing. We propose to start $p$ at 1, in which the problem is convex and a unique solution is possible under some conditions [38, 42]. Theorem 1 of this section reveals that the set of local optima of the $\ell_p$-norm problem is a continuous function of $p$. This means that by continuous/gradual reduction of $p$, it may be possible to track global optima of the problem (i.e., if they exist) with respect to $p$ in order to reach an enhanced sparser solution. Since the $\ell_p$-norm is not Lipschitz continuous and differentiable for $p < 1$, we use the smooth approximation [41, 95, 96, 97] by introducing a small positive number $\epsilon \geq 0$. This approximation tends to the $\ell_p$-norm as $\epsilon \to 0$. We propose decreasing updating rules for $p \in (0, 1]$ and $\epsilon$, respectively to minimize the chance that the tracked solution be trapped in the sub-optimal minima, and to ensure the $\ell_p$-norm is differentiable at 0. By using this concept, our proposed method aims to find a sparser solution [38] while the parameter $p$ is reduced smoothly. Moreover, we utilize the ADMM approach [86, 82] to solve our adaptive version of objective function in the sense of varying the value of norm $p$ based on one of our proposed updating methods. Also, we approximate the $\ell_p$-norm through an iteratively reweighted $\ell_2$-norm with varying $p$. We are able to solve the corresponding minimization problem with the closed form of the solution (i.e., for the unconstraint version of the problem) and update the involved parameters $(p, \epsilon)$ based on the mentioned proposed updating rules. Ultimately, we evaluate our proposed method over the synthetic data as well as real hyperspectral
data and show that they have a better performance such as the higher RSNR as well as the probability of success (PoS) compared with some other state-of-the-art algorithms by simulation results.

3.2.1 \( \ell_p \) regularized problem with variable \( p \)

Consider the \( \ell_p \)-norm \((0 < p \leq 1)\) problem as

\[
\begin{align*}
\min_{x \in S} f(x, p), \\
f(x, p) &= \frac{1}{2} ||y - Mx||_2^2 + \lambda ||x||_p^p,
\end{align*}
\]

where \( ||x||_p^p = \sum_{i=1}^{N} |x_i|^p \) and \( \lambda > 0 \) adjusts the trade-off between the sparsity of the potential solution and the squared error of the estimated fractional abundances of endmembers. In [98] it has been proved that for any constant values of \( p \) smaller than one, the above problem is strongly NP-hard. This means that as \( p \) is reduced, the number of local minima of \( f(x, p) \) increases. However, each of its local minimizer can be computed in a polynomial time, see Theorem 3 and Section 4 in [36]. Chen et al. obtained some lower bounds for non-zero entries of these local minimizers and proposed a hybrid orthogonal matching pursuit-smoothing gradient method for solving (3.10a) [36].

The above problem is convex for \( p \geq 1 \) and there are efficient algorithms for solving this problem. Unfortunately, the problem is not convex for \( p < 1 \). However, because of two folklores in high-dimensional statistic the problem is heavily investigated in the literature [38]. The first folklore is that the optimal solution for smaller \( p \) tends to be the optimal solution for the \( \ell_0 \)-norm problem. The second folklore is that using iterative methods and good initialization, the \( \ell_p \)-norm solution converges closer to
3.2. SPARSE HYPERSPECTRAL UNMIXING VIA VARYING \( \ell_p \)-NORM APPROXIMATION OF \( \ell_0 \)-NORM

the \( \ell_0 \)-norm solution than the \( \ell_1 \)-norm solution [38]. There are plenty of experimental results that support these folklores which are not theoretically established yet, see [38] and references therein.

By having the fact the \( \ell_p \)-norm term approaches to \( \ell_1 \)-norm and \( \ell_0 \)-norm while \( p \) goes to 1 and 0, respectively, we first propose the following weighted \( \ell_1 \)-norm approximation for \( ||x||^p \)

\[
\sum_{i=1}^{N} |x_i|^p \approx \sum_{i=1}^{N} |x_i| w_i^{(t)},
\]

\( w_i^{(t)} = (|x_i^{(t)}| + \epsilon^{(t)})^{p(t)-1} \)

where \( \epsilon \) is an small positive value to make the weights \( w_i \) Lipschitz continuous and \( t \) is the index of the last iteration.

Definition (Hausdorff distance [99, 100]). Let \( \mathcal{X} \) be the set of all finite subsets of \( \mathbb{R}^N \), then \( (\mathcal{X}, d) \) is a metric space where the Hausdorff distance \( d(A, B) \) of two sets of \( A \) and \( B \) belongs to \( \mathcal{X} \) is defined by

\[
d(A, B) = \max\{\sup_{x \in A} \inf_{y \in B} ||x - y||_{\infty}, \sup_{y \in B} \inf_{x \in A} ||x - y||_{\infty}\},
\]

where \( ||z||_{\infty} = \max_{i \in \{1, 2, \ldots, N\}} |z_i| \).

Using the above definition, the following theorem proves the continuity of the set of candidate local minima of (3.10a) with respect to norm \( p \). The continuity in Theorem 1 gives insight into the first folklore and states that the solution obtained using previous value of \( p \) is a good initialization for the next iteration with slightly smaller \( p \) as expected from the second folklore [38]. In other words, this continuity implies that
by starting \( p = 1 \) and iteratively reducing it, we expect to obtain enhanced sparse solution. Appropriate reduction of \( p \) results in tracking of better local optima.

**Theorem 1.** The set of all local optima of

\[
g(x, p, \epsilon) = \frac{1}{2} \|y - Mx\|^2 + \lambda \sum_{i=1}^{N} (|x_i| + \epsilon)^p,
\]

in \( \mathcal{S} \) denoted by \( X_{p,\epsilon} \) is a continuous function of \( p \in (0, 1] \) for all \( \epsilon > 0 \).

**Proof.** For the proof of the Theorem, see Appendix A.1. \( \square \)

In this theorem, \( X_{p,\epsilon} \) is a continuous function of \( p \) for all \( \epsilon > 0 \). Thus \( \lim_{\epsilon \to 0} X_{p,\epsilon} \) hereafter denoted by \( X_p \) exists and is also a continuous function of \( p \) which corresponds to the set of local optima of \( f(x, p) = \lim_{\epsilon \to 0} g(x, p, \epsilon) \). The continuity in this theorem means that the optimal trajectory \( x_p \in X_p \) versus \( p \) is a continuous function of \( p \), i.e., small variations of \( p \) could not result in large jumps in \( x_p \). For \( x_p \in X_p \), it is easy to show that

\[
\frac{\partial f(x_p, p)}{\partial p} = \frac{\partial f(x_p, p)}{\partial p} + \nabla_x f(x_p, p) = \lambda \sum_i \ln(|x_i|)|x_i|^p
\]

which is non-positive, bounded and continuous in terms of \( p \) for \( p > 0 \) and \( 0 \leq x_i \leq 1 \). Thus, the optimal solution of (3.10a) for a given value of \( p \) is a good initial point for another slightly reduced value of \( p \).

For the initial value \( p = 1 \) or \( p > 1 \), the problem in (3.10a) is convex and may have unique member under some conditions [42]. However, the number of local minima of \( f(x, p) \) certainly increases as \( p \) becomes smaller than 1. This theorem proves that the mapping \( X_p : p \in [0, 1] \mapsto \mathcal{X} \) which indicates the set of local minima of the problem is a continuous function of \( p \). Thus one may track the optimal solution provided that \( p \) could have been reduced continuously and gradually in each iteration. In practice, \( p \) can only take decreasing values from a countable set through an iterative approach.
3.2. SPARSE HYPERSPECTRAL UNMIXING VIA VARYING \( \ell_p \)-NORM APPROXIMATION OF \( \ell_0 \)-NORM

(e.g., as in (3.20a), (3.21) or (3.22a) where \( \alpha = -\ln p \)), however allows to find a better solution (in the sense of \( \ell_0 \) norm) in each iteration by reducing \( p \). For \( p < 1 \) the function \( ||x||_p \) in (3.10b) is not Lipschitz continuous which results in numerical problem. This is because for the zero elements of the sparse vector \( x \), the gradient term \( x^{p-1} \) is undefined where \( x^{p-1} = [\text{sign}(x_1)|x_1|^{p-1}, \ldots, \text{sign}(x_N)|x_N|^{p-1}]^T \). As remedy, several approximations were used in the literature for \( ||x||_p \) such as \( \sum_{i=1}^N (|x_i| + \epsilon)^p \) and \( \sum_{i=1}^N (|x_i|^2 + \epsilon^2)^{\frac{p}{2}} \) where \( \epsilon > 0 \) is an small fixed value [95, 96, 97]. Different versions of the efficient iterative reweighted \( \ell_1 \)-norm algorithms were proposed using such approximations, e.g. see [95, 97, 101] and references therein.

The approximation of \( \ell_p \)-norm by weighted \( \ell_1 \)-norm has already been proposed in [102, 87] where a fixed value for \( p \) is used in contrast to this paper. The right-hand-side (RHS) of (3.11) converges exactly to its left-hand-side (LHS) if \( x_i^{(t)} \) converges to \( x_i \) while \( \epsilon^{(t)} \) and \( p^{(t)} \) tend to 0 and \( p \), respectively. Theorem 1 implies that continuous variation of \( p \) and \( \epsilon \) toward 0 allows tracking of the optimal solution. Such a continuous variation is not possible, however, we expect the proposed algorithms in the limit to follow the optimal solution if the discrete variations of \( p \) and \( \epsilon \) toward 0 are small enough. Finding optimal updating rules for \((p, \epsilon)\) decreasing toward zero is still an open problem to guarantee the continuity of the tracked solution toward the optimal \( \ell_0 \)-norm solution. Hence, we will introduce three approximated updating rules in a practical view to propose our final unmixing methods.

Rewriting the approximated problem in (3.10a), we have

\[
\min_{x \in \mathbb{S}} \frac{1}{2}||y - Mx||_2^2 + \lambda||w^{(t)} \odot x||_1,
\]

where \( w^{(t)} = [w_1^{(t)}, \ldots, w_N^{(t)}] \) is the weight vector defined by (3.12) and \( \odot \) is the
3.2. SPARSE HYPERSONTICAL UNMIXING VIA VARYING

\( \ell_p \)-NORM APPROXIMATION OF \( \ell_0 \)-NORM

Hadamard (entrywise) product. This problem can be solved using the ADMM algorithm [82, 86] which is an efficient numerical optimization tool.

In [30] for a constant \( p \), the function \( ||x||^{p}_{p} \) is approximated with the weighted \( \ell_2 \)-norm. Here we propose to iteratively update this approximation as \( p \) varies in order to solve (3.10a) as follows

\[
\min_{x \in \mathbb{R}^N} ||y - Mx||^2_2 + \lambda ||d^{(t)} \odot x||^2_2, \tag{3.16}
\]

where \( d^{(t)} = [d_1^{(t)}, \ldots, d_N^{(t)}] \) is the weight vector calculated by

\[
d_i^{(t)} = \left( (x_i^{(t)})^2 + (\epsilon^{(t)})^2 \right)^{(p^{(t)} - 2)/4}, \quad i = 1, \ldots, N \tag{3.17}
\]

where \( \epsilon \) and \( t \) have the same descriptions as mentioned before. We propose to use the projection step on the constraint simplex to impose the ASC and ANC. A simple alternative method is renormalizing the ultimate solution and replacing the nonnegative components of the solution in each iteration to impose the ASC and ANC, respectively.

In our setting, we update the parameters \( p \) and \( \epsilon \) in each iteration to enhance the accuracy of our approximation iteratively. The goal is to adjust the existing parameters such that \( \epsilon^{(t)} \) tend to zero and \( p^{(t)} \) to be reduced to some small value while \( x^{(t)} \) converges iteratively.

3.2.2 Updating rules for pair \((p^{(t)}, \epsilon^{(t)})\)

In this section, we introduce three different ways to update \( p^{(t)} \) and \( \epsilon^{(t)} \).
3.2. SPARSE HYPERSPECTRAL UNMIXING VIA VARYING $\ell_p$-NORM APPROXIMATION OF $\ell_0$-NORM

Gradient descent

The only term in the relation (3.15) or (3.16) that depends on the parameter $p$ is the weight $w_i$’s or $d_i$’s, respectively. We use the gradient descent based method to choose the decreasing sequence for $p$. Let $p(t) = e^{-\alpha(t)}$ for $t = 0, 1, \cdots, T - 1$ and $\alpha(0) = 0$ where $T$ is the known integer value representing the maximum iteration number.

By denoting $L_1 = \|w^{(t)} \odot x\|_1 = \sum_{i=1}^{N} |x_i|(|x_i| + \epsilon)^{p-1}$, and $L_2 = \|d^{(t)} \odot x\|_2^2 = \sum_{i=1}^{N} |x_i|(|x_i|^2 + \epsilon^2)^{(p-2)/4}$, respectively, we have

$$
\frac{\partial L_1}{\partial \alpha} = \left. \frac{\partial L_1}{\partial p} \frac{\partial p}{\partial \alpha} \right|_{x = x^{(t)}} = -\sum_{i=1}^{N} |x_i^{(t)}| \ln(|x_i^{(t)}| + \epsilon^{(t)}) \left( |x_i^{(t)}| + \epsilon^{(t)} \right)^{p(t)-1} e^{-\alpha^{(t)}}. 
$$

and

$$
\frac{\partial L_2}{\partial \alpha} = -\frac{1}{2} \sum_{i=1}^{N} \left( |x_i^{(t)}| \ln( |x_i^{(t)}|^2 + (\epsilon^{(t)})^2) \right) 
\left( |x_i^{(t)}|^2 + (\epsilon^{(t)})^2 \right)^{(p(t)-2)/4} e^{-\alpha^{(t)}}. 
$$

Now, we use the gradient descent to present our first updating rule as follows

$$
\alpha^{(t+1)} = \alpha^{(t)} - \sigma \frac{\partial L_i}{\partial \alpha} 
$$

(3.20a)

where $L_i$ is either $L_1$ or $L_2$ and $\sigma$ is a selected constant learning rate value. Since $\frac{\partial L_i}{\partial \alpha} < 0$ in (3.18) and (3.19), the sequence $\alpha^{(t+1)}$ is increasing, i.e., $p^{(t+1)}$ is decreasing.

To update the parameter $\epsilon$, we may utilize the same method used for the convergence
3.2. SPARSE HYPERSPECTRAL UNMIXING VIA VARYING \( \ell_p \)-NORM APPROXIMATION OF \( \ell_0 \)-NORM

of IRLS algorithm in [26] as follows

\[
e^{(t+1)} = \min\{e^{(t)}, \frac{r(x^{(t)K+1})}{N}\} \tag{3.20b}
\]

where \( r(\cdot)_K \) is the \( K \)-th largest element of non-increasing rearrangement of the absolute values of its entries and \( K \) is a fixed integer.

**Taylor expansion**

In this approach we aim to control \( \frac{\partial \ln L_i}{\partial \alpha} \) to a constant value \( \tau \). The motivation behind this approach is intuitive; if we assume that the variation of \( L_i \) is somehow related to the chance of a bifurcation happening; thereby this approach aims to control the chance of a skipped bifurcation to be constant. Using the first term of Taylor expansion of \( L_i \) with respect to \( \alpha = -\ln p \) and some manipulation, we obtain:

\[
\alpha^{(t+1)} = \alpha^{(t)} + \tau \frac{|L_i|}{|\frac{\partial L_i}{\partial x}|} \tag{3.21}
\]

where \( L_i \) is calculated using (3.18) or (3.19) at the previous value \( x^{(t)} \). We update \( e^{(t+1)} \) using (3.20b).

**Exponential updating**

In this scheme, we use the exponential updating rules similar to [41] as follows

\[
p^{(t+1)} = p_0 e^{-\gamma t} \tag{3.22a}
\]

\[
e^{(t+1)} = e_0 e^{-\beta t} \tag{3.22b}
\]
where $p_0$ and $\epsilon_0$ are initial values, $\gamma > 0$ and $\beta > 0$ are two fixed decay rate values. The updating rules in (3.22) are initially proposed in [41] (i.e., for a different approximation term of $\ell_0$-norm from in this paper) in compressive sensing. Precisely, we propose to use different methods to solve the $\ell_p$-norm problem in (3.10a) as opposed to the BCG algorithm [103], and use the rule in (3.22) only for updating the pair $(p, \epsilon)$. In our simulations, we initialize $p_0 = 1$ and $\epsilon_0 = 1$.

3.2.3 Proposed methods for solving (3.15) and (3.16)

The weighted $\ell_1$-norm problem (3.15) is a convex approximation of problem (3.10a) around the last solution which is a quadratic problem under ANC and ASC. The Lagrangian for (3.15) is

$$
\min_x \frac{1}{2} \|y - Mx\|_2^2 + \lambda \|w^{(t)} \odot x\|_1 + \iota_{\{1\}}(1^T x) + \iota_{\mathbb{R}_+^N}(x),
$$

(3.23)

where $1$ is the column vector of 1’s and $\iota_Q(x)$ is the indicator function either 0 if $x \in Q$ or $\infty$ for $x \notin Q$. To solve this problem, we employ the scaled version of the ADMM algorithm [82] and construct the augmented Lagrangian form of the object function (3.23) as

$$
L_{\mu}(x, z, u, w) = C_1(\mu, x, z, u, w) + \iota_{\{1\}}(1^T x) + \iota_{\mathbb{R}_+^N}(x),
$$

(3.24)

where the cost function $C_1$ is defined by

$$
C_1(\mu, x, z, u, w) = \frac{1}{2} \|y - Mx\|_2^2 + \lambda \|z\|_1 + \frac{\mu}{2} \|w \odot x - z + u\|_2^2.
$$

(3.25)
3.2. SPARSE HYPERSPECTRAL UNMIXING VIA VARYING
ℓ_p-NORM APPROXIMATION OF ℓ_0-NORM

Following the steps in the ADMM approach, we derive our proposed Varying ℓ_p-
Norm Approximation (VNA) algorithms by using one of the updating rules in (3.20),
(3.21), or (3.22) for L_1 and refer to them respectively as VNA1 (using the Gradient
descent), VNA2 (using the Taylor expansion) and VNA3 (using the Exponential
approach). These are summarized in Algorithm 4. For stopping criterion, we set
ϵ_{thr} = 10^{-8} and T = 100 for all simulations.

Algorithm 4 Pseudocode of the ℓ_p approach with varying p obtained by solving
(3.15).

```
1: Initialize t = 0, and choose w(0), z(0), u(0), µ > 0, λ > 0.
2: while ϵ(t) > ϵ_{thr} and t < T do
3:   A ← M^T M + diag(µw(t) ⊙ w(t))
4:   b ← M^T y + µw(t) ⊙ (z(t) − u(t))
5:   x(t+1) ← max(A^† b − \frac{1}{T^2 A^{T-1} A^T} 1^T, 0)
6:   z(t+1) ← w(t) ⊙ x(t+1) + u(t) − λ/µ
7:   u(t+1) ← u(t) + w(t) ⊙ x(t+1) − z(t+1)
8:   Update the pair (p(t), ϵ(t)) using (3.20) or (3.21) where p(t) = e^{-α(t)} or (3.22)
9:   Update elements of w(t+1) using (3.12)
10: Using (3.25) if C_1(µ, x(t+1), z(t+1), u(t+1), w(t)) < C_1(µ, x(t+1), z(t+1), u(t+1), w(t)), do w(t+1) ← w(t) and stop.
11: t ← t + 1
12: end while
```

The weighted ℓ_2-norm problem (3.16) is a quadratic approximation of the problem
(3.10a) around the last solution at each iteration. The problem in (3.16) is convex and
has a unique solution with no closed form. As remedy, we first ignore the constraints
x_k ≥ 0 leading to the following closed form solution

\[ \tilde{x} = C^\dagger (M^T y - \kappa 1) \]  (3.26a)

where \( \tilde{x} \) is the estimated solution of (3.16) and \( (.)^\dagger \) is the Moore-Penrose pseudoinverse
3.2. SPARSE HYPERSPECTRAL UNMIXING VIA VARYING \( \ell_p \)-NORM APPROXIMATION OF \( \ell_0 \)-NORM

of a matrix. The value of \( \kappa \) and \( C \) are also computed as follows

\[
\kappa = (1^T C^T 1)^{-1} (1^T C^T M^T y - 1) \tag{3.26b}
\]
\[
C = M^T M + \lambda \text{diag}(d^{(t)} \odot d^{(t)}) \tag{3.26c}
\]

Note that \( C \) is full rank \(^1\). By projecting \( \tilde{x} \) on the nonnegativity constraint, we obtain \( \hat{x} = \max\{\tilde{x}, 0\} \) as an approximate solution for (3.16) where max is the entrywise maximization. Our extensive experiments show that the above solution gives an accurate approximate solution for (3.16) because of the accuracy of previous solution.

By considering the objective function (3.16) as follows

\[
C_2(x, d) = \|y - Mx\|_2^2 + \lambda \|d \odot x\|_2^2,
\]

we complete our second proposed method as summarized in Algorithm 5. We note that we derive the similar proposed VNA algorithms by using one of the updating rules in (3.20), (3.21), or (3.22) and refer to them respectively as VNA4 (using the Gradient descent), VNA5 (using the Taylor expansion) and VNA6 (using the Exponential updating rule).

3.2.4 The convergence analysis

The convergence of the proposed methods (VNA1 to VNA6) are enforced/proved by checking the cost functions to be decreasing and we will support the effectiveness of our proposed algorithms by the following experimental evidences.

\(^1\)This is because for all \( \|a\|^2 > 0 \), we have \( a^T C a = \|M a\|^2 + \lambda \|d^{(t)} \odot a\|^2 > \lambda \chi \|a\|^2 > 0 \) where \( \chi \) is a function of \( (\epsilon^{(t)}, p^{(t)}) \).
3.2. SPARSE HYPERSPECTRAL UNMIXING VIA VARYING $\ell_p$-NORM APPROXIMATION OF $\ell_0$-NORM

Algorithm 5 Pseudocode of the $\ell_p$ approach with varying $p$ obtained by solving (3.16).

1: Initialize $t = 0$, and choose $d^{(0)}$, $\lambda > 0$.
2: while $\epsilon(t) > \epsilon_{thr}$ and $t < T$ do
3: Update $x^{(t+1)}$ using (3.26)
4: Update the pair $(p^{(t)}, \epsilon^{(t)})$ using (3.20) or (3.21) where $p^{(t)} = e^{-\alpha^{(t)}}$ or (3.22)
5: Update elements of $d^{(t+1)}$ using (3.17)
6: Using (3.27) if $C_2(x^{(t+1)}, d^{(t)}) < C_2(x^{(t+1)}, d^{(t+1)})$, do $d^{(t+1)} \leftarrow d^{(t)}$ and stop.
7: $t \leftarrow t + 1$
8: end while

To illustrate the convergence in our simulations, we calculate five values of equation (3.25) respectively before Step 1, after Steps 3, 4, 5 and 7 of Algorithm 4 for VNA1, VNA2 and VNA3 for each iteration. Similarly, for the proposed methods of VNA4, VNA5 and VNA6, we calculate 3 values of (3.27) respectively before Step 1, after Steps 1 and 3 of Algorithm 5 for each iteration. In this experiment, we used the spectral library $M_1$ as introduced in the next section with the sparsity level of 5 and AWGN with SNR= 30dB. The nonzero elements of $x$ are randomly selected as $x(85) = 0.1877, x(299) = 0.3216, x(325) = 0.2398, x(393) = 0.1406$ and $x(491) = 0.1103$ where $x(k)$ represents the $k$-th value of vector $x$. Figure 3.3 shows the variations of the cost functions through different steps of these methods. We observe the convergence of the proposed after several iterations as $p$ reduces. Since the sequences of pair $(p, \epsilon)$ are a set of decreasing sequences versus the iteration number (i.e., by following up updating rules) and we enforce the corresponding conditions in our proposed methods, the convergence will be guarantee. However, it would not easy to mention which method gives the best performance with the small number of iterations due to the nature of the iteratively-based proposed methods as well as convergence restriction. Thus, there is not a direct relation between the number of...
3.2. SPARSE HYPERSPECTRAL UNMIXING VIA VARYING $\ell_p$-NORM APPROXIMATION OF $\ell_0$-NORM

Figure 3.3: The variation of the cost function through different steps and iterations of the proposed methods, VNA1 to VNA6 for AWGN at SNR = 30dB, sparsity level = 5: (a) The cost function (3.25) calculated for 5 steps of Algorithm 1 in each iteration, (b) The cost function (3.27) calculated for 3 steps of Algorithm 2 in each iteration.

iterations to convergence versus the method used to select $p$ and $\epsilon$.

In the next experiment, we show that how the proposed methods estimate the true values after several number of iterations. Figure 3.4 illustrates the results of the non-zero elements of $\mathbf{x}$ which are randomly selected as $x(12) = 0.1602, x(71) = 0.2428$ and $x(202) = 0.5970$ at SNR = 30dB. As expected, Figure 3.4 confirms that the proposed methods converge to the groundtruth values after small number of iterations.

At the end, we should note that selection of the value of $\mu$ for such methods (VNA1, 2, 3) is an active research topic (e.g., see [17, 82] and references therein) as $\mu$ impacts the rate of convergence of the ADMM as well. Here, we employ the equation $\mu = 10\lambda + 0.01$ which is used within the available codes for SUnSAL [3, 82].
3.2. SPARSE HYPERSPECTRAL UNMIXING VIA VARYING \( \ell_p \)-NORM APPROXIMATION OF \( \ell_0 \)-NORM

![Graphs showing estimated entries of x with 3 non-zero elements using proposed methods in AWGN at SNR = 30dB.](image)

**Figure 3.4:** Estimated entries of \( x \) with 3 non-zero elements using proposed methods in AWGN at SNR = 30dB.

### 3.2.5 Experimental results and analysis

We evaluated the proposed algorithms and updating rules in the following experiments.

**Experiments with Synthetic Data**

In the first experiment, we use the following six spectral libraries where the 1st, 4th, 5th and 6th are obtained from the U.S. Geological Survey (USGS) library [93]:

1. \( M_1 \in \mathbb{R}^{224 \times 498} \) chosen from the USGS library [93] contains 498 spectral signatures of minerals with 224 spectral bands provided for measured reflectance values distributed uniformly in the interval 0.4-2.5 \( \mu m \). This library which is
also used in [14, 16, 15, 30] allows to compare our results.

2. $\mathbf{M}_2 \in \mathbb{R}^{224 \times 498}$ randomly generated with i.i.d. components uniformly distributed in the interval $[0,1]$.

3. $\mathbf{M}_3 \in \mathbb{R}^{224 \times 498}$ randomly generated with i.i.d. zero-mean Gaussian components with variance of one.

4. $\mathbf{M}_4 \in \mathbb{R}^{224 \times 342}$ selected a subset of $\mathbf{M}_1$ such that the angles between its columns are larger than $3^\circ$. This is to prune the larger library which contains many similar signatures with very minor variations of the same material as in [14].

5. $\mathbf{M}_5 \in \mathbb{R}^{224 \times 240}$ selected a subset of $\mathbf{M}_1$ such that the angle between its columns are larger than $4.44^\circ$ as in [17].

6. $\mathbf{M}_6 \in \mathbb{R}^{224 \times 70}$ a smaller subset of $\mathbf{M}_1$ such that the angle between the columns are larger than $9.25^\circ$. This allows to further reduce the computational burden.

We generated the hyperspectral data by generating the mixtures (for different sparsity levels) and noise in the same way as in [14, 16, 30] which are some of the state-of-the-art methods and compared our results with theirs. The fractional abundances were generated randomly with a Dirichlet distribution as in [43] for above libraries.

To evaluate the performance of unmixing, we use the RSNR defined in (3.9) and the probability of success (PoS) defined as

$$
\text{PoS} = \Pr\left(\frac{||\mathbf{x} - \hat{\mathbf{x}}||_2}{||\mathbf{x}||_2} \leq \xi\right),
$$

(3.28)
where $\xi$ is a constant threshold, $\mathbf{x}$ and $\hat{\mathbf{x}}$ are the original simulated fractional abundance vector and the reconstructed fractional abundance vector obtained from different algorithms, respectively [14, 17]. We repeated our simulations for all state-of-the-art methods as well as our proposed methods between 200 and 10000 times. More precisely, after obtaining the vector of RSNR over previous runs denoted by RSNR, we calculated the following stopping criterion

$$C = \frac{(\text{std}(\text{RSNR}))^2}{K_1 \times (\text{mean}(\text{RSNR}))^2}$$

(3.29)

where $K_1$ is the number of performed runs and std is the standard deviation. We repeated simulations until $C$ for all methods became less than 0.03 to maintain simulation time manageable. This stop criterion yields that the ratio of the variance of the average RSNR over the average RSNR is 0.03. The number of runs heavily depends on the sparsity level, the location of non-zero components, the type and the size of the library.

In our experiments, we considered different sparsity levels $\|\mathbf{x}\|_0$ (that is the number of non-zero entries of $\mathbf{x}$) between 1 to 30 for SNR of 25 dB and $\xi = 0.65$. We set the parameters of several state-of-the-art methods either as it is reported in the literature or have adjusted them within their source code by try and error as in Table 3.2 as follows:

- wlasso [1]: $\lambda = 1$ and $\gamma = 0.2$.

- BiICE [16]: MaxIter = 50 and $a_{\text{Vita}} = b_{\text{Vita}} = a_\lambda = b_\lambda = 10^{-6}$.

- SUnSAL with activated ANC$^2$ [3, 14]: See Table 3.2.

$^2$We used the codes for SUnSAL from www.lx.it.pt/ bioucas/code/sunsal_demo.zip.
3.2. SPARSE HYPERSPECTRAL UNMIXING VIA VARYING $\ell_p$-NORM APPROXIMATION OF $\ell_0$-NORM

- CLS [3]: $\lambda = 0$.
- CZ [30]: $p = 0.2$ and $\log_{10} \lambda = 0.0008 \text{SNR}^2 - 0.1144 \text{SNR} - 0.9983$.
- PLA [41]: $T = 80, p_1 = 1, p_T = 0, 1, \epsilon_1 = 1, \epsilon_T = 0.01, E_t = 10^{-25}, \delta = 10^{-5}$.
- VNA1, VNA2, VNA4 and VNA5: See Table 3.2.
- VNA3 and VNA6: $\gamma = 0.0291, \beta = 0.0583$ and see Table 3.2 for $\lambda$.

One can tune the parameters of the aforementioned methods for different libraries, SNRs and sparsity levels by hit and trial. However, the sparsity level is not exactly known in practice. Thus, we performed extensive simulations and tuned the parameter setting for two Libraries $M_1$ and $M_5$ as in Table 3.2. We use these parameter setting in our following experiments for these libraries. For other libraries, we chose from these tables the one which results in a higher RSNR.

Figures 3.5 and 3.6 compare the RSNR and the PoS of these methods versus sparsity level for these six libraries. Our proposed methods outperform the other state-of-the-art methods in terms of RSNR. Precisely for libraries $M_2$ and $M_3$, our first three methods lead to higher RSNRs. For the USGS library, the next three methods have higher RSNRs than other methods. Moreover, the proposed methods outperform other methods in PoS values for all mentioned libraries. Our results reveal that the type of the library has significant impact on the RSNR and PoS values. We observed that the RSNR and PoS values are higher for the i.i.d. libraries (uniformly distributed library $M_2$ and the Gaussian distributed library $M_3$) than the USGS library $M_1$, specifically, for smaller sparsity levels. For the libraries with reduced dimensions, $M_4$, $M_5$ and $M_6$, we observe that the RSNR and the PoS curves exhibit a similar shape of variations versus the sparsity levels. Moreover the spectral libraries
Table 3.2: Parameter setting for our proposed methods (VNA\(i\); \(i = 1, \ldots, 6\)) and SUnSAL tuned via hit and trial for different sparsity levels at SNR = 25dB.

<table>
<thead>
<tr>
<th>Sparsity levels</th>
<th>SUnSAL</th>
<th>VNA1 (\lambda, \lambda, \sigma)</th>
<th>VNA2 (\lambda, \tau)</th>
<th>VNA3 (\lambda, \lambda, \sigma)</th>
<th>VNA4 (\lambda, \lambda, \sigma)</th>
<th>VNA5 (\lambda, \tau)</th>
<th>VNA6 (\lambda)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.05</td>
<td>0.005, 0.05</td>
<td>0.01, 0.15</td>
<td>0.003</td>
<td>0.1, 0.001</td>
<td>0.01, 0.3</td>
<td>0.05</td>
</tr>
<tr>
<td>2</td>
<td>0.005</td>
<td>0.001, 0.02</td>
<td>10^{-4}, 0.08</td>
<td>0.003</td>
<td>0.01, 0.001</td>
<td>0.01, 0.3</td>
<td>0.01</td>
</tr>
<tr>
<td>3</td>
<td>0.08</td>
<td>0.003, 0.01</td>
<td>10^{-4}, 0.02</td>
<td>8 \times 10^{-6}</td>
<td>0.008, 10^{-4}</td>
<td>8 \times 10^{-4}, 0.1</td>
<td>0.001</td>
</tr>
<tr>
<td>4</td>
<td>0.001</td>
<td>5 \times 10^{-6}, 0.001</td>
<td>10^{-4}, 0.1</td>
<td>8 \times 10^{-6}</td>
<td>0.008, 10^{-4}</td>
<td>8 \times 10^{-4}, 0.08</td>
<td>5 \times 10^{-6}</td>
</tr>
<tr>
<td>5</td>
<td>10^{-4}</td>
<td>10^{-6}, 0.001</td>
<td>5 \times 10^{-6}, 0.02</td>
<td>8 \times 10^{-6}</td>
<td>0.008, 2 \times 10^{-4}</td>
<td>8 \times 10^{-4}, 0.05</td>
<td>5 \times 10^{-6}</td>
</tr>
<tr>
<td>6</td>
<td>10^{-5}</td>
<td>10^{-6}, 0.001</td>
<td>5 \times 10^{-6}, 0.5</td>
<td>8 \times 10^{-6}</td>
<td>0.005, 5 \times 10^{-6}</td>
<td>10^{-4}, 0.1</td>
<td>5 \times 10^{-4}</td>
</tr>
<tr>
<td>7</td>
<td>5 \times 10^{-6}</td>
<td>10^{-6}, 0.002</td>
<td>10^{-3}, 0.05</td>
<td>8 \times 10^{-6}</td>
<td>0.05, 5 \times 10^{-6}</td>
<td>0.1, 10^{-4}</td>
<td>5 \times 10^{-6}</td>
</tr>
<tr>
<td>10</td>
<td>10^{-6}</td>
<td>10^{-6}, 0.002</td>
<td>10^{-3}, 0.01</td>
<td>8 \times 10^{-6}</td>
<td>0.05, 5 \times 10^{-6}</td>
<td>0.5, 10^{-6}</td>
<td>5 \times 10^{-6}</td>
</tr>
<tr>
<td>15</td>
<td>10^{-6}</td>
<td>10^{-6}, 8 \times 10^{-4}</td>
<td>10^{-6}, 0.01</td>
<td>10^{-6}</td>
<td>5, 10^{-6}</td>
<td>0.8, 10^{-6}</td>
<td>5 \times 10^{-6}</td>
</tr>
<tr>
<td>20</td>
<td>10^{-6}</td>
<td>10^{-6}, 8 \times 10^{-4}</td>
<td>10^{-6}, 0.01</td>
<td>10^{-6}</td>
<td>5, 10^{-6}</td>
<td>0.8, 10^{-6}</td>
<td>5 \times 10^{-6}</td>
</tr>
<tr>
<td>30</td>
<td>10^{-6}</td>
<td>10^{-6}, 8 \times 10^{-4}</td>
<td>10^{-6}, 0.01</td>
<td>10^{-6}</td>
<td>5, 10^{-6}</td>
<td>0.8, 10^{-6}</td>
<td>5 \times 10^{-6}</td>
</tr>
</tbody>
</table>

with reduced dimension exhibit enhanced RSNR and PoS. For \(\|x\|_0 = 1\), VNA6 lead to the highest RSNR values among all methods using \(M_1, M_4, M_5\) and \(M_6\) whereas VNA2 outperforms them all using other two libraries. The proposed methods have similar RSNR values for \(2 \leq \|x\|_0 \leq 7\) and outperform other state-of-the-art methods for this range of sparsity levels. For \(\|x\|_0 \geq 10\), VNA5 has the highest RSNR values for all types of USGS libraries except for \(M_5\). For the sparsity levels \(\|x\|_0 \geq 10\), our proposed methods have higher RSNR than other state-of-the-art methods using \(M_2\); where VNA3 has the best performance using \(M_3\). Figures 3.5 and 3.6 show that the proposed methods outperform other state-of-the-art methods in terms of RSNR and PoS in sparse environments.
3.2. SPARSE HYPERSPECTRAL UNMIXING VIA VARYING $\ell_p$-NORM APPROXIMATION OF $\ell_0$-NORM

Figure 3.5: The comparison of RNSR values between our proposed algorithms and the other state-of-the-art methods with respect to different sparsity levels $\|x\|_0$ for different libraries ($M_1, \cdots, M_6$).
3.2. SPARSE HYPERSPECTRAL UNMIXING VIA VARYING $\ell_p$-NORM APPROXIMATION OF $\ell_0$-NORM

Figure 3.6: The comparison of PoS values between our proposed algorithms and the other state-of-the-art methods with respect to different sparsity levels $\|x\|_0$ for different libraries $(M_1, \cdots, M_6)$.
To investigate the impact of the SNR on the RSNR, we use the similar methods as [16, 14] to construct the simulated hyperspectral data. We generate two simulated hyperspectral data with the size of 1000 pixels and 10000 pixels using the libraries $M_5$ and $M_6$, respectively. We set the sparsity level of the abundance vectors to five and randomly select the endmembers from $M_5$ and $M_6$ with the constraint of sum-to-one. The locations and the values of non-zero components are chosen randomly as well. For various SNRs, we have tuned the involved parameters in our proposed methods, CZ [30] method and SUnSAL [3, 14] via extensive simulations as reported in Table 3.3. For the rest of the state-of-the-art methods we used the same parameters in the previous experiment mentioned above. For these parameter setting, Figure 3.7 depicts the RSNR values versus the SNR for additive white and correlated Gaussian noise, respectively. We observe that all of proposed methods outperform the other state-of-the-art methods with the higher RSNR for the additive white gaussian noise for all range of SNRs except VNA6 for SNR values 45 and 50 dBs for $M_5$ and VNA3 for SNR values 45 and 50 dBs for $M_6$. For the colored noise, VNA5 and VNA6 recover best for all SNRs less than 40 dB and VNA5 becomes best at higher SNRs.

We should mention that the Lagrangian parameter $\lambda$ allows to trade-off between the sparsity of the solution and the quadratic error. However, there is no straightforward relation to find $\lambda$. Although an experimental relation is reported in [30] between $\lambda$ and SNR using the curve fitting, it does not give a suitable result with our algorithms. It needs further research on finding equations to set the parameters in our method as a function of the SNR and $p$ for best performance and we may take advantage of methods studied in [22] which is our future research.
Figure 3.7: The RSNR values obtained by different sparse unmixing methods versus SNRs for the simulated data with white- and correlated- noise over the USGS spectral library $M_5$ and $M_6$ for the sparsity level of 5.
### 3.2. SPARSE HYPERSPECTRAL UNMIXING VIA VARYING $\ell_p$-NORM APPROXIMATION OF $\ell_0$-NORM

Table 3.3: Parameter setting for SUnSAL algorithm, CZ method and our proposed methods (VNA$i(i = 1, \ldots, 6)$) for different SNRs of white Gaussian noise and colored noise (sparsity level = 5).

<table>
<thead>
<tr>
<th>SNR (dB)</th>
<th>SUnSAL [3, 14]</th>
<th>CZ [30]</th>
<th>VNA1</th>
<th>VNA2</th>
<th>VNA3</th>
<th>VNA4</th>
<th>VNA5</th>
<th>VNA6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\lambda$</td>
<td>$p$</td>
<td>$\lambda, \sigma$</td>
<td>$\lambda, \tau$</td>
<td>$\lambda$</td>
<td>$\lambda, \sigma$</td>
<td>$\lambda, \tau$</td>
<td>$\lambda$</td>
</tr>
<tr>
<td>20</td>
<td>0.001</td>
<td>0.5</td>
<td>$5 \times 10^{-3}, 0.005$</td>
<td>$10^{-3}, 0.02$</td>
<td>$10^{-3}$</td>
<td>0.03, $10^{-6}$</td>
<td>0.005, 0.01</td>
<td>$10^{-5}$</td>
</tr>
<tr>
<td>25</td>
<td>0.001</td>
<td>0.4</td>
<td>$3 \times 10^{-3}, 0.008$</td>
<td>$10^{-3}, 0.05$</td>
<td>$8 \times 10^{-4}$</td>
<td>0.03, $10^{-4}$</td>
<td>0.001, 0.08</td>
<td>$5 \times 10^{-4}$</td>
</tr>
<tr>
<td>30</td>
<td>$5 \times 10^{-3}$</td>
<td>0.3</td>
<td>$10^{-4}, 0.01$</td>
<td>$10^{-4}, 0.1$</td>
<td>$10^{-6}$</td>
<td>0.01, $10^{-4}$</td>
<td>0.001, 0.08</td>
<td>$5 \times 10^{-4}$</td>
</tr>
<tr>
<td>35</td>
<td>$10^{-4}$</td>
<td>0.35</td>
<td>$5 \times 10^{-4}, 0.01$</td>
<td>$10^{-4}, 0.05$</td>
<td>$10^{-6}$</td>
<td>0.01, $10^{-4}$</td>
<td>0.001, 0.05</td>
<td>$10^{-4}$</td>
</tr>
<tr>
<td>40</td>
<td>$5 \times 10^{-5}$</td>
<td>0.15</td>
<td>$10^{-5}, 0.005$</td>
<td>$5 \times 10^{-7}, 0.05$</td>
<td>$10^{-6}$</td>
<td>0.005, $10^{-4}$</td>
<td>0.001, 0.01</td>
<td>$5 \times 10^{-4}$</td>
</tr>
<tr>
<td>45</td>
<td>$10^{-5}$</td>
<td>0.1</td>
<td>$10^{-5}, 0.005$</td>
<td>$5 \times 10^{-6}, 0.05$</td>
<td>$10^{-6}$</td>
<td>0.001, $10^{-4}$</td>
<td>0.001, 0.01</td>
<td>$10^{-5}$</td>
</tr>
<tr>
<td>50</td>
<td>$5 \times 10^{-5}$</td>
<td>0.001</td>
<td>$10^{-5}, 0.005$</td>
<td>$10^{-6}, 0.05$</td>
<td>0.001</td>
<td>$5 \times 10^{-5}, 10^{-4}$</td>
<td>$5 \times 10^{-4}, 0.01$</td>
<td>$5 \times 10^{-6}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SNR (dB)</th>
<th>color noise</th>
<th>$\lambda$</th>
<th>$p$</th>
<th>$\lambda, \sigma$</th>
<th>$\lambda, \tau$</th>
<th>$\lambda$</th>
<th>$\lambda, \sigma$</th>
<th>$\lambda, \tau$</th>
<th>$\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>white noise</td>
<td>$\lambda$</td>
<td>$p$</td>
<td>$\lambda, \sigma$</td>
<td>$\lambda, \tau$</td>
<td>$\lambda$</td>
<td>$\lambda, \sigma$</td>
<td>$\lambda, \tau$</td>
<td>$\lambda$</td>
</tr>
<tr>
<td>20</td>
<td>0.04</td>
<td>0.3</td>
<td>$0.004, 0.01$</td>
<td>$10^{-3}, 0.05$</td>
<td>0.005</td>
<td>0.01, $5 \times 10^{-4}$</td>
<td>0.001, 0.1</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>$10^{-4}$</td>
<td>0.2</td>
<td>$0.004, 0.01$</td>
<td>$10^{-4}, 0.1$</td>
<td>0.005</td>
<td>0.01, $5 \times 10^{-4}$</td>
<td>0.001, 0.1</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>$5 \times 10^{-4}$</td>
<td>0.25</td>
<td>$0.001, 0.01$</td>
<td>$10^{-4}, 0.1$</td>
<td>0.001</td>
<td>0.01, $10^{-4}$</td>
<td>0.001, 0.08</td>
<td>$5 \times 10^{-4}$</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>$10^{-4}$</td>
<td>0.1</td>
<td>$5 \times 10^{-4}, 0.01$</td>
<td>$10^{-5}, 0.05$</td>
<td>$5 \times 10^{-4}$</td>
<td>0.01, $10^{-4}$</td>
<td>0.001, 0.05</td>
<td>$10^{-4}$</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>$5 \times 10^{-4}$</td>
<td>0.05</td>
<td>$10^{-5}, 0.01$</td>
<td>$10^{-5}, 0.1$</td>
<td>$10^{-5}$</td>
<td>0.005, $10^{-4}$</td>
<td>0.001, 0.01</td>
<td>$5 \times 10^{-4}$</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>$10^{-4}$</td>
<td>0.1</td>
<td>$10^{-5}, 0.005$</td>
<td>$10^{-5}, 0.05$</td>
<td>$5 \times 10^{-5}$</td>
<td>0.001, 0.001</td>
<td>0.001, 0.005</td>
<td>$10^{-4}$</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>$5 \times 10^{-4}$</td>
<td>0.05</td>
<td>$10^{-5}, 0.005$</td>
<td>$5 \times 10^{-5}, 0.01$</td>
<td>$10^{-4}$</td>
<td>$5 \times 10^{-5}, 10^{-4}$</td>
<td>$5 \times 10^{-4}, 0.01$</td>
<td>$5 \times 10^{-6}$</td>
<td></td>
</tr>
</tbody>
</table>

### Computational Complexity

The computational cost is an important issue for the unmixing process due to the large size of hyperspectral images and the large dimension of spectral libraries. The computation of Algorithm 1 requires about $(L + n + 2)n^2 + (L + 9)n$ operations per iteration where each operation means one multiplication and one addition. Algorithm 2 requires $(L + n + 1)n^2 + LN + 7n$ operations per iteration. We also measure the computational cost of the unmixing methods by their running times per pixel [3, 20, 14]. Table 3.4 compares the processing times in Seconds for different state-of-the-art algorithms with ours for three types of spectral libraries.
3.2. SPARSE HYPERSPECTRAL UNMIXING VIA VARYING $\ell_p$-NORM APPROXIMATION OF $\ell_0$-NORM

Table 3.4: Comparison of the measured processing time per pixel (in Seconds) for different unmixing algorithms over three different libraries on an Intel Core i7-2600 (at 3.5 GHz) and 8 GB of RAM Memory (SNR = 40 dB, Sparsity level = 5).

<table>
<thead>
<tr>
<th>method</th>
<th>Library $M_4$</th>
<th>Library $M_5$</th>
<th>Library $M_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BiICE [16]</td>
<td>10.746</td>
<td>4.636</td>
<td>0.476</td>
</tr>
<tr>
<td>SUnSAL [3, 14]</td>
<td>0.102</td>
<td>0.0517</td>
<td>0.0114</td>
</tr>
<tr>
<td>CZ [30]</td>
<td>4.014</td>
<td>1.999</td>
<td>0.0205</td>
</tr>
<tr>
<td>PLA [41]</td>
<td>0.546</td>
<td>0.374</td>
<td>0.101</td>
</tr>
<tr>
<td>VNA1</td>
<td>3.712</td>
<td>1.861</td>
<td>0.211</td>
</tr>
<tr>
<td>VNA2</td>
<td>3.661</td>
<td>2.001</td>
<td>0.212</td>
</tr>
<tr>
<td>VNA3</td>
<td>0.532</td>
<td>0.306</td>
<td>0.0371</td>
</tr>
<tr>
<td>VNA4</td>
<td>4.129</td>
<td>2.169</td>
<td>0.216</td>
</tr>
<tr>
<td>VNA5</td>
<td>4.059</td>
<td>2.098</td>
<td>0.237</td>
</tr>
<tr>
<td>VNA6</td>
<td>4.244</td>
<td>1.892</td>
<td>0.200</td>
</tr>
</tbody>
</table>

Experiments with real hyperspectral data

In our real data experiment, we use a subimage of the well known Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) hyperspectral image of Cuprite mining in Nevada used in different literatures for evaluation of unmixing methods [14, 30, 18]. Figure 3.8 shows the corresponding mineral map produced by a Tricorder 3.3 software product 2. Although the Tricorder map had been produced in 1995, the publicly collected AVIRIS Cuprite data has been available in 1997. Here, we just consider the mineral maps as an indicator to make a qualitative analysis of the fractional abundance maps derived from the different unmixing methods.

The scene consists of 224 spectral bands ranging from 400 nm to 2500 nm where the spectral sub-bands 1-2, 105-115, 150-170, and 223-224 were removed due to their low SNRs and the water-vapor absorption. Thereby, we used 188 remaining spectral bands of the subimage. To have a better impression for the AVIRIS cuprite hyperspectral data used in our experiments, we show two samples of the sub-bands of the
3.2. SPARSE HYPERSPECTRAL UNMIXING VIA VARYING \( \ell_p \)-NORM APPROXIMATION OF \( \ell_0 \)-NORM

Figure 3.8: USGS map showing the location of different minerals in the Cuprite mining district in Nevada. The map is available online at: http://speclab.cr.usgs.gov/cuprite95.tgif.2.2um_map.gif.
3.2. SPARSE HYPERSPECTRAL UNMIXING VIA VARYING \( \ell_p \)-NORM APPROXIMATION OF \( \ell_0 \)-NORM

scene in Figure 3.9.

![Figure 3.9: Bands 5 (a) and 40 (b) of the subimage of AVIRIS cuprite Nevada dataset.](image)

We employed the CLS [14], SUnSAL [3, 14], CZ method [30], and our six proposed methods using the USGS spectral library \( \mathbf{M}_5 \) to estimate the abundances. We use the same parameter setting as in Table 3.3 \( \text{SNR} = 30 \text{ dB} \) for our proposed methods and the other two state-of-the-art methods. In Figure 3.10, we have included the fractional abundances of five endmembers estimated by the mentioned algorithms. It is common to visually compare such results [20, 16, 30, 17]. The visual comparison between the fractional abundance maps is often considered as a qualitative comparison [15, 17, 16]. We observe that our proposed unmixing methods visually behave similar to the SUnSAL algorithm [3, 14] and the USGS Tricorder algorithm as reported in [20], see [20, 14] for more details.

The estimated fractional abundance of an endmember for each pixel is the output of the unmixing procedure shown by an image, e.g. Figure 3.10. The proportion of the contributions of each endmember is varied from the high contribution for a light pixel to the low one for a darker pixel [1, 16].
3.2. SPARSE HYPERSPECTRAL UNMIXING VIA VARYING $\ell_p$-NORM APPROXIMATION OF $\ell_0$-NORM

<table>
<thead>
<tr>
<th></th>
<th>Alunite</th>
<th>Buddingtonite</th>
<th>Kaolinite</th>
<th>Muscovite</th>
<th>Montmorillonite</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CLS</strong></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
</tr>
<tr>
<td><strong>SU4SAL</strong></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
</tr>
<tr>
<td><strong>CZ</strong></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
</tr>
<tr>
<td><strong>VNA1</strong></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
</tr>
<tr>
<td><strong>VNA2</strong></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
</tr>
<tr>
<td><strong>VNA3</strong></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
</tr>
<tr>
<td><strong>VNA4</strong></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
</tr>
<tr>
<td><strong>VNA5</strong></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
</tr>
<tr>
<td><strong>VNA6</strong></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
<td><img src="image" alt="" /></td>
</tr>
</tbody>
</table>

Figure 3.10: Fractional abundances estimated from different algorithms for five end-members appeared in the scene of AVIRIS Cuprite hyperspectral data.
3.2. SPARSE HYPERSPECTRAL UNMIXING VIA VARYING $\ell_p$-NORM APPROXIMATION OF $\ell_0$-NORM

3.2.6 Conclusion

We have investigated LSHU with the new approximation of the $\ell_0$-norm minimization problem. Our new proposed method updates the $\ell_p$-norm iteratively as an $\ell_0$-norm approximation. We took advantage of the Hausdorff distance to show the continuity for the set of solutions with respect to norm $p$ in order to result in tracking of better local minima in each iteration. Accordingly, we introduced three methods to update required parameters of $(\epsilon, p)$ based on the Gradient descent, Taylor expansion and Exponential equations. The ultimate minimization problem was formulated as the adaptive unconstrained $\ell_1$-norm and $\ell_2$-norm problems whose weights are updated properly. Using the ADMM algorithm provided an opportunity to solve the acquired optimization problem. Experiments on artificial data sets and a real hyperspectral image were used to compare the performance of our proposed methods VNAx ($x=1,2,\ldots,6$) with several state-of-the-art methods. The simulation results justified the improvements in terms of RSNR as well as PoS for both methods (for $\ell_p - \ell_1$-norm and $\ell_p - \ell_2$-norm problems).
Chapter 4

Semiblind Sparse Hyperspectral Unmixing

Through $\ell_0$-norm Approximation

4.1 Sparse Hyperspectral Unmixing via arctan Approximation of $\ell_0$-norm

In this section, we present a sparse unmixing procedure through the approximation of the $\ell_0$ norm as the sparsity regularization term. We employ a smooth function to approximate the $\ell_0$ norm along with the $\ell_2$ norm term as the error tolerance for the objective optimization problem. Although the most previous research works consider the sparse unmixing via the iterative methods [20, 94], the approximation of the $\ell_0$ norm problem using a smooth function can attain more accurate results than the mixed $\ell_1$ norm with the given error for LMM [31]. Hence, it motivates us to use a smoothed function to replace the $\ell_0$ norm term in order to provide better accuracy. We show that our approach help us to reconstruct the fractional abundances with higher RSNR as well.
4.1. SPARSE HYPERSPECTRAL UNMIXING VIA ARCTAN APPROXIMATION OF $\ell_0$-NORM

4.1.1 The proposed method

Substituting the $\ell_0$ norm term with the $\ell_1$ norm in a practical scenario may result in considerable inaccuracies and the solution using the $\ell_1$ norm may be far from that of using the $\ell_0$ norm. The $\ell_0$ norm is not a continuous function and has abrupt changes. This is the main reason that the problem in (2.6) is computationally complex. We alternatively approximate the $\ell_0$ norm with a $C^1$-smooth function\(^1\). The smoothness of the approximated problem allows iterative numerical to converge to acceptable solutions. Indeed, using such smooth function to approximate the $\ell_0$ norm term has some advantages [18]. First, it has a better phase transition than the greedy algorithms (e.g., the proposed smoothed $\ell_0$ norm method called SL0 algorithm in [31]) and requires considerably less computation time than the state-of-the-art linear program solvers [32]. Next, a smooth measure of sparsity can be provided by means of such definition of sparsity. Furthermore, such a representation could tolerate noise to some extend [31]. Finally, we can obtain more accurate results than the $\ell_1$ norm approach in the filed of sparse regression [32].

The desired smoothed function must satisfy the initial values as $f(0) = 0$ and $f(1) = 1$ as well as the continuity property to achieve the mentioned advantageous where $f(x)$ is a smooth function. The continuous gaussian function and logarithm-based function are two of them as introduced in [31, 32] and [18], respectively.

The logarithm-based approximation function has been used recently to approximate the $\ell_0$ norm term for the spectral unmixing application [18]. Indeed, an iterative weighted algorithm using the logarithm-based smoothed function was proposed in [18]. However, our objective is to find some other smoothed function to approximate

\(^1\)A differentiable function with a continuous derivative is called $C^1$-smooth.
4.1. SPARSE HYPERSPECTRAL UNMIXING VIA ARCTAN APPROXIMATION OF $\ell_0$-NORM

the $\ell_0$ norm term in more accurate manner to achieve the better reconstruction SNR defined later. Now, we propose the following smooth function

$$F(\delta, x) = \sum_{i=1}^{N} \frac{2}{\pi} \arctan \left( \frac{x_i}{\delta} \right),$$  \hspace{1cm} (4.1)$$

where $0 < \delta < 1$ and $0 \leq x_i \leq 1$.

It is easy to show that $F(\delta, x)$ converges to $||x||_0$ as $\delta$ tends to 0, i.e., $\lim_{\delta \to 0} F(\delta, x) = ||x||_0$.

By approximating the $\ell_0$ norm term with the smooth function, we can consider the following unconstrained version of the desired optimization problem

$$\min_{x} ||Mx - y||_2^2 + \lambda F(\delta, x), \text{ subject to ANC and ASC}$$  \hspace{1cm} (4.2)$$

where $\lambda > 0$ is the Lagrangian parameter.

Using the first order of Taylor expansion of the smooth function and simplifying the related terms, we have the following minimization problem

$$\min_{x} ||Mx - y||_2^2 + \lambda \sum_{i=1}^{N} w_i x_i, \text{ subject to ANC and ASC}$$  \hspace{1cm} (4.3)$$

where $w_i = \frac{2}{\pi} \frac{\delta}{\delta^2 + x_i^2}$ and $l$ shows the corresponding iteration number for the algorithm used to solve the minimization problem.

In our approach, we employ the ADMM algorithm [86, 82] that it has been recently used for solving a number of optimization problems for hyperspectral unmixing [3]. In the next section, we apply this method for the simulated hyperspectral data with a given spectral library.
4.1. SPARSE HYPERSPECTRAL UNMIXING VIA ARCTAN APPROXIMATION OF $\ell_0$-NORM

Table 4.1: The comparison results of RSNR values for different methods.

<table>
<thead>
<tr>
<th>SNR in dB</th>
<th>20</th>
<th>40</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCLS [104]</td>
<td>-6.92</td>
<td>5.32</td>
<td>16.92</td>
</tr>
<tr>
<td>L2-SL0 [18]</td>
<td>0.04</td>
<td>2.08</td>
<td>2.64</td>
</tr>
<tr>
<td>the proposed method</td>
<td>2.76</td>
<td>11.02</td>
<td>29.34</td>
</tr>
</tbody>
</table>

4.1.2 Simulation results

We examine the proposed method using simulated hyperspectral data from the USGS library [93]. We generate the simulated hyperspectral data cubes as the same approach used in [15]. The simulated data cube is produced using a linear mixture model with five randomly selected spectral signatures from the USGS library. The library contains 240 materials for 224 spectral bands and distributed uniformly in the interval 0.4-2.5 $\mu$m. The produced data cube has $15 \times 15$ pixels and 224 spectral bands per pixel and we impose the ASC condition for each simulated pixel as well.

Figure 4.1 displays the original generated data cube and the corresponding reconstructed pixels using the nonnegative constrained least squares (NCLS) method [104], the $\ell_0$ norm approximation method introduced in [18] and our proposed method. We use the RSNR metric defined in (3.9) to evaluate the reconstruction quality of a spectral mixture. Table 4.1 represents the RSNR values for the proposed method and the other mentioned methods. The proposed method exhibits the highest RSNR in dBs in comparison with the NCLS method and the proposed method of [18]. In particular, more improvement can be achieved at higher SNRs.

4.1.3 Conclusion

We proposed a new approximation for the problems involving the $\ell_0$ norm minimization and applied it for the hyperspectral unmixing. We introduced a method of
4.1. SPARSE HYPERSPECTRAL UNMIXING VIA ARCTAN
APPROXIMATION OF $\ell_0$-NORM

SNR=20dB  SNR=40dB  SNR=60dB

Original data cube

NCLS [104]

Method in [18]

Our proposed method

Figure 4.1: The original data pixels and the corresponding estimated maps by different methods for three SNRs.
unmixing for the linear mixture model based on the smooth function. The method is able to solve the sparse optimization problem and gave enhanced performance in the HSI process compared with the other state-of-the-art $\ell_0$ norm approximation method such as the proposed method of [18]. To evaluate the performance, the reconstruction SNRs have been computed for our method and the other methods over the USGS library. The simulation results indicated that our proposed method outperforms other methods. More experiments over real hyperspectral data sets can be performed to criticize the proposed method.
4.2 $\ell_0$-Norm Sparse Hyperspectral Unmixing Using Arctan Smoothing

In recent years, several approximation methods have been proposed for the $\ell_0$-norm minimization problem notwithstanding various unmixing methods, which employed $\ell_1$-norm instead of $\ell_0$-norm (e.g., [14, 20, 17, 3]). These may include iterative reweighted schemes (e.g., [26, 27]), greedy algorithms [28, 29], Bayesian learning algorithms [16], $\ell_q$ regularization [105] and compressive sensing schemes [19, 32]. Each of these methods has specific characteristics, e.g., the method proposed in [16] exploits Bayesian learning to control the parameters involved. Some algorithms have used better approximations of the $\ell_0$-norm, e.g., the $\ell_p$-norm is approximated as a weighted $\ell_2$-norm in [30]. Although these methods improve the sparsity, the $\ell_p$-norm function is not Lipschitz continuous for $p < 1$. As a result, these methods suffer from numerical problems for smaller values of $p$. Thus, an attractive solution is to employ Lipschitz continuous approximations, such as the exponential function, the logarithm function or sigmoid functions, e.g., [31, 18, 33]. The arctan function is also used in different literature works for sparse regularization such as approximating the sign function appearing in the derivative of the $\ell_1$-norm term in [81], introducing a penalty function for the sparse signal estimation by the maximally-sparse convex approach in [106] and approximating the $\ell_0$-norm term through a weighted $\ell_1$-norm term in [33].

In this section, we propose a new algorithm utilizing an arctan function, which allows us to start our search with the $\ell_1$-norm problem, which is convex and initially guarantees fast convergence to the unique optimal solution. This method allows us to iteratively update our problem to better approximate the $\ell_1$-norm problem and provides an enhanced separation of the zero components. The proposed arctan sum is a smooth approximation of the $\ell_0$-norm and $\ell_1$-norm as a function of $\sigma$. We
gradually increase the parameter $\sigma$ in order to allow the convergence and tracking of the best local optimal solution and iteratively find a better sparse solution. The arctan function is Lipschitz continuous; thus, the proposed method does not have additional considerations to avoid numerical problems, e.g., [25, 40, 41]. Moreover, our proposed algorithm improves the sparsity as $\sigma$ varies from zero to $\infty$, whereas in [18, 33], the value of $\sigma$ is constant. We use the ADMM approach to minimize the resulting objective function at each iteration [86, 82]. We prove that the set of local optima of our objective function is continuous with the Hausdorff metric versus $\sigma$. This implies that iterative minimization along with a gradual increase of $\sigma$ guarantees the convergence to the optimal solution. Finding the appropriate increasing sequence for $\sigma$ is an open problem to guarantee this convergence and to reduce the number of iterations. Thus, we simply propose to increase $\sigma$ exponentially. We compare our proposed method to several state-of-the-art methods [3, 14, 16, 30] over both the synthetic data and real hyperspectral data.

4.2.1 Our proposed unmixing method: Arctan approximation of the $\ell_1$- and $\ell_0$-norms

We propose the following function to approximate the $\ell_1$ or $\ell_0$-norms

$$F(\sigma, x) = g(\sigma) \sum_{i=1}^{N} \arctan(\sigma x_i), \quad (4.4)$$

where $\sigma > 0$ is a tunable parameter and $0 \leq x_i \leq 1$. We find an appropriate function for $g(\sigma)$, such that $F(\sigma, x)$ converges to the $\ell_1$ and $\ell_0$-norms, respectively, as $\sigma$ tends to zero and $\infty$. The basic idea behind this concept is to start at $\sigma = 0$ for which our problem becomes the $\ell_1$-norm problem in equation (2.7). Thus, the problem
becomes a convex optimization for $\sigma = 0$ that is known to be a good approximation of equation (3.1), e.g. [107]. By iteratively increasing $\sigma$, the proposed problem-minimizing equation (4.4) tends to the problem in equation (3.1).

**Remark 1.** We shall choose $g(\sigma)$, such that the following conditions are satisfied:

(i) $F(\sigma, x)$ tends to $||x||_1$ as $\sigma$ tends to zero.

(ii) $F(\sigma, x)$ tends to $||x||_0$ as $\sigma$ tends to $\infty$.

There are many such functions that satisfy the above conditions, such as follows

\[
g_1(\sigma) = \frac{2}{\pi} + \frac{1}{\sigma}, \quad (4.5a)
\]

\[
g_2(\sigma) = \frac{1}{\text{arctan}(\sigma)} \quad (4.5b)
\]

where $\sigma > 0$.

Figure 4.2 shows the curves of functions $\frac{\text{arctan}(\sigma x)}{\text{arctan}(\sigma)}$ and $x^p$ for $x \in [0, 1]$ and several different values of $\sigma$ and $p$. We observe that for $p = 1$ and $\sigma \to 0$, these functions become linear, and both yield the $\ell_1$-norm. As $p \to 0$ and $\sigma \to \infty$, these functions tend to the unit step function and both yield the $\ell_0$-norm. For values of $p$ between one and zero and $\sigma$ between zero and $\infty$, we observe that these curves are similar and that they can approximate each other. However, the important difference between these functions is in their derivatives for small values of $x$ around zero; in contrast to $\frac{\text{arctan}(\sigma x)}{\text{arctan}(\sigma)}$, the derivatives of $x^p$ are not bounded around $x = 0$. These unbounded derivatives cause numerical instabilities in iterative algorithms.
Figure 4.2: Comparison of \(\arctan(\sigma x)/\arctan(\sigma)\) with \(x^p\) for different values of \(\sigma\) and \(p\).

The approximation of the \(\ell_0\)-norm problem using the function \(F(\sigma, x)\) for the constrained \(\ell_0\)-norm problem can be considered as follows

\[
\min_{x \in S} F(\sigma, x) \quad \text{subject to} \quad \|y - Mx\|_2^2 \leq \epsilon, \tag{4.6}
\]

as \(\sigma\) increases. The unconstrained version of equation (4.6) using the Lagrangian method can be presented by

\[
\min_{x \in S} f(x, \sigma), \tag{4.7a}
\]

\[
f(x, \sigma) = \frac{1}{2}\|y - Mx\|_2^2 + \lambda g(\sigma) \sum_{i=1}^N \arctan(\sigma x_i), \tag{4.7b}
\]

where there exists some \(\lambda > 0\), such that equations (4.6) and (4.7a) are equivalent.

Now, we prove the continuity of the set of candidate local minima of equation (4.7a) with respect to the parameter \(\sigma\) to guarantee that our proposed method
reaches the possible sparse solution (i.e., if it exists) while \( \sigma \) is varying. The motivation behind Theorem 1 is to give insight to the solution obtained using the previous value of \( \sigma \) as a good initialization for the next iteration with the larger value of \( \sigma \).

Using the definition of Hausdorff distance mentioned in (3.13), the following theorem proves the desired continuity of the set of all candidate local minima.

**Theorem 2.** Let \( X_\sigma \subset \mathcal{X} \) be the set of all solutions of:

\[
\nabla_x f(x, \sigma) = \lambda v(x, \sigma) + M^T M x - M^T y = 0,
\]

where \( v(x, \sigma) = \left[ \frac{\sigma g(\sigma)}{1 + \sigma^2 x_1^2}, \ldots, \frac{\sigma g(\sigma)}{1 + \sigma^2 x_N^2} \right]^T \). Then, \( X_\sigma \) is a continuous function of \( \sigma \in [0, \infty) \).

**Proof.** See Appendix B.1.

For simplicity, the above theorem is written for the simplified case where \( S \) is relaxed into \( \mathbb{R}^N \). However, the proof in Appendix B.1 includes the ANC, as well as the ASC. For \( \sigma \to 0 \), the problem equation (4.7a) is indeed a kind of \( \ell_1 \)-norm problem, which is convex, and thus, \( X_\sigma \) has a unique solution provided that \( M \) has the full rank property. We should point out that the strong condition of restricted isometry property of \( M \) leads to have \( \ell_0 \)-norm solution at the initial step of our proposed method automatically [108, 40], i.e., our proposed method will solve an equivalent \( \ell_1 \)-norm problem at the beginning when \( \sigma \to 0 \).

The continuity of \( X_\sigma \) versus \( \sigma \) implies that there is a neighborhood around \( \sigma = 0 \) for which \( X_\sigma \) still has a unique member. Thus, we could increase \( \sigma \) within this neighborhood. As \( \sigma \) further increases, the number of local minima (i.e., \( |X_\sigma| \)) may increase by splitting the members, i.e., bifurcation might happen. Our algorithm
tracks only one member of $\mathbf{X}_\sigma$ as the solution, which has a lower value for $f(\mathbf{x}, \sigma)$. As $\sigma$ increases, we anticipate obtaining a sparser solution. Appropriate increment values for the sequence of $\sigma$ allow one to track the best local optima. Aggressive increasing of $\sigma$ in each iteration may result in missing the tracking of the best local optima, which translates into some performance loss. On the other hand, conservatively increasing $\sigma$ results in additional computational cost. Optimal selection of the increasing sequence of values for $\sigma$ is the focus of our future research and remains an open challenging problem, since this sequence must avoid missing the best minima in each iteration.

Now, we propose to update $\sigma$ iteratively as follows:

$$
\sigma^{(j+1)} = \sigma^{(j)} \exp(\alpha), j = 1, \ldots, I_{\text{max}}
$$

(4.9)

where $\sigma^{(j)}$ is exponentially increasing versus the iteration index $j$, $I_{\text{max}}$ is the maximum number of iterations, $\sigma^{(1)}$ is a small initial value and $\alpha$ is the increasing rate.

The values for $\sigma^{(1)}$ and $\alpha$ are selected via hit and trial using extensive simulations. To choose the initial value for $\sigma^{(1)}$, we first set the value of $\alpha$ equal to zero. Then, we gradually increase the value of $\sigma^{(1)}$ from zero up to the largest value, such that the behavior of the algorithm remains the same as for $\sigma = 0$ (the $\ell_1$-norm problem). Indeed, we propose to choose $\sigma^{(1)}$ as the largest value for which the problem behaves similarly to the $\ell_1$-norm problem in terms of their RSNR, as defined in equation (3.9).

The problem in equation (4.7a) is an approximation of the original $\ell_0$-norm problem under the ANC and ASC constraints, i.e. $\mathbf{x} \in \mathbb{S}$. The unconstrained Lagrangian
of equation (4.7a) can be also rewritten as:

$$\min_{x} \frac{1}{2} \|y - Mx\|_2^2 + \lambda g(\sigma) \sum_{i=1}^{N} \arctan(\sigma x_i) + \iota_{\{1\}}(1^T x) + \iota_{\mathbb{R}^N_{+}}(x),$$

(4.10)

where $1$ is the column vector of ones and $\iota_{Q}(x)$ is the indicator function, either zero or $\infty$ if $x \in Q$ or $x \not\in Q$, respectively.

We use the ADMM approach [86, 82] to solve equation (4.10). By constructing the augmented Lagrangian multipliers and assigning $f_1(x) = \frac{1}{2} \|y - Mx\|_2^2 + \iota_{\{1\}}(1^T x)$, the primary minimization problem is:

$$\arg \min_{x} \frac{1}{2} \|y - Mx\|_2^2 + \iota_{\{1\}}(1^T x) + \mu \frac{1}{2} \|x - z^{(j)} - u^{(j)}\|_2^2.$$ 

(4.11)

The solution of the above is updated by:

$$x^{(j+1)} \leftarrow A^{-1} B - A^{-1} 1(1^T A^{-1} 1)^{-1} (1^T A^{-1} B - 1),$$

(4.12a)

where $A$ and $B$ are first calculated as follows:

$$A \leftarrow M^T M + \mu I,$$

(4.12b)

$$B \leftarrow M^T y + \mu (z^{(j)} - u^{(j)}),$$

(4.12c)

and $z^{(j)}$ represents the value of vector $z$ at the $j$-th iteration.

By assigning the remaining terms of equation (4.10) to $f_2(z)$, i.e., $\lambda g(\sigma) \sum_{i=1}^{N} \arctan(\sigma z_i) +$
4.2. $\ell_0$-NORM SPARSE HYPERSPECTRAL UNMIXING USING ARCTAN SMOOTHING

To find the updating equation for $z$, we take the derivative of equation (4.13) with respect to $z$ and set it to zero, which leads to following equations:

$$z_i = x_i^{(j+1)} - u_i^{(j)} - \frac{\lambda \sigma g(\sigma)}{\mu(1 + \sigma^2 z_i^2)}, \quad (4.14)$$

where $z = [z_1, \cdots, z_N]^T$. We are interested in the positive root of these polynomial equations in equation (4.14) of degree three, which can be computed numerically. However, to reduce the computational cost, we propose to approximate the last term, $\frac{\lambda \sigma g(\sigma)}{\mu(1 + \sigma^2 z_i^2)}$, with its value from the previous iteration, which leads to the following update equation:

$$z^{(j+1)}_i \leftarrow (x^{(j+1)}_i - u^{(j)}_i - \frac{\lambda \sigma g(\sigma)}{\mu(1 + \sigma^2 z^{(j)}_i)})^+ \quad (4.15)$$

where $a^+ = \max(a, 0)$ and $z^{(j)^2}$ denotes the vector of the squared of elements of $z^{(j)}$, and the division is an element-wise operation, i.e., the division of elements of two vectors or a scalar divided by elements of a vector.

To prove the convergence of equation (4.15), we define the function $\theta(z) = x_i^{(j+1)} - u_i^{(j)} - \frac{\lambda \sigma g(\sigma)}{\mu(1 + \sigma^2 z_i^2)}$. It is easy to show that $\theta(z)$ is a contraction mapping for $z > 0$ and $\lambda \sigma^2 g(\sigma) < 2\mu$. Thus, by virtue of the fixed point theorem for contraction mapping functions, the convergence of $z_i^{(j+1)} = \theta(z_i^{(j)})$ to the optimal solution is guaranteed under the sufficient (not necessary) condition $\lambda \sigma^2 g(\sigma) < 2\mu$. This sufficient condition
is not imposed in our simulation.

Now, the pseudocode of the proposed algorithm can be considered as follows.

Algorithm 6 Pseudocode of the proposed method.

1: Initialize \( j = 1 \), and choose \( z^{(1)}, u^{(1)}, \mu > 0, \lambda > 0 \).
2: \textbf{while} \( j < I_{\text{max}} \) and \( \left( \min \left\{ \| x^{(j)} - z^{(j)} \|_2, \mu \| z^{(j)} - z^{(10 \lfloor j/10 \rfloor)} \|_2 \right\} > 10^{-4} \) \textbf{do}
3: \textbf{end while}
4: \textbf{end while}

4.2.2 Updating the regularized parameter \( \lambda \)

The Lagrangian parameter \( \lambda \) weights the sparsity term \( F(\sigma, x) \) in combination with the squared errors \( \| y - Mx \|_2^2 \) produced by the estimated fractional abundances. The expression in equation (4.7b) or equation (4.10) reveals that the larger values of the Lagrange multiplier lead to the sparser solutions. Moreover, the smaller \( \lambda \) leads to the smaller squared error. Hence, the parameter \( \lambda \) must be chosen to trade-off between the sparsity and the smaller squared error.

In our evaluations, we have first simulated the algorithms using several constant values for \( \lambda \) and chosen the value of \( \lambda \), which leads to the highest RSNR defined in equation (3.9). Hereafter, we refer to the proposed algorithm using a constant \( \lambda \) and equation (4.15) as the smoothing arctan (SA1) algorithm.

The drawback of using a constant value for \( \lambda \) is that it requires \textit{a priori} knowledge or simulations to adjust \( \lambda \) for each environment and signal-to-noise ratio. As an alternative, following the expectation-maximization (EM) approach in [109], we
propose to update $\lambda$ as follows:

$$
\lambda \leftarrow \frac{1}{L} \| y - Mx \|_2^2 + \frac{\lambda}{L} \sum_{k=1}^{L} \left( \frac{d_k^2}{\lambda + d_k^2} \right)
$$

(4.16)

where $Mx = [d_1, \cdots, d_L]^T$. Hereafter, we refer to this unmixing method as SA2.

There are several updating rules for the regularizing parameters, e.g. [110, 111, 112, 113] and we have examined three existing methods for updating $\lambda$, which have been proposed for other similar optimization problems, i.e., the L-curve method [110], the normalized cumulative periodogram (NCP) method [111] and the generalized cross-validation (GCV) method [112]. Our performance evaluations of our proposed algorithm revealed that the GCV updating rules for $\lambda$ result in the best performance amongst these methods in terms of RSNR. Hereafter, we refer to this combination as SA3.

The convergence

The ADMM is a powerful recursive numerical algorithm for various optimization problems [82]. In this paper, we employ this method for solving the minimization problem in equation (4.7a). If the conditions of Theorem 1 of [86] are met, the convergence of the ADMM is guaranteed. However, $f(x, \sigma)$ in the objective function of equation (4.7a) is not convex for all $\sigma$, and for these non-convex problems, the ADMM may converge to suboptimal/non-optimal solutions depending on the initial values ([82], page 73). Note that the primary minimization problem in equation (4.11) is always convex and, hence, leads to a converging solution to its optimum. In contrast, the secondary minimization problem in equation (4.13) is not convex for all $\sigma$. As we discussed earlier, it is easy to show that this term is convex for some small values of
4.2. \( \ell_0 \)-NORM SPARSE HYPERSONTAL UNMIXING USING ARCTAN SMOOTHING

\( \sigma \) and is not for large values.

The problem equation (4.13) is convex if its Hessian is non-negative, i.e., \( \mu I - 2\lambda g(\sigma)\text{diag}[\frac{\sigma^3 z_1}{(1+\sigma^2 z_1^2)^2}, \cdots, \frac{\sigma^3 z_q}{(1+\sigma^2 z_q^2)^2}] > 0 \). This means that for equation (4.13) to be convex, it is sufficient that \((1 + \sigma^2 z_i^2)^2 \geq 2\frac{\lambda}{\mu} \sigma^3 g(\sigma) z_i \) for all \( i \), which guarantees the convergence of the proposed algorithm. Since \( z_i \in [0,1] \), the condition \( 2\frac{\lambda}{\mu} \sigma^3 g(\sigma) \leq \inf_{z \in [0,1]} \frac{(1+\sigma^2 z^2)^2}{z} \) is sufficient for equation (4.13) to be convex and guarantees the convergence of the proposed algorithm to its optimal solution.

The upper bound for which \( \sigma \) leads to the convergence of our algorithm can be obtained by finding the maximum value of the RHS of the sufficient condition. Hence, it can be simplified to \( \max(\frac{9}{16\sqrt{3}} \sigma^2 g(\sigma), \frac{\sigma^3 g(\sigma)}{(1+\sigma^2)^2}) \leq 0.5\frac{\mu}{\lambda} \). Thus, given \( \frac{\mu}{\lambda} \), this condition easily gives us the largest value of \( \sigma \) for which our algorithm converges to its unique optimal solution. As the value of \( \sigma \) increases beyond this condition, the objective function in equation (4.7a) will have multiple local optima. Our numerical method attempts to track the best one on the basis that the set of local optima is continuous versus \( \sigma \).

Within initial iterations, \((z, x)\) will be around the unique optimal solution. We expect \( z \) to be sparse, i.e., most of its elements are close to zero. Thus, the corresponding diagonal elements of the Hessian matrix, i.e., \( \mu - 2\lambda g(\sigma)\frac{\sigma^3 z_i}{(1+\sigma^2 z_i^2)^2} \), will be close to \( \mu \), which is non-negative. In the next iterations, we gradually increase \( \sigma \) allowing equation (4.13) to become non-convex and locally track a sparser solution as \( \sigma \) increases.
4.2. $\ell_0$-NORM SPARSE HYPERSPECTRAL UNMIXING USING ARCTAN SMOOTHING

4.2.3 Experimental results and analysis

Here, we first evaluate our proposed algorithms SA1, SA2 and SA3, via different simulations. For our experiments, we take advantage of the USGS library [93] having 224 spectral bands in the interval 0.4 to 2.5 $\mu$m. For convenience, in simulations following [14, 16, 15, 30], we choose a subset of 240 spectral signatures of minerals from the original spectral signatures similar to [14], i.e., we discard the vectors of spectral signatures of materials that the angle between all remaining pairs is greater than $4.44^\circ$. This selection allows us to compare the results to [14, 16, 15, 30]. This library has similar properties with the original one, i.e., it has a very close mutual coherence value to the original library, which contains 498 spectral signatures of the endmembers. The mutual coherence (MC) is defined by

$$MC(M) = \max_{1 \leq i, j \leq N, i \neq j} \frac{|m_i^T m_j|}{||m_i||_2 ||m_j||_2},$$

(4.17)

where $m_i$ is the $i$-th column of $M$. We have also generated two additional libraries based on the uniform and Gaussian distributions. The examined libraries are:

1. $M_{\text{Original}} \in \mathbb{R}^{224 \times 498}$ is obtained from the USGS library [93] by selecting the spectral library, which contains 498 spectral signatures of minerals with 224 spectral bands with the MC of 0.999 in the same way as in [14].

2. $M_{\text{Prune}} \in \mathbb{R}^{224 \times 240}$ is a selected subset of $M_{\text{Original}}$, such that the angle between its columns is larger than $4.44^\circ$, and its MC is 0.996.

3. $M_{\text{Unif}} \in \mathbb{R}^{224 \times 240}$ is randomly generated with i.i.d. components uniformly distributed in the interval [0,1], and its MC is 0.823.
4.2. $\ell_0$-NORM SPARSE HYPERSPECTRAL UNMIXING USING ARCTAN SMOOTHING

$M_{\text{Gauss}} \in \mathbb{R}^{224 \times 240}$ is randomly generated with i.i.d. zero-mean Gaussian components with the variance of one, and its MC is 0.278.

We compare our proposed methods SA1, SA2 and SA3 to several existing state-of-the-art methods, including the nonnegative constrained least square (NCLS) [14], the SUnSAL algorithm [3, 14], the novel hierarchical Bayesian approach (BiICE (Bayesian inference iterative conditional expectations) algorithm [16]) and the method based on the $\ell_p - \ell_2$ minimization problem proposed in [30], we call it CZ method.

It should be noted that we report the experimental results only for $g(\sigma) = g_2(\sigma)$. In fact, one approach is to let the ADMM converge for a given $\sigma$ and, upon convergence, update $\sigma$. However, our experiments reveal that gradual updating of $\sigma$ in Step 3 of Algorithm 2 during the iteration of the ADMM leads to a significantly faster convergence. The expressions of the algorithm using equation (4.5a) or equation (4.5b) can be derived in a similar way, and our extensive simulation results show that using equation (4.5b) for $g(\sigma)$, the algorithm slightly outperforms the one using equation (4.5a). Thus, the experimental results are given for $g(\sigma) = g_2(\sigma)$. Finally, we must mention that we initialize $x^{(1)} = z^{(1)} = [\frac{1}{N}, \cdots, \frac{1}{N}]^T$ and $u^{(1)} = [0, \cdots, 0]^T$. This uniform initialization gives equal chance to all elements of the primary and secondary minimization problems to converge their optimal values.

Experiments with synthetic data

In the first experiment, we generate the fractional abundances for vector $x$ randomly with the Dirichlet distribution [114, 43] by generating independent and uniformly-distributed random variables and dividing their logarithms by the minus of sum of their logarithms. These vectors have different sparsity levels ranging from one to 10
that are compatible in practice for the mixed pixels, e.g., [12]. We generate 2500 data randomly for each sparsity level between one and 10. For each data sample, we first randomly select the location of nonzero abundances and generate the nonzero abundances following the Dirichlet distribution mentioned above. Then, we add the white Gaussian noise (AWGN) at different signal-to-noise ratios (SNRs), 15 dB (low SNR), 30 dB (medium SNR) and 50 dB (high SNR).

We generate 100 randomly-fractional abundances with the Dirichlet distribution for different types of libraries, while the sparsity levels is set to four. We should mention that the values of fractional abundances are varied during this experiment because of the consistency of the results for the experiment. The SNR is also set to 30 dB. We compare the performance of these unmixing methods using two criteria, the RSNR and the probability of success (PoS) defined by (3.9) and (3.28), respectively.

In our experiments, we select the threshold value $\xi = 0.316$ following the experimental approach in [14, 17]. We have chosen the parameters of these state-of-the-art methods either as they are reported in their proposed literature works or have adjusted them within the source code provided by the authors by trial and error for the best performance as follows:

- **SUnSAL [3, 14]:** maximum iteration = 200, $\lambda = 5 \times 10^{-2}$ for lower SNRs and $\lambda = 10^{-4}$ for higher SNRs.

- **NCLS:** only ANC is applied in the SUnSAL method, and set $\lambda = 0$ in [14].

- **BiICE [16]:** MaxIter = 50 and $a_{\text{Vita}} = b_{\text{Vita}} = a_{\lambda} = b_{\lambda} = 10^{-6}$.

- **CZ [30]:** $p = 0.2$ and $\log_{10} \lambda = 0.0008 \text{SNR}^2 - 0.1144 \text{SNR} - 0.9983$.

- **SA1, SA2, SA3:** $I_{\text{max}} = 100$, $\sigma^{(1)} = 0.1$, $\alpha = 0.07$. 
• SA1: $\lambda = 10^{-2}$.

Figure 4.3 shows the RSNR values and the corresponding PoS values for these methods versus different sparsity levels. Our proposed methods outperform the other state-of-the-art methods specifically for very sparse conditions in terms of RSNR values. Moreover, the PoS values of our proposed methods are superior to other methods, except for the SUnSAL algorithm. Besides, the results reveal that our third proposed method gives the best performance amongst our three methods for both RSNR and PoS values. Moreover, it is obvious that the values of RSNR and PoS are decreasing and increasing by raising the number of nonzero components and SNRs, respectively.

In the second experiment, we evaluate the impact of the SNR on the reconstruction quality of these methods for three sparsity levels, non-mixed (pure) pixels, for pixels with three and five nonzero elements, as illustrated in Figure 4.4. Again, we produce the fractional abundances based on the Dirichlet distribution for different ranges of SNRs from 10 dB to 50 dB. Similar to the first experiment, we only set the sparsity level to the desired values and their locations are chosen randomly. Then, we generate 2500 sample data and add the AWGN noise. For the pure pixel, our second proposed method outperforms the other state-of-the-art methods, as well as two other methods in terms of reconstruction errors. For the mixed pixels, SA1 and SA2 have the highest RSNRs from the low SNR (e.g., 10 dB) to the medium SNR (e.g., 30 dB). However, SA3 outperforms the other methods for an SNR greater than 30 dB. Furthermore, we have similar performances for the PoS exclusive of the SUnSAL method. Note that we may enhance the PoS curves by increasing the threshold $\xi$.

In the third experiment, we investigate the effect of the mutual coherence of
4.2. \( \ell_0 \)-NORM SPARSE HYPERSONTRAL UNMIXING USING ARCTAN SMOOTHING

Figure 4.3: The comparison of RSNR values and their corresponding probability of success (PoS) between our proposed methods and the other state-of-the-art methods with respect to different sparsity levels using \( \mathbf{M}_{\text{Prune}} \). (a) SNR = 15 dB; (b) SNR = 30 dB; (c) SNR = 50 dB.
4.2. $\ell_0$-NORM SPARSE HYPERSPECTRAL UNMIXING USING ARCTAN SMOOTHING

Figure 4.4: The RSNR values obtained by different sparse unmixing methods versus SNRs for the simulated data. (a) Sparsity level = 1; (b) sparsity level = 3; (c) sparsity level = 5.
the employed library (e.g., the type of library), as well as the number of available spectral signatures of endmembers (e.g., the size of the library) for the unmixing methods. Similar to the previous experiments, we generate 1000 randomly-fractional abundances with the Dirichlet distribution for different types of libraries, while the sparsity levels is set to four. The locations of these four abundances are selected at random. The SNR is also set to a medium value of 30 dB following [14]. Then, we compute the RSNR and the corresponding PoS values for different unmixing methods. Figure 4.5 depicts these results. They reveal that our proposed methods outperform the other state-of-the-art methods for different types of libraries in the sense of RSNRs. Indeed, all of three proposed methods outperform the other state-of-the-art methods; specifically, our third proposed method, i.e. SA3, has the best performance for the recovered fractional abundances compared to the other methods. For the PoS values, we have the same trend, except for the SUnSAL method. It is obvious that the library with the lower MC values results in the higher RSNR values. Moreover, we can observe that our second proposed method has better reconstruction error in comparison to the other state-of-the-art methods while the noise is colored. It also has very similar performance of the success for reconstruction with the SUnSAL algorithm in this experiment. Finally, the last bar chart shows that the values of RSNR and PoS for all unmixing methods have higher values by assuming the colored noise compared to the white Gaussian noise over \( \textbf{M}_{\text{Prune}} \).
4.2. $\ell_0$-NORM SPARSE HYPERSONTICAL UNMIXING USING ARCTAN SMOOTHING

To evaluate the impact of the noise type on these methods, we generated a colored noise following [14]. In this experiment, the colored noise is the output of a low pass filtering with a cut-off frequency of $5\pi / L$ where the input is generated as an independent and identically distributed (i.i.d.) Gaussian noise. We observe that the unmixing

Figure 4.5: The impact of library properties and colored noise over (a) RSNR and (b) PoS values obtained by different sparse unmixing methods when $||x||_0 = 4$ and SNR = 30dB.
4.2. $\ell_0$-NORM SPARSE HYPERSPECTRAL UNMIXING USING ARCTAN SMOOTHING

Table 4.2: The processing time of different algorithms per pixel (in seconds) using 4 different libraries and an i7-2600-3.5-GHz Intel Core processor with 8 GB of RAM memory.

<table>
<thead>
<tr>
<th>Method</th>
<th>$M_{\text{Original}}$</th>
<th>$M_{\text{Prune}}$</th>
<th>$M_{\text{Unif}}$</th>
<th>$M_{\text{Gauss}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCLS [14]</td>
<td>0.39</td>
<td>0.062</td>
<td>0.091</td>
<td>0.094</td>
</tr>
<tr>
<td>SUnSAL [3, 14]</td>
<td>0.44</td>
<td>0.066</td>
<td>0.11</td>
<td>0.10</td>
</tr>
<tr>
<td>BiICE [16]</td>
<td>60.48</td>
<td>4.53</td>
<td>5.06</td>
<td>4.61</td>
</tr>
<tr>
<td>CZ [30]</td>
<td>34.12</td>
<td>6.91</td>
<td>8.32</td>
<td>8.98</td>
</tr>
<tr>
<td>SA1</td>
<td>3.59</td>
<td>0.54</td>
<td>0.59</td>
<td>0.55</td>
</tr>
<tr>
<td>SA2</td>
<td>4.19</td>
<td>0.81</td>
<td>0.88</td>
<td>0.83</td>
</tr>
<tr>
<td>SA3</td>
<td>5.24</td>
<td>1.77</td>
<td>1.86</td>
<td>1.81</td>
</tr>
</tbody>
</table>

performance is improved as the noise becomes colored, i.e., in Figure 4.5, the performance using the library $M_{\text{Prune}}$ is superior in the case of colored noise compared to the case of white noise.

Computational complexity

Our proposed method uses the ADMM method and has the same order of computational complexity as the methods in [3, 14, 17, 20, 33, 82, 50]. Table 4.2 compares the running time of these algorithms in seconds per pixel, which is commonly used [3, 20, 14] as a measure of the computational efficiency of these algorithms.

We implemented the NCLS in our simulation following [14], which has a similar running time as the SUnSAL. The comparison shows that our proposed method is faster than other state-of-the-art methods, except SUnSAL. Besides, the size of the library has a significant impact on the running time.
Experiments with real hyperspectral data

For the real data experiments, we utilize a subimage of the hyperspectral data set of the AVIRIS cuprite mining in Nevada. Indeed, this hyperspectral data cube is very common in different literature works for the evaluation of unmixing methods [14, 30, 18]. This scene contains 224 spectral bands ranging from 0.4 µm to 2.5 µm. However, we remove the spectral Sub-bands 1 to 2, 105 to 115, 150 to 170 and 223 to 224 due to the water-vapour absorption, as well as low SNRs in the mentioned sub-bands. Thus, we applied all unmixing methods over the rest of the 188 spectral bands of the hyperspectral data scene.

Figure 4.6 illustrates six samples of the estimated fractional abundances by different unmixing methods. We exploited the pruned hyperspectral library (i.e., \( M_{\text{Prune}} \)) for the unmixing process and used the same parameter setting described in Section (4.2.3). Indeed, we can produce a visual description of the fractional abundances in regards to each individual pixel by means of unmixing methods. At the point of visual comparison, the darker pixels exhibit a smaller proportion of the corresponding spectral signatures of the endmembers. Conversely, the higher contribution of the endmember in the specific pixel can be presented by a lighter pixel. Eventually, we can infer that our proposed unmixing methods can share a high degree of similarity to the SUnSAL algorithm in which its performance was evaluated in [20] compared to the Tricorder maps.
4.2. $\ell_0$-NORM SPARSE HYPERSPECTRAL UNMIXING USING ARCTAN SMOOTHING

<table>
<thead>
<tr>
<th>Method</th>
<th>Kaolinite</th>
<th>Buddingtonite</th>
<th>Montmorillonite</th>
<th>Chalcedony</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCLS</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
</tr>
<tr>
<td>SuSaL</td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
</tr>
<tr>
<td>BICE</td>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
</tr>
<tr>
<td>CZ</td>
<td><img src="image13.png" alt="Image" /></td>
<td><img src="image14.png" alt="Image" /></td>
<td><img src="image15.png" alt="Image" /></td>
<td><img src="image16.png" alt="Image" /></td>
</tr>
<tr>
<td>SA1</td>
<td><img src="image17.png" alt="Image" /></td>
<td><img src="image18.png" alt="Image" /></td>
<td><img src="image19.png" alt="Image" /></td>
<td><img src="image20.png" alt="Image" /></td>
</tr>
<tr>
<td>SA2</td>
<td><img src="image21.png" alt="Image" /></td>
<td><img src="image22.png" alt="Image" /></td>
<td><img src="image23.png" alt="Image" /></td>
<td><img src="image24.png" alt="Image" /></td>
</tr>
<tr>
<td>SA3</td>
<td><img src="image25.png" alt="Image" /></td>
<td><img src="image26.png" alt="Image" /></td>
<td><img src="image27.png" alt="Image" /></td>
<td><img src="image28.png" alt="Image" /></td>
</tr>
</tbody>
</table>

Figure 4.6: Estimated abundance fraction maps for the subimage of AVIRIS cuprite using different unmixing methods.
4.2. $\ell_0$-NORM SPARSE HYPERSPECTRAL UNMIXING USING ARCTAN SMOOTHING

For each of these methods, we concatenated the output abundances fractions of all pixels (four abundances are shown in Figure 4.6) into one vector. Using these experimental output vectors for AVIRIS cuprite mining in Nevada, Figure 4.7 shows the estimated cumulative distribution function (CDF) of the estimated fractional abundances of different methods in order to compare the sparsity of the output of those methods. Figure 4.7 reveals that the outputs of SA3, SA1 and SA2 have the highest sparsity, respectively, among the considered methods. More specifically, 3%, 1% and 0.3% of the estimated fractional abundances are non-zero, respectively, using SA3, SA1 and SA2; whereas, about 7.9%, 7.6%, 4.7% and 3.2% of them are more than $10^{-3}$, respectively, for SUnSAL, NCLS, BiICE and CZ.

Figure 4.7: The estimated CDF of the fractional abundances of different methods over AVIRIS cuprite mining in Nevada.
4.2. \( \ell_0 \)-NORM SPARSE HYPERSONTICAL UNMIXING USING ARCTAN SMOOTHING

4.2.4 Conclusion

We have considered the LSSU with an iterative approximation of the \( \ell_0 \)-norm problem through an arctan function. Our approximation starts with the \( \ell_1 \)-norm problem, which is convex and has a unique optimal solution. As the algorithm converges to its initial optimal solution, we iteratively update our approximation toward the \( \ell_0 \)-norm problem. The superiority of this method is because our objective function is initially convex and initially converges to the optimal solution of the \( \ell_0 \)-norm problem. By updating this function iteratively, we iteratively make accurate approximation of the \( \ell_0 \)-norm minimization. The proposed approximation is controlled by updating parameter \( \sigma \). Furthermore, we have proven that the set of local optima of our objective function is a continuous set versus \( \sigma \) with the Hausdorff distance metric. This means that the gradual increase of \( \sigma \) along with iterative minimization of the proposed objective function leads to the optimal solution. By virtue of this theorem, the algorithm tracks the local optima of the current approximation and most local minima of the \( \ell_0 \)-norm problem. This is affirmed by our experiments that the number of non-zero elements of the solution using our method is significantly less than that of existing methods while the RSNR is improved. We must note that finding an optimal increasing sequence for \( \sigma \) is still an open problem, as a more conservative increasing sequence results in more computational cost, and an aggressive increasing sequence leads to a suboptimal solution. Moreover, we evaluated the role of Lagrangian multiplier \( \lambda \) and investigated two update rules for \( \lambda \). We applied the ADMM method to solve the minimization problem. We compared our proposed methods to several state-of-the-art methods using a simulated dataset, as well as the cuprite AVIRIS data cube. Our results illustrate that the proposed method outperforms these methods.
in terms of the achieved RSNR and, in terms of PoS, outperforms all of them, except the SUnSAL method for the synthetic data. For the subimage of cuprite AVIRIS, 3%, 1% and 0.3% of estimated abundances are non-zero using our proposed methods, whereas about 7.9%, 7.6%, 4.7% and 3.2% of them are more than $10^{-3}$ using other competitive algorithms.
Chapter 5

Blind Sparse Hyperspectral Unmixing Through Nonnegative Matrix Factorization

5.1 Collaborative Unmixing Hyperspectral Imagery via NMF

In this part, we are initiated to achieve the fact that the fractional abundances of endmembers can be effected by imposing sparsity among the endmembers collaboratively for all hyperspectral pixels. In fact, in hyperspectral images, all the pixels share the same set of spectral signatures of materials lying into a lower dimensional subspace [17]. Although the motivation of simultaneous sparse technique [115] is used in [17] for the unmixing purpose, the spectral library was assumed to be known. Besides, the averaging is applied as an $\ell_1$-norm term. Inspired by this motivation, we introduce a new minimization problem that considers the collaborative $\ell_{2,q}$-norm term along with a sparse approximation term for the abundances when the spectral library is also unknown. To solve the acquired minimization problem, we apply the multiplicative updating rule used for the standard NMF problem [7]. We also show that the updating rules will be guarantee to reach a local minimum. The simulation results demonstrate the effectiveness of our proposed method and outperform the
other state-of-the-art methods for both metrics of SAD and AAD metrics.

5.1.1 Proposed collaborative NMF-based unmixing method

In this section, we first reformulate the equation (2.15) as the following cost function in order to propose our method of an appropriate way:

\[ c(A, X) = \frac{1}{2} ||Y - AX||^2_2 + \alpha ||X||^\frac{1}{2}_1, \]  

(5.1)

where \( ||X||^\frac{1}{2}_1 = \sum_{i=1}^{N} \sum_{j=1}^{P} x_{ij}^\frac{1}{2} \) and \( x_{ij} \) is the corresponding abundance for the \( i \)-th endmember at the \( j \)-th pixel and \( \alpha > 0 \) is the Lagrangian parameter. More \( \ell_1 \)-based methods are also studied for the hyperspectral unmixing purpose, e.g., [54, 8, 55]. The multiplicative updating rules, which has been proposed for the standard NMF in [7], was also used in [5] to solve the corresponding minimization problem of equation(5.1) under nonnegativity constraints of \( A \) and \( X \) as follows

\[
A \leftarrow A \ast \frac{YX^T}{AXX^T} \\
X \leftarrow X \ast \frac{A^TY}{(A^TAX + \frac{\alpha}{2}X^{-\frac{1}{2}})}
\]

(5.2)

(5.3)

where \( \ast \) and \( \ast / \) are the element-wise multiplication and division, respectively.

Now, we define the following \( \ell_{2,q} \)-norm minimization problem

\[
\min_{A \geq 0, X \geq 0} \ f(A, X),
\]

(5.4a)

\[
f(A, X) = \frac{1}{2} ||Y - AX||^2_2 + \alpha ||X||^\frac{1}{2}_1 + \beta ||X||^q_{2,q},
\]

(5.4b)
5.1. COLLABORATIVE UNMIXING HYPERSPECTRAL IMAGERY VIA NMF

where

\[ ||X||_{2,q} = \left( \sum_{i=1}^{N} \left( ||x^i||_2^q \right)^{\frac{1}{q}} \right), \quad (5.5) \]

and \( ||x^i||_2^q = \left( \sum_{j=1}^{P} |x_{ij}|^2 \right)^{\frac{q}{2}} \) [116] and \( \alpha > 0 \) and \( \beta > 0 \) are the Lagrangian regularizes.

In fact, the term of \( ||X||_{2,q}^q \) in (5.4b) is the \( q \)-th power of \( ||X||_{2,q} \) and results in the sum of \( ||x^i||_2^q \) over \( i \)'s. It is obvious that we have the exact averaging of the \( \ell_2 \)-norm of vectors \( \{x^i\} \) when \( q = 1 \), i.e., \( \ell_{2,1} \)-norm.

The updating rule for the spectral signatures \( A \) keeps the same with (5.2). To find the updating rule for \( X \), we first take the partial derivative of (5.4b) with respect to \( X \) which gives the following result

\[
\frac{\partial f(A, X)}{\partial A} = A^TAX - A^TY + \frac{\alpha}{2}X^{-\frac{1}{2}} + \beta F \cdot \text{sign}(X). \quad (5.6)
\]

where

\[
F = q \left( \begin{array}{cccc}
(||x^1||_2^q)^{\frac{2}{q}-1} & (||x^1||_2^q)^{\frac{2}{q}-1} & \cdots & (||x^1||_2^q)^{\frac{2}{q}-1} \\
(||x^2||_2^q)^{\frac{2}{q}-1} & (||x^2||_2^q)^{\frac{2}{q}-1} & \cdots & (||x^2||_2^q)^{\frac{2}{q}-1} \\
\vdots & \vdots & \ddots & \vdots \\
(||x^N||_2^q)^{\frac{2}{q}-1} & (||x^N||_2^q)^{\frac{2}{q}-1} & \cdots & (||x^N||_2^q)^{\frac{2}{q}-1}
\end{array} \right) \cdot \text{sign}(X) \quad (5.7)
\]

and \( \text{sign function operates element-wisely over } X. \)

Following the Karush-Kuhn-Tucker (KKT) conditions and applying the transposition and division, the updating rule can be determined as follows:

\[
X \leftarrow X \cdot A^TY / (A^TAX + \frac{\alpha}{2}X^{-\frac{1}{2}} + \beta F \cdot |X|). \quad (5.8)
\]
5.1. COLLABORATIVE UNMIXING HYPERSONTRAL IMAGERY VIA NMF

The value of Lagrangian parameter $\alpha$ depends on the degree of sparseness for the fractional abundances of endmembers and it can be estimated based on the proportion of $\ell_1$-norm of the observed signals to the corresponding $\ell_2$-norm as the similar way in [85]. In our simulations, we set a fraction of the value defined in [85] for our proposed method as follows

$$
\alpha = \frac{\eta}{\sqrt{L}} \sum_{k=1}^{L} \frac{\sqrt{P - ||y_k||_2^2}}{\sqrt{P - 1}},
$$

(5.9)

where $y_k$ is the $k$-th band of the observed hyperspectral images and $0 < \eta \leq 1$ is a known constant value.

To impose the ASC over the fractional abundances, we append $\delta 1_T$ to $Y$ and $A$, respectively, during the iterative process as the same way used in [104] and many literature afterwards, e.g., [54, 5, 8, 55]. Then, the matrices $\bar{Y}$ and $\bar{A}$ take the place of $Y$ and $A$ as follows:

$$
\bar{Y} = \begin{bmatrix}
Y \\
\delta 1_T^T
\end{bmatrix},
\bar{A} = \begin{bmatrix}
A \\
\delta 1_N^T
\end{bmatrix}
$$

(5.10)

where $\delta$ is a known constant value that controls the effect of ASC and $1_l$ is the $l \times 1$ column vector with all elements equal to one.

Now, we show that our proposed method is converging.

**Proposition 1.** The loss function in (5.4b) is nonincreasing under (5.2) and (5.8).

**Proof.** The convergence of the method under the updating rule for $A$ in (5.2) can be shown similar to [7]. For the updating rule of $X$ in (5.8), we can follow up the same procedure in [5] and shows the cost function in (5.4b) is nonincreasing function under (5.8).

Our proposed $\ell_{2,q}$-NMF based unmixing method is summarized as the following
5.1. COLLABORATIVE UNMIXING HYPERSPECTRAL IMAGERY VIA NMF

Algorithm 7 Pseudocode of the collaborative NMF-based unmixing method.

1: **Input:** The observed matrix data $Y$
2: Set parameters $\alpha$, $\beta$, and $\delta$.
3: Initialize $A$ using VCA [43] or randomly from interval $[0, 1]$.
4: Initialize $X$ using $(A^T A)^{-1} A^T Y$.
5: **Repeat:**
6: Replace $A$ and $Y$ by (5.10).
7: Compute $f_{\text{old}} = f(\bar{A}, \bar{X})$ using (5.4b).
8: Update $A$ using (5.2).
9: Update $X$ using (5.8).
10: Replace $A$ and $Y$ by (5.10).
11: Compute $f_{\text{new}} = f(\bar{A}, \bar{X})$ using (5.4b).
12: Continue if the iteration number is less than $I_{\text{max}}$ or $|f_{\text{new}} - f_{\text{old}}| > \varepsilon$.
13: **Output:** Spectral signature matrix of endmembers $A$ and their fractional abundances $X$.

5.1.2 Experimental results

In this section, we evaluate our proposed method through different experiments. First, we use the USGS library [93] to generate the synthetic data as follows.

We select $n_N$ random spectral signatures from [93] for all of the following experiments. It should be noted that these signatures must be linearly independent. Then, we consider $P = r^2 \times r^2$ pixels of entire image to produce linear mixtures where $r$ is an integer known value. In fact, these pixels are divided into $r \times r$ patches. Each patch is assigned randomly by an integer value, say $n_E$ that is between 2 (to avoid generating the pure pixels) and $m_E$ shows the maximum number of endmember for pixels involved in the patch. Moreover, the fractional abundances for these spectral signatures are generated randomly with the Dirichlet distribution [114] based on the assigned number of materials for each patch, i.e., $n_E$. In order to make sure that
the number of spectral signatures constructs a mixed pixel has enough contribution to build such pixel, we replace the abundances of all pixels whose their fractional abundances are larger than a threshold, say $n_T$, with the equal contributions, i.e., $\frac{1}{n_E}$.

Afterwards, we passed the generated pixels through the Additive White Gaussian Noise (AWGN) and observed the outputs. For the evaluation purpose, we use the SAD metric to measure the similarity between the recovered spectral signatures and the ground-truth samples. Besides, we use the AAD metric to find the similarity of the estimated fractional abundances with their ground-truth values. They are defined as follows

$$SAD_i = \arccos\left(\frac{\hat{a}_i^T a_i}{||\hat{a}_i|| ||a_i||}\right),$$  

$$AAD_i = \arccos\left(\frac{\hat{x}_i^T x_i}{||\hat{x}_i|| ||x_i||}\right),$$

where $\hat{a}_i$ and $\hat{x}_i$ represent the estimated spectral signature of $i$-th material and the corresponding estimated fractional abundances.

**Implementation setting:** In our simulations, we set $q = 0.01$ in (5.4b). Also, we set $\eta = 0.5$ and $\beta = 0.2\alpha$ in (5.9) due to the performance consideration and set $\eta = 1$ for $\ell_{\frac{1}{2}}$-NMF method as recommended in [5]. Selecting the larger value of $\delta$ gives the closer the columns of $X$ to the full additivity constraint e.g., [5, 55] which leads to more time for unmixing process in simulations. We choose $\delta = 5$ for our experiments over all unmixing methods. Moreover, we set $r = 8$, $n_N = 12$, $m_E = 5$ and $n_T = 0.7$ to generate $P = 4096$ pixels from 12 spectral signatures chosen from [93] with $L = 224$ spectral bands. Thus, each pixel has at most 5 mixed materials where the maximum purity is set to 0.7. Finally, we set $I_{max} = 1000$ and $\varepsilon = 10^{-4}$ in
Algorithm 7 as well as $\ell_1$-NMF method proposed in [5].

For the evaluation, first, we prepare the unmixing results for a fixed value of SNR. Figure 5.1 compares the estimated spectral signatures by our proposed method and the other state-of-the-art methods for 4 sample materials out of 12 while the SNR is set to 25 dB. We can observe that our proposed method gives the closest results to
the original spectral signatures of materials compared with the other two unmixing methods.

Also, we evaluate the performance of our proposed method for different values of the noise power. Figure 5.2 shows the values of SAD and AAD as the functions of SNRs for our proposed method as well as the other two state-of-the-art methods. The value of inf represents noise-free observation. Our proposed method outperforms the other methods from lower SNRs to the noise-free environments.

![Figure 5.2: Simulation results of unmixing methods of (a) SAD (b) AAD as functions of SNRs.](image)

**5.1.3 Conclusion**

In this part, we proposed a method of unmixing hyperspectral images based on the collaborative property of the fractional abundances of endmembers through the NMF problem. We introduced a new cost function based on the NMF and $\ell_{2,q}$-norm to formulate our minimization problem. Then, we applied the multiplicative updating rules to solve the desire objective function. We showed that our proposed method converges and evaluated it by different experiments over the USGS spectral library.
5.1. COLLABORATIVE UNMIXING HYPERSPECTRAL IMAGERY VIA NMF

Our simulation results illustrated that the proposed method outperformed the other state-of-the-art methods in terms of SAD and AAD metrics.
5.2 Multilayer NMF Sparse Hyperspectral Unmixing via Adaptive $\ell_p$-norm

In this section, we first introduce our objective function through multilayer NMF (MNMF) problem [117] and $\ell_p$-norm ($0 < p \leq 1$) term for the sparsity constraint. In fact, the MNMF method gives an enhanced performance compared with the classic usage of NMF specifically for badly scaled or ill-conditioned data [117, 118] and we utilize this achievement for our problem formulation. The original MNMF proposed method of [117] used the $\ell_2$-norm or $\ell_1$-norm terms for both dictionary matrix as well as the corresponding coefficients. However, we are interested in $\ell_0$-norm term for the sparse constraints. We propose to use an adaptive $\ell_p$-norm term where the value of $p$ is adaptively decreasing during iterations for each layer of MNMF. To solve the acquired minimization problem, we apply the multiplicative updating rules used for the standard NMF problem [7]. By having this approach, we can show that the updating rules will be guarantee to reach a local minimum at each layer. Thus, the resulted methods are converging. The simulation results demonstrate the effectiveness of our proposed methods and outperform several state-of-the-art methods for both estimated dictionary matrix and the corresponding abundances matrix.

5.2.1 Problem formulation and proposed methods

By considering the linear system model described in equation (2.12) for the blind sparse unmixing scenario, we define the following two cost functions

\begin{align}
    c_1(A, X) &= \frac{1}{2} \| Y - AX \|_F^2 + \alpha \| X \|_p^p, \\
    c_2(A, X) &= \frac{1}{2} \| Y - AX \|_F^2 + \alpha \| X \|_p^p + \frac{\beta}{2} \| A \|_F^2,
\end{align}

(5.13) (5.14)
5.2. MULTILAYER NMF SPARSE HYPERSPECTRAL UNMIXING VIA ADAPTIVE $\ell_p$-NORM  

where $||.||_F$ denotes the frobenius norm and $||X||_p^p = \sum_{i=1}^P \sum_{j=1}^N x_{ij}^p$, $\alpha > 0$ and $\beta > 0$ are the regularization parameters to control the sparsity constraint and to impose smoothness of solution for matrices $X$ and $A$, respectively. Indeed, we can bound the errors of the estimated dictionary matrix by adding the term $||A||_F^2$ in (5.14) and have an enhancement for the estimation of matrix $X$ consequently. We show this intuition in the experimental results in Section 5.2.2 in order to recover $X$.

We employ the multilayer NMF approach [117] to solve the corresponding minimization problems based on the cost functions in (5.13) and (5.14). Using the multiplicative updating rules [7], we derive the main updating steps for $X$ and $A$. Thus, by partial differentiating of (5.13) and (5.14) with respect to $A$ and $X$ for each layer $l (1 \leq l \leq B)$ and following the KKT conditions and applying the transposition and division, we obtain the following updating rules

$$A_l \leftarrow A_l \cdot (Y_l \cdot X_l^T) \cdot (A_l \cdot X_l \cdot X_l^T)$$

$$X_l \leftarrow X_l \cdot (A_l^T \cdot Y_l) \cdot (A_l^T \cdot A_l \cdot X_l + \alpha p X_l^{p-1})$$

$$A_l \leftarrow A_l \cdot (Y_l \cdot X_l^T) \cdot (A_l \cdot X_l \cdot X_l^T + \beta A_l)$$

where $\cdot$ and $\cdot.$ are the multiplication and division element-wise, $(.)^T$ is the transpose of the involved matrix and $l$ is the layer number. Also, it should be noted that (5.15) and (5.16) are the solutions of the mentioned minimization problem of equation (5.13) and the updating rules of (5.16) and (5.17) are the solutions of the minimization problem of the cost function (5.14), respectively.

We employ an exponential function to update the parameter $p$ similar to [41, 119]
as follows

\[ p = e^{-\gamma(t-1)} \] (5.18)

where \( \gamma > 0 \) is the fixed decay rate value and \( t \) is the iteration number. Moreover, we update the regularization parameters \( \alpha \) and \( \beta \) similar to the temperature schedule in the simulated annealing technique used in [117, 118] as follows

\[ \alpha = \beta = \delta_0 e^{-t/\tau} \] (5.19)

where \( \delta_0 \) and \( \tau \) are the constant values.

Our proposed \( \ell_p \)-MNMF based methods are summarized in the following algorithms.

**Algorithm 8** \( \ell_p \)-MNMF algorithm for the cost function (5.13)

1. **Input** The observed matrix data \( Y \)
2. Set parameters \( \delta_0, \tau, B, T_{\text{max}} \) and \( \varepsilon \)
3. Initialize \( A_0 \) and \( X_0 \) randomly
4. \( Y_l \leftarrow Y \)
5. while \( l \leq B \) do
6. for \( t \leftarrow 1, T_{\text{max}} \) do
7. Update \( A_t \) using (5.15)
8. Update \( X_t \) using (5.16)
9. Update \( p \) using (5.18)
10. Update \( \alpha \) using (5.19)
11. if \( \varepsilon > |c_1(A_t, X_t)_{(t)} - c_1(A_t, X_t)_{(t-1)}| \) then
12. Stop \( \triangleright c_1(.,.)_{(t)} \) denotes the value of \( c_1(.,.) \) at the iteration \( t \)
13. end if
14. end for
15. \( Y_{l+1} \leftarrow X_l \)
16. \( l \leftarrow l + 1 \)
17. end while
18. **Output** \( A = \prod_{l=1}^{B} A_l \) and \( X = X_B \)
5.2. MULTILAYER NMF SPARSE HYPERSPECTRAL UNMIXING VIA ADAPTIVE $\ell_p$-NORM

Algorithm 9 $\ell_p$-MNMF algorithm for the cost function (5.14)

1: **Input** The observed matrix data $Y$
2: Set parameters $\delta_0$, $\tau$, $B$, $T_{\text{max}}$ and $\varepsilon$
3: Initialize $A_0$ and $X_0$ randomly
4: $Y_l \leftarrow Y$
5: while $l \leq B$ do
6:   for $t \leftarrow 1, T_{\text{max}}$ do
7:     Update $A_t$ using (5.17)
8:     Update $X_t$ using (5.16)
9:     Update $p$ using (5.18)
10:    Update $\alpha$ and $\beta$ using (5.19)
11:   if $\varepsilon > |c_2(A_t, X_t)(t) - c_2(A_l, X_l)(t-1)|$ then
12:       Stop $\triangleright c_2(\cdot)(t)$ denotes the value of $c_2(\cdot)$ at the iteration $t$
13:   end if
14: end for
15: $Y_{l+1} \leftarrow X_l$
16: $l \leftarrow l + 1$
17: end while
18: **Output** $A = \prod_{l=1}^{B} A_l$ and $X = X_B$

5.2.2 Experiment over real data

For the evaluation of our proposed methods over real data, we employ a subset of Samson hyperspectral data set available from [120] and used in [121] for the unmixing process. This data set contains $95 \times 95$ pixels. Three different materials are interested in recovery: soil, tree and water. Then, we consider the following methods to compare the performance of them with our propose methods:

- ADMM-NMF [122]
- MLNMF [8]
- Classical NMF (CNMF) [7]
5.2. MULTILAYER NMF SPARSE HYPERSPECTRAL UNMIXING VIA ADAPTIVE $\ell_p$-NORM

- $\ell_{\frac{1}{2}}$-NMF [5]
- $\ell_1$-NMF using multiplicative updating rules

It should be noted that we use the VCA algorithm [43] for the initialization for all of the state-of-the-art methods and only for the first layer of our proposed methods. For our proposed methods, we set $\gamma = 3.45$, $\tau = 2$, $B = 5$, $T_{\text{max}} = 200$ and $\varepsilon = 10^{-4}$ for both proposed algorithms and set $\delta_0$ to $10^{-4}$ and 0.9 for Algorithm 8 and Algorithm 9, respectively.

Table 5.1 compares the results of our proposed methods with the other methods in terms of SAD. The bold and italic values display the first and the second best performance of the methods to estimate the spectral mixing matrix. Algorithm 8 gives the best result in terms of mean values of SAD for all recovered materials and Algorithm 9 has the second rank for the recovery purpose. Besides, both methods have the considerable enhancement for the dictionary matrix reconstruction compared with the other methods.

Table 5.1: Comparison of the SAD results of the various methods on the Samson hyperspectral data set

<table>
<thead>
<tr>
<th>Unmixing methods</th>
<th>soil</th>
<th>tree</th>
<th>water</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algorithm 8</td>
<td>0.0389</td>
<td><strong>0.0326</strong></td>
<td>0.0463</td>
<td><strong>0.0393</strong></td>
</tr>
<tr>
<td>Algorithm 9</td>
<td>0.0374</td>
<td>0.0367</td>
<td>0.0487</td>
<td><strong>0.0409</strong></td>
</tr>
<tr>
<td>ADMM-NMF [122]</td>
<td>0.9249</td>
<td>0.0730</td>
<td>0.3567</td>
<td>0.4515</td>
</tr>
<tr>
<td>MLNMF [8]</td>
<td><strong>0.0320</strong></td>
<td>0.0463</td>
<td>0.2852</td>
<td>0.1212</td>
</tr>
<tr>
<td>$\ell_{\frac{1}{2}}$-NMF [5]</td>
<td>0.0493</td>
<td>0.0426</td>
<td>0.3237</td>
<td>0.1385</td>
</tr>
<tr>
<td>$\ell_1$-NMF</td>
<td>0.2350</td>
<td>0.0490</td>
<td>0.0617</td>
<td>0.1152</td>
</tr>
</tbody>
</table>

Also, we show a visualization of the estimated fractional abundances for these three materials. We show these results for our proposed methods and two other specified unmixing methods in Figure 5.3.
Figure 5.3: A visualization of the abundances estimated from different methods for three components of Samson hyperspectral data set: #1 Soil, #2 Tree #3 Water.
5.2.3 Conclusion

We proposed a NMF method of blind sparse hyperspectral unmixing using $\ell_p$-norm by exponential reduction of $p$. We studied two different scenarios for a linear system model: (i) only the sparsity property is exploited (ii) both the sparsity of $X$ and smoothness of the solution is exploited. Then, we employed a multilayer NMF approach to solve the acquired minimization problems. Hence, we introduced two algorithms and evaluate the proposed methods through real data. Our simulations over both types of data confirm the superiority of the proposed methods on several state-of-the-art unmixing methods.
5.3 Smooth and Sparse Regularization for NMF Hyperspectral Unmixing

In this section, we use the NMF approach and propose a new minimization problem by introducing a Lipschitz continuous function and new regularizer term related to the characteristics of the spectral signatures of materials. First, we use the Lipschitz continuous function instead of $\ell_p$ (for $0 \leq p < 1$) to better exploit the sparse property of the abundances and enhance the numerical stability and the immunity to noise corruption. Following [22], we employ an arctan function as a measure of the sparsity instead of $\ell_p$-norm term which is not continuous and not differentiable for $0 \leq p < 1$. Through a parameter $\omega$, this function allows to adjust the desired level of sparsity of the obtained solution. Second, the smooth regularization term is introduced based on the observation of that the energy distribution of the Discrete Cosine Transform (DCT) of existing measured spectral libraries decay exponentially for higher DCT subbands, e.g., 99% of the cumulative energy of the spectral signatures of the materials is contained in that part of subbands [14]. This observation basically means that the spectrum of endmembers is an smooth function which is missing for unmixing processes to the best of our knowledge. Thus, to efficiently exploit this smooth property, we introduce a quadratic weighted norm of the DCT of the spectral library as a regularizer term. Indeed, we assign a filtering matrix includes 0's and 1's where the components that are expected to have less energy are mapped to 1’s and vise versa, thereby the latter components are suppressed more at the output. The DCT is suitable for this application since it can be implemented with a low order of computational cost and the higher components of the DCT of the spectral library are expected to have insignificant values. We derive a set of multiplicative updating rule [7] to solve our proposed NMF problem. We also prove that the proposed algorithm converges.
Our experiments on simulated as well as real hyperspectral data demonstrates the effectiveness of our proposed method.

5.3.1 Arctan-NMF: our proposed method for unmixing

The arctan function is recently used for some sparse minimization problems in different literatures. Approximating either the sign function [81] and [106] or the $\ell_0$-norm term [33] and [22] are such examples where some similarities between the $\ell_p$-norm term ($p < 1$) and arctan function are studied in [22] as well. In this part, we explore the use of arctan function as a sparse regularizer to approximate the $\ell_0$-norm term to solve our NMF problem. Indeed, we employ this function for the sparsity term of the fractional abundances.

Besides, we enforce an enhanced $\ell_2$-norm of the spectral library as a roughness penalty in order to smooth the spectral signatures and to restrict the related estimation error. To achieve so, we first look at some properties of the spectral signatures. Figure 5.4 and 5.5 show five spectral signatures and their corresponding DCT coefficients randomly chosen from splib06 [93]. Those figures illustrate all of the spectral signatures curves are relatively smooth and nonnegative. Figure 5.6 shows the mean squared value of DCT of all spectral signatures provided in [93] and we can observe that around 99% of the values are contained in the first 20 coefficients. Hence, we define a kind of regularizer coefficients denoted by $\Lambda$ with this observation. More precisely, we construct $\Lambda$ with the elements of 0 and 1 that they represent the components of normalized average energy are bigger or smaller than an specified threshold, respectively. Afterwards, we apply another transformation onto the spectral library under study denoted by matrix $\Gamma$. 
5.3. SMOOTH AND SPARSE REGULARIZATION FOR NMF
HYPERSPECTRAL UNMIXING

Figure 5.4: Signatures of 5 sample materials from USGS library (splib06): Almandine, Beryl, Carnallite, Meionite, and Strontianite.

Figure 5.5: The squared of the DCT of the spectral signatures of the mentioned materials in Figure 5.4.

Figure 5.6: The squared of the DCT of all materials signatures averaged over USGS library (splib06).
By translating the observation of smooth property of the spectral signatures of materials into $\ell_2$-norm term and considering the sparse property of fractional abundances through smooth function, we introduce the following cost function

$$f(A, X) = \frac{1}{2} ||Y - AX||_F^2 + \alpha \sum_{i,j} \frac{\arctan(\omega x_{ij})}{\arctan(\omega)} + \frac{\beta}{2} ||\Lambda \Gamma A||_F^2,$$  

(5.20)

where the Lagrangian parameters $\alpha > 0$ and the constant $\omega$ control the sparsity terms of the fractional abundances and $\beta > 0$ controls the smoothness of $A$. The matrix $\Gamma$ is the DCT. The diagonal $L$-dimensional matrix $\Lambda$ is designed as a filter by setting the diagonal elements of $\Lambda$ either to 0 or to 1, i.e., only if the average of the squared DCT of spectral signatures are greater than some threshold value, say $t$, the corresponding diagonal element is zero.

To solve the minimization problem regarding to the cost function defined in equation (5.20) subject to the nonnegative constraints for $A$ and $X$, we apply the multiplicative updating method. Thus, we must take the derivative of equation (5.20) with respect to $A$ and $X$ as follows:

$$\frac{\partial f(A, X)}{\partial A} = -YX^T + AXX^T + \beta \Gamma^T \Gamma A,$$  

(5.21)

$$\frac{\partial f(A, X)}{\partial X} = -A^T Y + A^T A X + \frac{\alpha \omega}{\arctan(\omega)} \left[ \frac{1}{1 + (\omega x_{ij})^2} \right],$$  

(5.22)

where $[.]$ represents the matrix form of its inside elements, $(.)^T$ denotes the transpose of the matrix, and $x_{ij}$ is the $ij$-th element of matrix $X$. 
5.3. SMOOTH AND SPARSE REGULARIZATION FOR NMF HYPERSPECTRAL UNMIXING

According to the Karush Kuhn Tucker (KKT) conditions and applying the transposition and division, the updating rules are achieved as follows:

\[
A \leftarrow A \cdot YX^T / (AXX^T + \beta \Gamma^T \Lambda^2 \Gamma A) \tag{5.23}
\]

\[
X \leftarrow X \cdot A^T Y / \left( A^T AX + \frac{\alpha \omega}{\arctan(\omega)} / (1 + \omega^2 X^2) \right) \tag{5.24}
\]

The following proposition guarantees the convergence of the Arctan-NMF in (5.23) and (5.24).

**Proposition 2.** The loss function in (5.20) is nonincreasing under (5.23) and (5.24).

**Proof.** See Appendix C.1. \qed

The ASC can be imposed over the fractional abundances within the updating rules by appending \(\delta 1^T\) to \(Y\) and \(A\) as mentioned earlier in (5.10).

Our proposed arctan-NMF based unmixing method is summarized as the following algorithm.

The regularization parameters \(\alpha\) and \(\beta\) in (5.20), respectively, are introduced to control the sparsity of the fractional abundances \(X\) and the smooth property of spectral signatures \(A\). After each iteration of (5.24), the zero elements of \(X\) remain zero which guarantees the iterative reduction of \(\|X\|_0\); larger values of \(\alpha\) speeds up this process. Thus, after a number of iteration, once \(\|X\|_0\) is small enough, since \(\alpha\) and \(\beta\) tend to zero, the cost function (5.20) becomes \(\frac{1}{2}\|Y - AX\|_F^2\) which will be further reduced. The number of unknowns is \(LN + \|X\|_0\), starts as the total number of elements of \(X\) and \(A\), i.e., \((L + P)N\) during initial iterations and iteratively reduces (since zero elements of \(X\) remain zero in (5.24). A larger \(\beta\) allows to exploit the
5.3. SMOOTH AND SPARSE REGULARIZATION FOR NMF
HYPERSPECTRAL UNMIXING

Algorithm 10 Pseudocode of the arctan-NMF based Unmixing Method.

1: **Input:** The measured hyperspectral data matrix $Y$ and parameters $\tau$, $\xi$, $\omega$, $I_{\text{max}}$.
2: Initialize $A$ using random initialization from interval $[0, 1]$ or using VCA [43].
3: Initialize $X$ using $(A^T A)^{-1} A^T Y$.
4: **Repeat:**
5: Substitute $A$ and $Y$ in (5.10).
6: Compute $f_{\text{old}} = f(A, X)$ using (5.20).
7: Update $A$ using (5.23).
8: Update $X$ using (5.24).
9: Substitute $A$ and $Y$ in (5.10).
10: Compute $f_{\text{new}} = f(A, X)$ using (5.20).
11: Stop if either $|f_{\text{new}} - f_{\text{old}}| < \varepsilon$ or the iteration number exceeds $I_{\text{max}}$.
12: **Output:** Spectral signature matrix of endmembers $A$ and their fractional abundances $X$.

side information about smoothness of $A$ during initial iterations to deal with larger number of unknowns. As the number of unknowns reduces the solutions for $X$ and $A$ become accurate alleviating the need to exploit the side information. Thus, we propose the following exponential decay functions for $\alpha$ and $\beta$ in (5.20)

$$\alpha = e^{-\frac{i}{\tau}}, \quad (5.25)$$
$$\beta = e^{-\frac{i}{\xi}}, \quad (5.26)$$

where $i$ is the iteration number, $\tau$ is the constant value and $\xi$ denotes the rate of vanishing regularization parameter $\beta$.

5.3.2 Experimental results

In this section, we evaluate our Arctan-NMF method through different experiments over both simulated data and real hyperspectral data. We use SAD and AAD metrics to measure how much the recovered spectral signatures are close to the corresponding
ground truth samples and to measure the similarity of the estimated fractional abundances with their ground truth values as defined in (5.11) and (5.12), respectively. Also, we use the USGS library [93] to generate the synthetic data and use two different types of real hyperspectral data including Samson [120] and Cuprite e.g., [43, 14] for our evaluation. We consider the following methods to compare the performance of them with our propose algorithm:

- Classic NMF method (CNMF) using multiplicative updating rules [7]
- $\ell_1$-NMF unmixing algorithm [6]
- $\ell_{\frac{1}{2}}$-NMF [5]
- VCA-FCLS

It should be noted that VCA-FCLS is the two stage unmixing method based on VCA [43] and fully constrained least square (FCLS) [104] and used in different literature (e.g., [8, 55, 54]) as the state-of-the-art method. Besides, we use the same parameters setting for the above methods as mentioned in the corresponding references for the simulation purposes unless the ones we point out in the simulations.

**Experiments on simulated data**

In this Section, we examine our proposed unmixing method for a quantitative analysis based on different metrics mentioned above. Then, we compare the performance of the Arctan-NMF with the other state-of-the-art methods via various experiments.

There are different methods to generate the synthetic hyperspectral data, e.g. [123] and [34], and we use the similar method proposed in our previous publication [34] due to its flexibility to produce data for different following scenarios of experiments.
That is, we first choose $n_S$ random spectral signatures from the spectral library [93] where they are linearly independent. We produce $P$ pixels of entire image with size $r^2 \times r^2$ where $r$ is an integer known value and the generated pixels are supposed as $r \times r$ regions (patches). We assign a random integer value $n_E$ to each patch where $1 < n_E \leq E_m$ and $E_m$ denotes the maximum number of endmembers may build mixed pixels in the patch. To produce the fractional abundances for each pixel, we use the Dirichlet distribution [114] based on the given number of materials of each patch. Moreover, we apply a simple threshold filtering over these fractional abundances to ensure that each mixed pixel has enough contribution of spectral signatures. Hence, the set of abundances for each pixel whose their fractional are larger than the threshold $n_T$ where $0 < n_T < 1$ are replaced by $\frac{1}{n_E}$. In fact, $n_T$ adjusts the percentage of purity of a material in each mixed pixel in which the smaller value of $n_T$ gives the higher degree of mixed materials.

**Experiment 1. Sparsity approximation parameter $\omega$**

The first experiment aims at evaluating the estimated abundances with respect to the changes of sparsity parameter $\omega$. First, we generate two different synthetic data sets (DS) as follows:

- **DS1:** Set $n_S = 12$, $r = 5$, $E_m = 5$ and $n_T = 0.7$ to produce the synthetic data with $P = 625$ pixels from 12 spectral signatures.
- **DS2:** Set $n_S = 12$, $r = 6$, $E_m = 7$ and $n_T = 0.8$ to produce the synthetic data with $P = 1296$ pixels from 12 spectral signatures.

Then, we pass both data sets DS1 and DS2 through the additive white gaussian noise (AWGN) with SNR = 20 dB and 30 dB, respectively. We select $t = 0.01$ and $2 \times 10^{-4}$
for the filtering threshold to construct two different types of $\Lambda$ and applied them over DS1 and DS2, respectively. Also, we set $\xi = 0.4$, $\tau = 100$, $\delta = 0.1$ and $I_{\text{max}} = 2000$. The value of $\omega$ is varied from 1 to $10^4$ in order to calculate the second term of (5.20) as the approximate of $\ell_0$-norm term. Then, the corresponding metrics are computed. This experiment is repeated for 100 trials.

Figure 5.7 shows these results for these two data sets and different noise powers. We observe that the best values of $\omega$ in order to minimize the mentioned metrics are in the range of 400 to 1000. Thus, we set $\omega = 700$ to approximate the fractional abundances term in (5.20) for the rest of experiments.

**Experiment 2. Iteration number**

In this experiment, we investigate the effect of the number of iterations over different mentioned metrics. We employ the same data sets in 5.3.2 with AWGN and set $\delta = 0.1$ and $\omega = 700$. The value of $I_{\text{max}}$ is varied from 1 to 4000 with the step size of 100. Then, we compute the metrics and present the results in Figure 5.8.

We observe that the metrics have their lowest values by choosing the maximum number of iteration for the range of greater than 1700. Then, we choose $I_{\text{max}} = 2000$ for the remaining experiments.

**Experiment 3. Decay parameter $\xi$**

In the third experiment, the performance of the algorithm is investigated in terms of $\xi$ which has a direct impact for the regularization parameter $\beta$ defined by (5.26). We use the same data sets DS1 and DS2 and choose $\delta = 0.1$, $\omega = 700$ and $I_{\text{max}} = 2000$. We change the value of $\xi$ from $10^{-3}$ to 20 and compute their corresponding metrics.
Figure 5.7: The metric values as a function of $\omega$ for different data sets and SNRs.

Figure 5.9 shows the results and it can be observed that the lowest values for the metrics can be found for the range of $0.3 \leq \xi \leq 5$ and we set $\beta = 0.4$ for our experiments.

**Experiment 4. Robustness to the number of endmembers and mixing degree**

This experiment aims to study the effect of the maximum number of endmembers as well as different mixing degree of purities on the performance of Arctan-NMF. For
5.3. SMOOTH AND SPARSE REGULARIZATION FOR NMF HYPERSPECTRAL UNMIXING

Figure 5.8: The metric values as a function of iteration numbers for different data sets and SNRs.

For this purpose, the value of $E_m$ and $n_T$ are varied from 2 to 10 and 0.1 to 0.9, respectively. We set $n_S = 12$ and $r = 5$ to generate 625 pixels of synthetic data from 12 spectral signatures. Also, we set the parameter setting of the proposed methods with $\xi = 0.4$, $\omega = 700$, $\delta = 15$ and $I_{max} = 2000$.

We show these results in Figure 5.10 and observe that the value of different metrics are decreasing as purity levels are increasing for a given maximum number of end-members. This trend is almost the same when the number of maximum endmembers
are decreasing for a given value of purity.

Furthermore, we compare the results of our proposed unmixing method with several state-of-the-art methods mentioned earlier. Figure 5.11 shows a comparison of the performance of these methods through different metrics when the degree of purities for the materials varies for a given maximum number of endmembers, i.e. $E_m = 5$. The Arctan-NMF outperforms the other unmixing methods for all values of SAD. Also, it gives the better results for AAD specifically the degree of purities, $n_T$, is greater than 0.6.
5.3. SMOOTH AND SPARSE REGULARIZATION FOR NMF
HYPERSPECTRAL UNMIXING

![Image of graph](image)

Figure 5.10: The metric values as a function of maximum number of endmembers and degree of purities contributed at each pixel at SNR = 20 dB.

Alternatively, Figure 5.12 compares the results of the proposed unmixing method with the other methods for the range of maximum number of endmembers from 2 to 10, present at each mixed pixel for a given maximum degree of purities, $n_T = 0.8$. We observe that our purposed method has the lowest value of metrics compared with the other methods for all metrics except the maximum number of endmembers is greater than 7 for the AAD metric.
5.3. SMOOTH AND SPARSE REGULARIZATION FOR NMF

HYPERSPECTRAL UNMIXING

Figure 5.11: The metric values as a function of purities for $E_m = 5$ at SNR = 20 dB for different unmixing methods.

**Experiment 5. Noise power robustness**

In this experiment, we compare the performance of the Arctan-NMF method with several state-of-the-art unmixing methods with respect to the SNR values over two data sets mentioned in the first experiment. The noise levels are varied between 15 dB to 35 dB with the step size of 5 dB. We measure the two metrics of SAD and AAD and the results are shown in Figure 5.13.

We observe that the Arctan-NMF method outperforms the other state-of-the-art
Figure 5.12: The metric values as a function of the maximum number of endmembers contributed at each mixed pixel for \( n_T = 0.8 \) at SNR = 20 dB for different unmixing methods.

unmixing methods. Obviously, the performance will be improved as the noise power decreases for all of unmixing methods.

**Experiment 6. Robustness to the number of pixels**

This experiment is taken to compare the performance of the Arctan-NMF with the other methods when the image size is varied. To sake this experiment, we set \( n_S = \)
5.3. SMOOTH AND SPARSE REGULARIZATION FOR NMF
HYPERSPECTRAL UNMIXING

Figure 5.13: The metric values as a function of the noise power (SNR in dB) for different unmixing methods over two different data sets.

12, $E_m = 6, n_T = 0.7$ and $r \in \{4, 5, \ldots, 10\}$ to produce the synthetic data with $P = r^2 \in \{256, 625, \ldots, 10000\}$ pixels from 12 spectral signatures. For our proposed unmixing method, we set $t = 2 \times 10^{-4}$ for the filtering threshold to construct $\Lambda$ and set $\xi = 0.4, \tau = 100, \delta = 0.1, \omega = 700$ and $I_{\text{max}} = 2000$.

Figure 5.14 shows the performances of the Arctan-NMF as well as other unmixing methods versus the number of pixels. We observe that the Arctan-NMF outperforms the other methods.
Figure 5.14: The metric values as a function of the number of pixels in the scene for $mE = 6$ and $aT = 0.7$ at SNR = 20 dB.

Experiment 7. Robustness to the initialization

In this experiment, the performance of different unmixing methods are evaluated in terms of the initialization setup. Thus, we initialize all of unmixing methods with either uniformly random distribution or VCA algorithm [43]. We use the similar data sets as well as parameter setting described in Section 5.3.2 for this experiment. Figure 5.15 shows the results. We observe that the Arctan-NMF outperforms the other methods for both type of initialization over two different data sets. Furthermore, the
5.3. SMOOTH AND SPARSE REGULARIZATION FOR NMF

HYPERSONTICAL UNMIXING

Figure 5.15: The impact of the initialization of the unmixing method over different metrics for two data sets DS1 and DS2.

The initialization with the endmember extraction algorithm such as VCA improve the performance for all methods compared with the random initialization.

Experiments on real data

For the evaluation of the Arctan-NMF over real data, we applied the unmixing methods on two different types of hyperspectral images. In the first experiment, we utilize the Samson hyperspectral image [120] where the original data set contains $952 \times 952$. 
5.3. SMOOTH AND SPARSE REGULARIZATION FOR NMF HYPERSONTICAL UNMIXING

Table 5.2: Comparison of the SAD results of the various methods on the Samson hyperspectral data set.

<table>
<thead>
<tr>
<th>Method</th>
<th>( \xi = 4 )</th>
<th>( \xi = 0.6 )</th>
<th>( \delta )</th>
<th>Soil</th>
<th>Tree</th>
<th>Water</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arctan-NMF</td>
<td>0.0106</td>
<td>0.0369</td>
<td>0.0583</td>
<td>0.0353</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \xi = 0.6 )</td>
<td>0.0163</td>
<td>0.0272</td>
<td>0.1315</td>
<td>0.0583</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \ell_1 \text{-NMF} ) [5]</td>
<td>0.0318</td>
<td>0.0522</td>
<td>0.3118</td>
<td>0.1319</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \ell_1 \text{-NMF} ) [6]</td>
<td>0.0301</td>
<td>0.0563</td>
<td>0.2188</td>
<td>0.1017</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CNMF [7]</td>
<td>0.0382</td>
<td>0.0632</td>
<td>0.2659</td>
<td>0.1024</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLNMF [8]</td>
<td>0.0185</td>
<td>0.0419</td>
<td>0.1646</td>
<td>0.0750</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VCA-FCLS</td>
<td>0.0207</td>
<td>0.0459</td>
<td>0.1299</td>
<td>0.0667</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

pixels with 156 spectral bands ranging from 0.4 \( \mu \text{m} \) to 0.89 \( \mu \text{m} \). In this experiment, we use a subset of the original data set with 95 \( \times \) 95 pixels similar to one studied in [121]. The scene covers three different materials: soil, tree and water. We compare the SAD results of the Arctan-NMF with the state-of-the-art methods mentioned in Section 5.3.2 as well as Multilayer NMF (MLNMF) [8] (i.e., that is proposed based on the NMF problem and uses multilayer approach to solve the acquired problem) as shown in Table 5.2.

Table 5.2 compares the SAD values of the proposed method with methods in [7, 6, 5] and VCA-FCLS using \( \delta = 0.1 \) and also using the recommended \( \delta \) in these references. For the Arctan-NMF method we use \( \xi = 4 \) and \( \xi = 0.6 \) chosen from the suggested range obtained from the previous experiment over the synthetic data. More precisely, the value \( \xi = 4 \) is the suggested upper range of \( \xi \) which yields to the best performance in this experiment. Moreover, the Arctan-NMF exhibits competitive performance for the lower range \( \xi = 0.6 \) which is also used in our next experiment on real data. We note that we use the VCA algorithm [43] for initializing these methods.
5.3. SMOOTH AND SPARSE REGULARIZATION FOR NMF HYPERSONCTRAL UNMIXING

Figure 5.16: A visualization of fractional abundance maps of endmembers estimated by different unmixing methods over Samson hyperspectral image: a) Arctan-NMF ($\xi = 4, \delta = 0.1$) b) $\ell_2$-NMF [5] ($\delta = 15$) c) $\ell_1$-NMF [6] ($\delta = 15$) d) CNMF [7] ($\delta = 15$) e) MLNMF [8] ($\delta = 25$) f) VCA-FCLS ($\delta = 15$)
suggested range. Thus, we have only reported the results for $\delta = 15$. In Tables 5.2 and 5.3, the bold and italic fonts show the best and the second best performances, respectively. We observe that the Arctan-NMF method outperforms other methods in terms of mean value of SAD where both materials of Soil and Water are estimated with the lowest SAD values and Tree is estimated with a reasonable value of SAD, i.e., by considering $\xi = 4$. Figure 5.16 shows a visualization of the estimated fractional abundances for these three materials. Due to space limit, we show the corresponding results of the unmixing methods whose the $\delta$ parameter is recommended in their original references (i.e., for the first row of each cell corresponding to the methods in Table 5.2).

For the second experiment, we apply the unmixing methods over the hyperspectral data set for Cuprite in Southern Nevada collected by the AVIRIS which has been commonly used for evaluation of various methods, e.g., [43, 14, 22, 54, 8, 121]. As mentioned in the previous chapters, the image contains of 224 spectral bands ranging from 0.4 $\mu$m to 2.5 $\mu$m where the spectral sub-bands 1-2, 104-113, 148-167, and 221-224 are removed due to their low SNRs and the water-vapor absorption. Hence, we used 188 remaining spectral bands of the scene for our experiment.

The study in [43] shows that there are 14 materials in this scene, however some of them have similar spectra. Thus in some recent literature, e.g., [5, 54, 8, 121], a smaller number of endmembers are considered. For sake of performance comparison, we select 12 endmembers as true signatures. Table 5.3 compares the SAD results of the Arctan-NMF method with the other methods using the same parameter setting mentioned earlier and for $\delta = 0.5$ and the recommended value of $\delta$ in those references. We observe that the proposed Arctan-NMF method outperforms the state-of-the-art
5.3. SMOOTH AND SPARSE REGULARIZATION FOR NMF HYPERSONTICAL UNMIXING

Table 5.3: Comparison of the SAD results of the various unmixing methods on the AVIRIS cuprite data

Using the default parameter settings in [6, 5, 8, 7]

<table>
<thead>
<tr>
<th></th>
<th>Arctan-NMF, $\xi = 0.6$</th>
<th>$\ell_1$-NMF [5] $\delta = 0.5$</th>
<th>$\ell_1$-NMF [6] $\delta = 15$</th>
<th>CNMF [7] $\delta = 15$</th>
<th>MLNMF [8] $\delta = 25$</th>
<th>VCA-FCLS $\delta = 15$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alunite</td>
<td>0.0640</td>
<td>0.1017</td>
<td>0.1609</td>
<td>0.1677</td>
<td>0.1461</td>
<td>0.3174</td>
</tr>
<tr>
<td>Andradite</td>
<td>0.0951</td>
<td>0.0895</td>
<td>0.0792</td>
<td>0.1806</td>
<td>0.0652</td>
<td>0.0571</td>
</tr>
<tr>
<td>Buddingtonite</td>
<td>0.0810</td>
<td>0.1119</td>
<td>0.0845</td>
<td>0.0819</td>
<td>0.0914</td>
<td>0.0750</td>
</tr>
<tr>
<td>Dumortierite</td>
<td>0.1651</td>
<td>0.1404</td>
<td>0.1586</td>
<td>0.1618</td>
<td>0.1555</td>
<td>0.1393</td>
</tr>
<tr>
<td>Kaolinite #1</td>
<td>0.0715</td>
<td>0.1216</td>
<td>0.0803</td>
<td>0.0830</td>
<td>0.0755</td>
<td>0.0753</td>
</tr>
<tr>
<td>Kaolinite #2</td>
<td>0.0613</td>
<td>0.1899</td>
<td>0.1724</td>
<td>0.1809</td>
<td>0.1630</td>
<td>0.0664</td>
</tr>
<tr>
<td>Muscovite</td>
<td>0.1294</td>
<td>0.1289</td>
<td>0.2265</td>
<td>0.2026</td>
<td>0.2340</td>
<td>0.1143</td>
</tr>
<tr>
<td>Montmorillonite</td>
<td>0.0639</td>
<td>0.0654</td>
<td>0.0920</td>
<td>0.0936</td>
<td>0.0880</td>
<td>0.0738</td>
</tr>
<tr>
<td>Nontronite</td>
<td>0.0801</td>
<td>0.0706</td>
<td>0.0877</td>
<td>0.0955</td>
<td>0.0823</td>
<td>0.0733</td>
</tr>
<tr>
<td>Pyrope</td>
<td>0.0681</td>
<td>0.0628</td>
<td>0.1541</td>
<td>0.0781</td>
<td>0.1005</td>
<td>0.0890</td>
</tr>
<tr>
<td>Sphene</td>
<td>0.0549</td>
<td>0.1771</td>
<td>0.1261</td>
<td>0.1220</td>
<td>0.1021</td>
<td>0.0654</td>
</tr>
<tr>
<td>Chalcedony</td>
<td>0.1660</td>
<td>0.1714</td>
<td>0.0441</td>
<td>0.0607</td>
<td>0.1044</td>
<td>0.1623</td>
</tr>
<tr>
<td>Mean</td>
<td>0.0918</td>
<td>0.1193</td>
<td>0.1222</td>
<td>0.1257</td>
<td>0.1173</td>
<td>0.1082</td>
</tr>
</tbody>
</table>

Using the default parameter settings in [6, 5, 8, 7] except for $\delta = 0.5$

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Alunite</td>
<td>0.0658</td>
<td>0.0747</td>
<td>0.1783</td>
<td>0.1432</td>
<td>0.2103</td>
<td>0.3174</td>
</tr>
<tr>
<td>Andradite</td>
<td>0.0993</td>
<td>0.0708</td>
<td>0.2215</td>
<td>0.1653</td>
<td>0.0652</td>
<td>0.0571</td>
</tr>
<tr>
<td>Buddingtonite</td>
<td>0.0779</td>
<td>0.0909</td>
<td>0.1523</td>
<td>0.0798</td>
<td>0.1706</td>
<td>0.0750</td>
</tr>
<tr>
<td>Dumortierite</td>
<td>0.1755</td>
<td>0.1716</td>
<td>0.1905</td>
<td>0.1677</td>
<td>0.1188</td>
<td>0.1393</td>
</tr>
<tr>
<td>Kaolinite #1</td>
<td>0.0729</td>
<td>0.0789</td>
<td>0.0739</td>
<td>0.0769</td>
<td>0.0750</td>
<td>0.0753</td>
</tr>
<tr>
<td>Kaolinite #2</td>
<td>0.0662</td>
<td>0.1272</td>
<td>0.0836</td>
<td>0.1930</td>
<td>0.0657</td>
<td>0.0664</td>
</tr>
<tr>
<td>Muscovite</td>
<td>0.1691</td>
<td>0.1324</td>
<td>0.2414</td>
<td>0.2101</td>
<td>0.2092</td>
<td>0.1143</td>
</tr>
<tr>
<td>Montmorillonite</td>
<td>0.0743</td>
<td>0.0722</td>
<td>0.2370</td>
<td>0.0987</td>
<td>0.0616</td>
<td>0.0738</td>
</tr>
<tr>
<td>Nontronite</td>
<td>0.1074</td>
<td>0.0796</td>
<td>0.1087</td>
<td>0.0877</td>
<td>0.1081</td>
<td>0.0733</td>
</tr>
<tr>
<td>Pyrope</td>
<td>0.0703</td>
<td>0.0608</td>
<td>0.0731</td>
<td>0.0845</td>
<td>0.0680</td>
<td>0.0890</td>
</tr>
<tr>
<td>Sphene</td>
<td>0.0683</td>
<td>0.1509</td>
<td>0.0603</td>
<td>0.1034</td>
<td>0.0626</td>
<td>0.0554</td>
</tr>
<tr>
<td>Chalcedony</td>
<td>0.1900</td>
<td>0.1507</td>
<td>0.2713</td>
<td>0.1149</td>
<td>0.1830</td>
<td>0.1623</td>
</tr>
<tr>
<td>Mean</td>
<td>0.1031</td>
<td>0.1050</td>
<td>0.1577</td>
<td>0.1271</td>
<td>0.1055</td>
<td>0.1082</td>
</tr>
</tbody>
</table>

methods in estimating the spectral signatures of materials in most of the cases and in terms of the average SAD value.

Also, we show the comparison between the spectral signatures of the existing endmembers in the AVIRIS cuprite image recovered by the Arctan-NMF (red dotted lines) with the ground-truth USGS spectral library (black solid lines) in Figure 5.17. These two curves are very close in most cases. Moreover, the abundance maps of corresponding endmembers using the Arctan-NMF are shown in the same figure.
5.3. SMOOTH AND SPARSE REGULARIZATION FOR NMF

HYPERSPECTRAL UNMIXING

Figure 5.17: Cuprite AVIRIS results: Comparison of the (black and solid lines) original spectral signatures with the (red and dotted lines) endmember signatures extracted by our Arctan-NMF and their corresponding abundance maps of endmembers.
Table 5.4: Number of operations of the Arctan-NMF in terms of required flops in each iteration $(2N < L < P)$.

<table>
<thead>
<tr>
<th>Operation</th>
<th>Classic-NMF</th>
<th>Arctan-NMF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiplication</td>
<td>$2LPN + 2N^2(P + L) + N(L + P)$</td>
<td>$2LPN + N^2(2P + 3L) + 2N(L + P) + 1$</td>
</tr>
<tr>
<td>Division</td>
<td>$N(L + P)$</td>
<td>$N(L + 2P) + 1$</td>
</tr>
<tr>
<td>Addition</td>
<td>$2LPN + 2N^2(P + L) - 2N(L + P + N)$</td>
<td>$2LPN + N^2(2P + 3L) - N(L + 3N)$</td>
</tr>
<tr>
<td>Exponent</td>
<td>$NP$</td>
<td>$2LN\log(L)$</td>
</tr>
<tr>
<td>DCT</td>
<td>$-$</td>
<td>$O(LPN)$</td>
</tr>
<tr>
<td>Order of complexity</td>
<td>$O(LP)$</td>
<td>$O(LPN)$</td>
</tr>
</tbody>
</table>

5.3.3 Computational complexity

The computational complexity of unmixing algorithms is always crucial due to the large size of hyperspectral data. The Arctan-NMF is computationally similar to the classical NMF method both using the multiplicative updating approach. We analyze the computational cost by counting the number of required floating-point operations (flop) per iteration. For $\alpha = 0$, the updating rules in (5.2) and (5.3) are reduced to the classical NMF method. These rules use three arithmetic operations, i.e. addition, multiplication and division. The Arctan-NMF in (5.23) and (5.24) requires two additional operations which are exponent and DCT/IDCT operation in comparison with the classical NMF. Table 5.4 summarizes the total operations of the classical NMF and the proposed method. We observe that they both have the same complexity order of $O(LPN)$ operations per iteration.

5.3.4 Conclusion

We proposed a method of hyperspectral unmixing using through a new sparse cost function using the arctan function with the parameter $\omega$. We introduced a term which represents the out of band energy of spectral signatures of materials. To minimize the proposed cost function, we employed the multiplicative updating approach and
applied the KKT conditions. We proved the convergence of the resulting proposed Arctan-NMF algorithm and evaluated it through the synthetic data and two different types of real hyperspectral data sets. Our simulations revealed that our proposed method outperforms several state-of-the-art methods in estimating both fractional abundances and spectral library.
Chapter 6

Conclusions

In this chapter, the system model of sparse unmixing problems are summarized and the main contributions are drawn based on the principal results of the previous chapters. At the end, we suggest some issues and recommendations for the further research works.

6.1 Summary and Concluding Remarks

This thesis followed the sparse hyperspectral unmixing methods for the linear mixing model through two main scenarios: dictionary-aided semiblind scenario and nondictionary-aided blind scenario. The first scenario was studied in the both chapters 3 and 4 and we proposed several sparse hyperspectral unmixing methods through \( \ell_0 \)-norm problem approximations. In Chapter 5, we studied the second scenario for the unmixing methods and we introduced our proposed methods based on some statistical and physical properties of the spectral signatures of materials. The following is a summary of the main contributions of the thesis:

In Chapter 3, we first proposed a method of hyperspectral unmixing for the linear
regression model that employs an adaptive lasso problem using ADMM. We formulated a weighted $\ell_1$ norm problem under the reasonable given error to reconstruct the fractional abundances and to avoid inconsistent endmember selection in a sparse semiblind hyperspectral imaging process. We illustrated that this problem can be efficiently solved by appropriate selection of functions and parameters appearing in the ADMM approach.

In the second part of Chapter 3, we proposed another approach to approximate the $\ell_0$-norm term. In particular, we approximated the $\ell_0$-norm with the $\ell_p$-norm and iteratively reduced $p$ to enhance the results. This was motivated by our theorem that the set of minima for the $\ell_p$-norm problem is continuous in terms of $p$, which implies that smooth iterative reduction of $p$ results in an enhanced solution. We introduced a weighted $\ell_1$-norm approximation of the $\ell_p$-norm problem involving a parameter $\epsilon$ to deal with the fact that the $\ell_p$-norm problem is not Lipschitz continuous for $p < 1$. We proposed to use three different rules for updating the pair $(p, \epsilon)$. To solve this approximated problem using the previous solution, we employed the ADMM approach to derive the remaining updating steps. Alternatively, we considered the iteratively reweighted $\ell_2$-norm to approximate the $\ell_p$-norm and determined a closed form of the solution and updated the involved parameters of $p$ and $\epsilon$ based on our proposed updating rules. Experimental results for both parts exhibited that our proposed algorithms outperform several state-of-the-art algorithms in terms of the reconstruction errors and their probability of success.

In Chapter 4, we introduced a method of hyperspectral unmixing for LMM that employs an arctan function to approximate the $\ell_0$-norm term. This approximation makes the objective function smooth and facilitates the convergence and results in
reduced reconstruction errors. Then, we proposed another iterative method, which starts as an $\ell_1$-norm optimization that is convex, has a unique solution, converges quickly and iteratively tends to be an $\ell_0$-norm problem. We used the arctan function with the parameter $\sigma \geq 0$ in our optimization. This function is Lipschitz continuous and approximates $\ell_1$-norm and $\ell_0$-norm for small and large values of $\sigma$, respectively. We proved that the set of local optima of our problem is continuous versus $\sigma$. Thus, by a gradual increase of $\sigma$ in each iteration, we may avoid being trapped in a suboptimal solution. We proposed to use the ADMM approach for our minimization problem iteratively while increasing $\sigma$ exponentially. Our evaluations revealed the superiorities and shortcomings of the proposed methods compared to several state-of-the-art methods. We performed such evaluations in different experiments over both synthetic and real hyperspectral data, and the results of our proposed methods exhibited the sparsest estimated abundances compared to other competitive algorithms for the subimage of AVIRIS cuprite data.

In Chapter 5, a new method of hyperspectral unmixing for LMM was proposed while both the spectral signatures of endmembers and their fractional abundances are unknown. The proposed algorithm employed the NMF method as well as simultaneous (collaborative) sparse regression model. We formulated the NMF problem along with an averaging over the $\ell_2$-norm of the fractional abundances so-called $\ell_{2,q}$-norm term. We showed that this problem can be efficiently solved by using the KKT conditions. Our simulations revealed that the proposed algorithm outperformed the state-of-the-art methods in terms of SAD and AAD.

In the second part of Chapter 5, we proposed another new NMF based unmixing algorithm using $\ell_p$-norm (for $p \leq 1$) to exploit the sparse property of the contribution
of each member of the unknown library. We formulated our cost function by reducing $p$ iteratively and employed the multilayer NMF method to solve our minimization problem. We showed that our proposed methods outperformed several state-of-the-art methods by simulations. Moreover, we evaluated our proposed methods on the real hyperspectral image data and demonstrated the enhancement in recovery results of the proposed method.

In the last part of Chapter 5, we investigated the smooth property of the spectral signatures of materials and applied this feature to introduce a new cost function. Indeed, most of the energy of spectral signatures of materials is concentrated around the first few subbands resulting in smooth spectral signatures. To exploit this smoothness, we added a weighted norm of the spectral signatures of the materials and to limit their non-smooth errors. Also, we added the arctan functions of the endmembers to the $\ell_2$-norm of the error in order to exploit the sparse property of the fractional abundances. Then, we proposed to use the multiplicative iterative algorithm to solve the acquired minimization problem as a NMF problem. We applied our proposed Arctan-NMF method on the synthetic data from real spectral library and compared the performance of Arctan-NMF method with several state-of-the-art unmixing methods. Finally, we evaluated the efficiency of Arctan-NMF on two different types of real hyperspectral data. Our simulations revealed that the Arctan-NMF was more effective than the state-of-the-art methods in terms of SAD and AAD.

6.2 Future Work

The work in this thesis may be extended in many directions and we highlight some of them as follows:
- In Chapter 3, we introduced three updating rules to adjust the pair \((p, \epsilon)\) involved in the proposed methods which are not optimistic. Finding more appropriate methods is still an open problem, specifically to update the value of \(p\) at each iteration. As a future research work, we will look for provable method/methods to control the bifurcation of the candidate solutions at each iteration in order to achieve more accurate/reliable approximation and sparser solution.

- In Chapter 4, our proposed methods through arctan functions used the parameter \(\sigma\) to approximate the \(\ell_0\) term. We proposed to use the exponential mapping to increase \(\sigma\). However, finding procedures to tune adaptively the involved parameter for the sparser solution is left as an open problem.

- Searching for other smooth functions to replace the \(\ell_0\)-norm term rather than arctan function used in Chapter 4 and reformulate the objective function with new constraints/metrics for the hyperspectral unmixing are the potential future works.

- In Chapter 5, we observed that the energy of the SSoM for a given big library is mostly appeared in the first several subbands and we used this property in order to introduce our minimization problem and showed some improvements of the results in the point of view of accuracy. A possible future plane is to find more intrinsic features of SSoM to be included in the objective functions in order to determine more accurate results.

- In Chapter 5, we observed that the acquired optimization problems of the blind sparse unmixing schemes are sensitive to the choice of mathematical tools for
solving. Thus, finding relations between the acquired optimization problem and
the type of mathematical tools such as NMF approach, ADDM and Multilayer
NMF is another direction of future research.

• For the blind scenario of spectral unmixing methods mentioned in Chapter 5, we
may incorporate statistical features of observed data such as grouping the pixels
into the minimization problem and enhance the results as a further research
work.

• In all chapters of the thesis, we proposed our new methods for the linear system
model. Expanding the proposed idea for the nonlinear model of hyperspectral
unmixing such as bilinear model is another potential area of research for the
future work. Moreover, finding differences between the proposed hyperspectral
unmixing methods for LMM and NLMM for the same hyperspectral data set is
the other future research direction.

• Another research direction is to extend the approach of proposed unmixing
methods to the other types of data in different applications with the similar
problems, e.g., medical imaging data such as Electroencephalography (EEG)
signals and compressive sensed data.

• The computational cost is an important issue of the new proposed methods
in addition to the accuracy and reliability of the results. All the aforemen-
tioned methods in previous chapters have been examined to be accurate and we
illustrated the superiorities of proposed methods in comparison with the state-
of-the-art unmixing methods in terms of reconstruction errors such as RSNR,
AAD and SAD. However, it is definitely beneficial to verify if the new methods
are competitive with the other methods in the point of commotional complexity view. Thus, the possibility of fast implementation of the proposed methods needs further research work.


Appendix A

Proof for Chapter 3

A.1 Proof of Theorem 1

Proof. Without loss of generality, we first prove Theorem 1 without the ANC and ASC, i.e., where the members of $X_p$ are interior points of $S$. We shall show for any $\zeta > 0$ there exists a $\delta > 0$ such that $|p - q| < \delta$ yields $d(X_{p,\epsilon}, X_{q,\epsilon}) < \zeta$ where $X_{p,\epsilon}$ and $X_{q,\epsilon}$ are the set of solutions for $\nabla_x g(x, p, \epsilon) = 0$ and $\nabla_x g(x, q, \epsilon) = 0$, respectively in which $\nabla_x g(x, p, \epsilon) = \lambda_p x_{\epsilon}^{p-1} + M^T M x - M^T y = 0$, where $x_{\epsilon}^{p-1} = [\text{sign}(x_1)(|x_1| + \epsilon)^{p-1}, \ldots, \text{sign}(x_n)(|x_n| + \epsilon)^{p-1}]^T$. We prove this by contradiction by assuming that the function is not continuous, i.e. there is a $\zeta > 0$ such that for any $\delta > 0$, there always exist some $p,q > 0$ with $|p - q| < \delta$ and $d(X_{p,\epsilon}, X_{q,\epsilon}) > \zeta$. Then we draw a contradiction. From $d(X_{p,\epsilon}, X_{q,\epsilon}) > \zeta$, we conclude that either

$$\sup_{x_{p,\epsilon} \in X_{p,\epsilon}} \inf_{x_{q,\epsilon} \in X_{q,\epsilon}} \|x_{p,\epsilon} - x\|_\infty > \zeta,$$  \hfill (A.1)
A.1. PROOF OF THEOREM 1

or

$$\sup_{x_{q,\epsilon} \in X_{q,\epsilon}} \inf_{x \in X_{p,\epsilon}} ||x - x_{q,\epsilon}||_\infty > \varsigma. \quad (A.2)$$

Since the set of solutions $X_{p,\epsilon}$ and $X_{q,\epsilon}$ are closed sets, $d(X_{p,\epsilon}, X_{q,\epsilon}) > \varsigma$ means that there exist a $x_{q,\epsilon} \in X_{q,\epsilon}$ such that $||x_{p,\epsilon} - x_{q,\epsilon}||_\infty > \varsigma$ for any $x_{p,\epsilon} \in X_{p,\epsilon}$ or there must exist a $x_{p,\epsilon} \in X_{p,\epsilon}$ such that $||x_{p,\epsilon} - x_{q,\epsilon}||_\infty > \varsigma$ for any $x_{q,\epsilon} \in X_{q,\epsilon}$. Moreover, the solutions $x_{p,\epsilon} \in X_{p,\epsilon}$ and $x_{q,\epsilon} \in X_{q,\epsilon}$ must satisfy the following equations

$$\lambda_p x_{p,\epsilon} - 1 = \lambda_q x_{q,\epsilon} - 1 + M^T M x_{p,\epsilon} - M^T y = 0, \quad (A.3)$$

$$\lambda_q x_{q,\epsilon} - 1 = \lambda_q x_{q,\epsilon} - 1 + M^T M x_{q,\epsilon} - M^T y = 0. \quad (A.4)$$

By defining $h(x) = M^T M x + \lambda_p x_{p,\epsilon}^{-1}$, we have

$$h(x_{p,\epsilon}) - h(x_{q,\epsilon}) = M^T M (x_{p,\epsilon} - x_{q,\epsilon}) + \lambda_p (x_{p,\epsilon}^{-1} - x_{q,\epsilon}^{-1}). \quad (A.5)$$

Now, by subtracting (A.5) from (A.3) and adding (A.4) and taking the infinity norm of the result, we obtain

$$||h(x_{p,\epsilon}) - h(x_{q,\epsilon})||_\infty = ||\lambda_p (x_{p,\epsilon}^{-1} - x_{q,\epsilon}^{-1})||_\infty. \quad (A.6)$$

Since $h(x)$ is continuous in terms of $x$ for fixed $p$ and $\epsilon$, from $||x_{p,\epsilon} - x_{q,\epsilon}||_\infty > \varsigma$ we conclude that there exist $\eta(\varsigma)$ such that $||h(x_{p,\epsilon}) - h(x_{q,\epsilon})||_\infty > \eta(\varsigma)$ i.e. the LHS of (A.6) must be greater than $\eta(\varsigma)$. This is contradiction with the RHS of (A.6) which tends to zero as $p$ tends to $q$ because $px_{p,\epsilon}^{-1}$ is a continuous function of $p$ for all $\epsilon > 0$. 
A.1. PROOF OF THEOREM 1

The proof of the continuity of the solutions under the ASC is similar since adding the ASC involves additional Lagrangian terms in (A.3) and (A.4) which are eliminated after substraction in (A.6). The ANC can be also included in the proof via indicator functions. Thus, the proof under the ANC involves one additional Lagrangian term for each element of $\mathbf{x}$ in both (A.3) and (A.4). These terms are eliminated after substraction in (A.6). This proves the continuity over the boundary of $S$. We easily can include ANC and ASC conditions in this Theorem by adding $(\mathbf{1}^T \mathbf{x} - 1)\zeta_1 + \iota_{\mathbb{R}^n_+}(\mathbf{x})$ to $g(\mathbf{x}, p, \epsilon)$ where $\zeta_1$ is the nonnegative Lagrangian multiplier. The gradient of $g(\mathbf{x}, p, \epsilon) + (\mathbf{1}^T \mathbf{x} - 1)\zeta_1 + \iota_{\mathbb{R}^n_+}(\mathbf{x})$ has two additional terms compared with the LHS of (A.3) and (A.4) which are independent from $p$ and $\mathbf{x}$. These terms are also eliminated after substraction in (A.6).
Appendix B

Proof for Chapter 4

B.1 Proof of Theorem 2

Proof. We shall show that, for any $\epsilon > 0$, there exists a $\delta > 0$, such that $|\sigma - \hat{\sigma}| < \delta$ yields $d(X_\sigma, X_{\hat{\sigma}}) < \epsilon$, where $X_\sigma$ and $X_{\hat{\sigma}}$ are the set of solutions for $\nabla_x f(x, \sigma) = 0$ and $\nabla_x f(x, \hat{\sigma}) = 0$, respectively. We prove this by contradiction. Hence, we assume that the function is not continuous, i.e. there is a $\epsilon > 0$, such that for any $\delta > 0$, there always exist some $\sigma, \hat{\sigma} > 0$ with $|\sigma - \hat{\sigma}| < \delta$ and $d(X_\sigma, X_{\hat{\sigma}}) > \epsilon$. Then, we draw a contradiction.

From $d(X_\sigma, X_{\hat{\sigma}}) > \epsilon$, we conclude that either:

$$\sup_{x_{\sigma} \in X_\sigma} \inf_{x_{\hat{\sigma}} \in X_{\hat{\sigma}}} ||x_{\sigma} - x_{\hat{\sigma}}||_{\infty} > \epsilon,$$

or:

$$\sup_{x_{\hat{\sigma}} \in X_{\hat{\sigma}}} \inf_{x_{\sigma} \in X_\sigma} ||x_{\sigma} - x_{\hat{\sigma}}||_{\infty} > \epsilon.$$
Since the set of solutions $X_\sigma$ and $\hat{X}_\sigma$ are closed sets, $d(X_\sigma, \hat{X}_\sigma) > \epsilon$ yields that there must exist a $x_\sigma \in X_\sigma$, such that $||x_\sigma - x_\hat{\sigma}||_\infty > \epsilon$ for any $x_\sigma \in X_\sigma$, or there must exist a $x_\hat{\sigma} \in X_\hat{\sigma}$, such that $||x_\sigma - x_\hat{\sigma}||_\infty > \epsilon$ for any $x_\hat{\sigma} \in \hat{X}_\sigma$. Moreover, the solutions $x_\sigma \in X_\sigma$ and $x_\hat{\sigma} \in \hat{X}_\sigma$ must satisfy the following equations:

\begin{align}
\lambda v(x_\sigma, \sigma) + M^T M x_\sigma - M^T y &= 0, \quad (B.3) \\
\lambda v(x_\hat{\sigma}, \hat{\sigma}) + M^T M x_\hat{\sigma} - M^T y &= 0. \quad (B.4)
\end{align}

By defining $h(x) = M^T M x + \lambda v(x_\sigma, \sigma)$, we have:

\begin{align}
h(x_\sigma) - h(x_\hat{\sigma}) &= M^T M (x_\sigma - x_\hat{\sigma}) + \lambda (v(x_\sigma, \sigma) - v(x_\hat{\sigma}, \sigma)). \quad (B.5)
\end{align}

Now, by subtracting equation (B.5) from equation (B.3), adding equation (B.4) and taking the infinity norm of the result, we obtain:

\begin{align}
||h(x_\sigma) - h(x_\hat{\sigma})||_\infty &= ||\lambda (v(x_\sigma, \sigma) - v(x_\hat{\sigma}, \sigma))||_\infty. \quad (B.6)
\end{align}

Since $h(x)$ and $h^{-1}(x)$ are continuous in terms of $x$ for fixed $\sigma$, from $||x_\sigma - x_\hat{\sigma}||_\infty > \epsilon$, we conclude that there exist $\eta(\epsilon)$, such that $||h(x_\sigma) - h(x_\hat{\sigma})||_\infty > \eta(\epsilon)$, i.e., the LHS of equation (B.6) must be greater than $\eta(\epsilon)$. This is a contradiction with the RHS of equation (B.6), which tends to zero as $\sigma$ tends to $\hat{\sigma}$, since $v(x_\sigma, \sigma)$ is a continuous function with respect to $\sigma$ for the fixed value of $x$.

To prove the continuity of the solutions under the ASC, we have to add an additional Lagrangian term using the indicator functions in equations (B.3) and (B.4) that
are eliminated after subtraction in equation (B.6). The proof under the nonnegativity constraints is also similar, since representing the ANC via indicator functions involves one additional Lagrangian term for each element of $x$ in both equations (B.3) and (B.4). These additional terms are also omitted after subtraction in equation (B.6). Thus, the proof of the continuity over the boundary of $S$ is completed. \qed
Appendix C

Proof for Chapter 5

C.1 Proof of Proposition 2

Proof. Without loss of generality, we first consider only the fractional abundances terms, i.e., $X$, in (5.20) and vectorize it as follows:

$$f_c(\tilde{x}) = \frac{1}{2} ||\tilde{y} - \tilde{A}\tilde{x}||^2 + \alpha \sum \left( \arctan(\omega\tilde{x}) / \arctan(\omega) \right),$$

(C.1)

where $\tilde{x} = \text{vec}(X)$ and $\tilde{y} = \text{vec}(Y)$ are $NP \times 1$ and $LP \times 1$ column vectors, $\tilde{A} = I_P \otimes A$ is $LP \times NP$ matrix and $I_P$ is the identity matrix with the size of $P$, $\otimes$ and $\odot$ are the kronecker product and the element-wise multiplication, respectively, and arctan function operates element-wisely.

Then, we follow the similar approach used in the Expectation Maximization (EM) algorithm and later in [7, 5] to make the following auxiliary function for (C.1):

$$G(\tilde{x}, \tilde{x}^t) = f_c(\tilde{x}^t) + (\tilde{x} - \tilde{x}^t)^T \nabla f_c(\tilde{x}^t) + \frac{1}{2}(\tilde{x} - \tilde{x}^t)^T K(\tilde{x}^t)(\tilde{x} - \tilde{x}^t)$$

(C.2)
where $K(\bar{x}^t)$ is the diagonal matrix as

$$K(\bar{x}^t) = \text{diag}\left(\left(\tilde{A}^T\tilde{A}\bar{x}^t + \alpha \omega./\arctan(\omega) \cdot \right)
\left(1 + (\omega \cdot \bar{x}^t)^2\right)\right)./\bar{x}^t)$$ \hspace{1cm} (C.3)

satisfying both conditions $G(\bar{x}, \bar{x}) = f_c(\bar{x})$ and $G(\bar{x}, \bar{x}^t) \geq f_c(\bar{x})$ such that

$$\bar{x}^{t+1} = \arg\min_{\bar{x}} G(\bar{x}, \bar{x}^t).$$ \hspace{1cm} (C.4)

The Taylor expansion of (C.1) is

$$f_c(\bar{x}) = f_c(\bar{x}^t) + (\bar{x} - \bar{x}^t)^T\nabla f_c(\bar{x}^t) + \frac{1}{2}(\bar{x} - \bar{x}^t)^T
\left(\tilde{A}^T\tilde{A} - \frac{2\alpha \omega^3}{\arctan(\omega)} \text{diag}\left(\bar{x}^t./\left(\left(1 + \omega^2\bar{x}^t^2\right)\right)\right)\right)(\bar{x} - \bar{x}^t)$$ \hspace{1cm} (C.5)

where the higher order of the Lagrange reminder is omitted. To satisfy the condition $G(\bar{x}, \bar{x}^t) \geq f_c(\bar{x})$, we must have

$$(\bar{x} - \bar{x}^t)^T\left(K_1(\bar{x}^t) + \alpha \text{diag}\left(\omega./\arctan(\omega)\cdot\left(1 + (\omega \cdot \bar{x}^t)^2\right)\right)\right)(\bar{x} - \bar{x}^t) \geq 0,$$ \hspace{1cm} (C.6)

where $K_1(\bar{x}^t) = \text{diag}\left(\tilde{A}^T\tilde{A}\bar{x}^t./\bar{x}^t\right) - \tilde{A}^T\tilde{A}$.

The involved terms in the diag operation in (C.6) are nonnegative due to nonnegativity of $\bar{x}^t$. Besides, $K_1(\bar{x}^t)$ was proved to be the positive semidefinite in [7]. Thus, the left-hand-side (LHS) terms in (C.6) must be nonnegative.
Then, by taking derivative of $G(\tilde{x}, \tilde{x}^t)$ in (C.2) with respect to $\tilde{x}$ and substituting into (C.4), we obtain the updating rule as $\tilde{x}^{t+1} = \tilde{x}^t - K^{-1}(\tilde{x}^t)\nabla f_c(\tilde{x}^t)$. This is simplified to

$$\tilde{x}^{t+1} = \tilde{x}^t * \tilde{A}^T \tilde{y} / \left( \tilde{A}^T \tilde{A} \tilde{x}^t + \alpha \frac{\omega}{\arctan(\omega) / (1 + (\omega \tilde{x}^t)^2)} \right)$$

which is the vectorization form of the updating rule in (5.24).

The proof for the updating rule in (5.23) is similar and we can consider the corresponding terms contain only the spectral signatures library, i.e., $A$, in (5.20). It should be noted that $\text{vec}(AX) = (X^T \otimes I_L)\text{vec}(A)$ and the rest of the proof follows the similar way.

\qed