



Libraries and Learning Services

# University of Auckland Research Repository, ResearchSpace

## Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognize the author's right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

## General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the [Library Thesis Consent Form](#) and [Deposit Licence](#).

# **Compressive Hyperspectral Imaging:**

## **Acquisition, Recovery, and Unmixing**

---

William Yen-Lei LEE

A thesis submitted in fulfilment of the requirements for the degree of  
Doctor of Philosophy in Electrical and Electronic Engineering,  
The University of Auckland, 2015



# Abstract

Hyperspectral imaging provides non-invasive, quantitative, characterization of materials' compositions via their spectral responses, demonstrating prospective developments in a wide range of scientific fields where contaminations from external sources could have a sizeable impact on what follows. The substantial overheads associated in handling the inherently massive volumes of hyperspectral data have unfortunately limited the number of practical applications by the stringent economic trade-offs. Nonetheless, recent advances in the field of compressive sensing (CS) have demonstrated great promise in alleviating the challenges faced in dealing with large data sets. In favour of this, the focus of this thesis is on the development of methods for integrating CS into hyperspectral imaging.

A new pragmatic acquisition strategy based on CS is considered. Instead of directly extending CS acquisition to three-dimensions, where the sensing methods are currently impractical or unknown, the new strategy exploits the various acquisition modalities available in existing systems to compressively sense hyperspectral imagery using a hybrid of two-dimensional (2D) measurements, captured via the different modes. Using convex optimization, the necessary algorithm for recovery from the resulting CS measurements is developed and is shown to be easily integrable with existing recovery models. A detailed analysis revealed that practical single-mode 2D acquisitions that use a unique sensing matrix for hyperspectral CS fails to facilitate models that exploit sparsity beyond the measured dimensions, while for the same setup, the multi-modal approach achieved superior reconstruction, leading to an improvement between 3 to 14 dB in the signal-to-reconstruction error of real scenes over state-of-the-art recovery models.

Two decoding models for the reconstruction of compressively sensed hyperspectral data are proposed which linearly combine the total variation (TV) of each spectral band with smoothness-inducing priors along each voxel of the hyperspectral image to achieve accurate reconstructions. The models are posed as convex optimization problems in the form of an  $l_1$ - $l_2$  gradient-based minimization, and an  $l_1$ - $l_1$  sparsity-based minimization to capture both the spatial uniformity exhibited in the bands and the spectral smoothness inherent in the voxels. Simulations with multiple datasets were conducted and comparisons were made against similar methods from the existing literature. The models were shown to be effective in all tested data with the gradient-based model outperforming by at least 6 dB in the reconstruction error metric.

Finally, the problem of linear unmixing directly from CS measurements is considered. A blind unmixing technique is presented that forgoes strict, unrealistic, assumptions about the underlying data imposed in prior works including the absence of spectral mixing, and the absolute knowledge of endmember presence. The technique uses the smoothness characteristic exhibited by, and the external geometry of, the participating endmembers under the linear mixing model (LMM) for effective unmixing in a highly mixed scenario. The resulting algorithm alternately minimizes the TV of the abundances under the LMM, and the second central moment of the endmembers in conjunction to the their relative Euclidean distances. The technique was shown to be promising through experimentation on simulated dataset with various controllable parameters.

While a number of challenges still remains to be further resolved in each area, the progress made in this thesis represents a significant step towards bridging the gap between hyperspectral imaging and the numerous applications that could benefit from it.

*To my parents.*

*Thank you for your unconditional support and patience.*



# Acknowledgements

First and foremost, I wish to express my deepest gratitude to my supervisor, Dr. Mark Andrews for his invaluable support throughout this journey. Your expertise and humour in academia, research, and life have greatly influenced my beliefs on just about everything. I am grateful for the freedom you have entrusted me with, enough to let me follow my nose without ever feeling leashed. Thank you for your constant optimism in my work, for your enthusiasm in reading my thesis (despite the frequent misplacement of it), and for the fruitful distractions that never ceased to put a smile on my face.

I would like to thank my co-supervisor, Dr. Catherine Watson for her advice and support. Your encouragements have been invaluable to me. I apologize for the lack of correspondence in times of great despair.

I would also like to thank the signal processing research group and the postgraduate community for your friendship during my PhD studies. In particular, I am grateful to Sahil Jain for his prompt troubleshooting advice with LaTeX; to Hui Zhi (Zak) Beh, for regularly checking up on my progress and timely reminders to my writings; to Timothy Roper, Robert Dunn, and Duncan Eason, for their academic companionship under the same roof. I extend my thanks to my friend David Huang for his persistent company at the gym every day, without which I would have most likely given up my well-being to writing this thesis. Finally, I am thankful to Cindy Wang for her kind support and for always believing in me even when I doubted myself. You have brought the full spectrum of colours into my otherwise monochromatic life.

Last, but certainly not least, I would like to express my sincere gratitude to my parents for

their patience throughout my studies, for their understanding during rough times, and most of all, for their unconditional support over the years. This endeavour would not have been possible without them and I therefore dedicate this thesis to my parents: I appreciate all the sacrifices that you have made for me, and I am forever grateful.

William Lee

December 2015

## Co-Authorship Form

This form is to accompany the submission of any PhD that contains published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements. Co-authored works may be included in a thesis if the candidate has written all or the majority of the text and had their contribution confirmed by all co-authors as not less than 65%.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 4, Section 4.1, 4.2.3, Pages 92-95,110-116 - Extracted from W. Y.-L. Lee, and M. Andrews, "Blind spectral unmixing for compressive hyperspectral imaging of highly mixed data", IEEE International Conference on Image Processing, 2014

Nature of contribution by PhD candidate	Development of methodologies, discussion of ideas, literature survey
Extent of contribution by PhD candidate (%)	75%

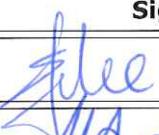
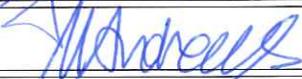
### CO-AUTHORS

Name	Nature of Contribution
Mark Andrews	Discussion of ideas, guidance, proofreading

### Certification by Co-Authors

The undersigned hereby certify that:

- ❖ the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- ❖ that the candidate wrote all or the majority of the text.

Name	Signature	Date
William Yen-Lei Lee		16/12/2015
Mark Andrews		16 DEC 2015

## Co-Authorship Form

This form is to accompany the submission of any PhD that contains published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements. Co-authored works may be included in a thesis if the candidate has written all or the majority of the text and had their contribution confirmed by all co-authors as not less than 65%.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 2, Section 2.3-2.6, Pages 37-62 - Extracted from W. Y.-L. Lee, and M. Andrews, "Multimodal Compressive Hyperspectral Imaging", Submitted for publication, 2015.

Nature of contribution by PhD candidate	Development of methodologies, discussion of ideas, literature survey
Extent of contribution by PhD candidate (%)	75%

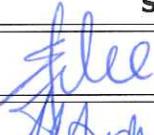
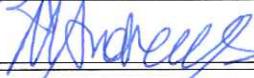
### CO-AUTHORS

Name	Nature of Contribution
Mark Andrews	Discussion of ideas, guidance, proofreading

### Certification by Co-Authors

The undersigned hereby certify that:

- ❖ the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- ❖ that the candidate wrote all or the majority of the text.

Name	Signature	Date
William Yen-Lei Lee		16/12/2015
Mark Andrews		16 DEC 2015

## Co-Authorship Form

This form is to accompany the submission of any PhD that contains published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements. Co-authored works may be included in a thesis if the candidate has written all or the majority of the text and had their contribution confirmed by all co-authors as not less than 65%.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 3, Section 3.1.1-3.2 Pages 70-75 - Extracted from W. Y.-L. Lee, D. Eason, and M. Andrews, "Compressive Hyperspectral Image Sensing Restoration via Joint TV-L1 Regularisation", Electronics New Zealand Conference, 2012

Nature of contribution by PhD candidate	Development of methodologies, discussion of ideas, literature survey
Extent of contribution by PhD candidate (%)	65%

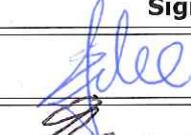
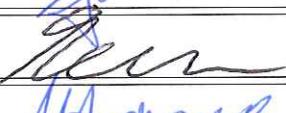
### CO-AUTHORS

Name	Nature of Contribution
Duncan Eason	Development of methodologies, discussion of ideas, guidance, literature survey
Mark Andrews	Development of methodologies, discussion of ideas, guidance, proofreading

### Certification by Co-Authors

The undersigned hereby certify that:

- ❖ the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- ❖ that the candidate wrote all or the majority of the text.

Name	Signature	Date
William Yen-Lei Lee		16/12/2015
Duncan Eason		16/12/2015
Mark Andrews		16 DEC 2015

## Co-Authorship Form

This form is to accompany the submission of any PhD that contains published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements. Co-authored works may be included in a thesis if the candidate has written all or the majority of the text and had their contribution confirmed by all co-authors as not less than 65%.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 3, Section 3.1.1-3.2 Pages 70-75 - Extracted from D. Eason, W. Y.-L. Lee, and M. Andrews, "Compressively Sensed Hyperspectral Image Recovery using Total Variation Minimisation by Approximation", Electronics New Zealand Conference, 2012

Nature of contribution by PhD candidate	Development of methodologies, discussion of ideas, literature survey,
Extent of contribution by PhD candidate (%)	33.3%

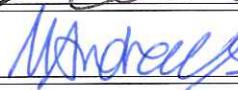
### CO-AUTHORS

Name	Nature of Contribution
Duncan Eason	Development of methodologies, discussion of ideas, guidance, literature survey
Mark Andrews	Development of methodologies, discussion of ideas, guidance, proofreading

### Certification by Co-Authors

The undersigned hereby certify that:

- ❖ the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- ❖ that the candidate wrote all or the majority of the text.

Name	Signature	Date
William Yen-Lei Lee		16/12/2015
Duncan Eason		16/12/2015
Mark Andrews		16 DEC 2015

# Contents

<b>Abstract</b>	ii
<b>Acknowledgements</b>	vi
<b>Acronyms</b>	xv
<b>Notations and Symbols</b>	xvii
<b>1 Introduction</b>	1
1.1 Compressive Sensing . . . . .	4
1.1.1 Main Results . . . . .	8
1.1.2 Matrices Satisfying the RIP . . . . .	11
1.2 Hyperspectral Imaging . . . . .	13
1.2.1 Spectral Image Capture . . . . .	16
1.2.2 Spectral Mixing . . . . .	19
1.3 Direction and Original Contributions . . . . .	22
1.3.1 Summary of Contributions . . . . .	24
1.3.2 Publications . . . . .	24
<b>2 Compressive Hyperspectral Image Acquisition</b>	27
2.1 Notations . . . . .	28
2.2 Survey of CS Hyperspectral Imaging Techniques . . . . .	29
2.2.1 The Single-Pixel Hyperspectral Camera . . . . .	30
2.2.2 Compressive Pushbroom/Whiskbroom Hyperspectral Imagers . . . . .	32

2.2.3	Coded Aperture Snapshot Spectral Imager . . . . .	35
2.2.4	Compressive Hyperspectral Imaging by Separable Spatial and Spectral Operators . . . . .	37
2.3	Discussion of Limitations and Difficulties . . . . .	39
2.4	Multi-modal Compressive Hyperspectral Imaging . . . . .	40
2.4.1	Problem Formulation . . . . .	42
2.5	Multi-modal CoHSI Reconstruction Algorithm . . . . .	44
2.5.1	Alternating Direction Method of Multiplier . . . . .	45
2.5.2	Multi-modal Recovery via Augmented Lagrangian . . . . .	47
2.6	Simulations and Performance Analysis . . . . .	55
2.6.1	Experimental Results with Synthetic Hyperspectral Data . . . . .	56
2.6.2	Experimental Results with Real Hyperspectral Data . . . . .	64
2.7	Conclusions . . . . .	65
<b>3</b>	<b>Compressive Hyperspectral Image Recovery</b>	<b>71</b>
3.1	Existing CoHSI Reconstruction Models . . . . .	72
3.1.1	Channel-by-Channel Reconstruction . . . . .	72
3.1.2	Joint Spatial-Spectral Reconstruction . . . . .	75
3.1.3	Low-Rank Prior Reconstruction . . . . .	76
3.2	CoHSI Recovery via Joint Spatio-Spectral Prior . . . . .	77
3.3	Simulations and Performance Analysis . . . . .	82
3.3.1	Experimental Results with Synthetic Hyperspectral Data . . . . .	82
3.3.2	Experimental Results with Real Hyperspectral Data . . . . .	85
3.4	Conclusions . . . . .	89
<b>4</b>	<b>Compressive Hyperspectral Unmixing</b>	<b>93</b>
4.1	Prior Art . . . . .	95
4.1.1	Compressive Unmixing with Endmember Priors . . . . .	95
4.1.2	Blind Compressive Unmixing . . . . .	98
4.2	A Blind Compressive HSI Unmixing Scheme . . . . .	99

4.2.1	Additivity Regularized Least-Squares . . . . .	99
4.2.2	Additivity-TV Regularized Least-Squares . . . . .	108
4.2.3	Joint Endmember-Abundance Regularized Least-Squares . . . . .	113
4.3	Conclusions . . . . .	120
<b>5</b>	<b>Future Work</b>	<b>125</b>
5.1	Acquisition . . . . .	125
5.2	Recovery . . . . .	128
5.3	Unmixing . . . . .	130
<b>6</b>	<b>Conclusions</b>	<b>133</b>
	<b>Bibliography</b>	<b>137</b>

# Acronyms

<b>ADMM</b>	Alternating Direction Method of Multipliers
<b>BP</b>	Basis Pursuit
<b>BPDN</b>	Basis Pursuit De-Noising
<b>CASSI</b>	Coded Aperture Snapshot Spectral Imager
<b>CFA</b>	Colour Filter Array
<b>CS</b>	Compressive Sensing
<b>DCT</b>	Discrete Cosine Transform
<b>DD-CASSI</b>	Dual-Disperser Coded Aperture Snapshot Spectral Imager
<b>DFT</b>	Discrete Fourier Transform
<b>DMD</b>	Digital Micro-mirror Device
<b>EM</b>	Electromagnetic
<b>FFT</b>	Fast Fourier Transform
<b>FOV</b>	Field of View
<b>HALS</b>	Hierarchical Alternating Least Squares
<b>HSI</b>	Hyperspectral Image
<b>i.i.d.</b>	Independently and Identically Distributed
<b>LMM</b>	Linear Mixing Model
<b>MM</b>	Method of Multiplier
<b>MuRecAL</b>	Multi-modal Recovery via Augmented Lagrangian
<b>PCA</b>	Principal Component Analysis
<b>RIP</b>	Restricted Isometry Property
<b>SD-CASSI</b>	Single-Disperser Coded Aperture Snapshot Spectral Imager

<b>SNR</b>	Signal-to-Noise Ratio
<b>SPC</b>	Single-Pixel Camera
<b>SPHC</b>	Single-Pixel Hyperspectral Camera
<b>SRE</b>	Signal-to-Reconstruction Error Ratio
<b>STU</b>	Sum-to-Unity
<b>SVD</b>	Singular Value Decomposition
<b>TV</b>	Total Variation

# Notations and Symbols

The mathematical notations used throughout this thesis are as follow:

Scalars are represented using unembellished letters, e.g.,  $n$ ,  $K$ ,  $\lambda$ . Vectors are represented using lower-case bold letters, e.g.,  $\phi$ ,  $e$ . If an individual element of a vector is required, it is indexed by the vector in bracket with a lower subscript, e.g.,  $(e)_1$  represents the first element of vector  $e$ . Unless specified otherwise, all vectors are by default column vectors. If a row vector is required, it is written as the transpose, e.g.,  $\phi^\top$ .

Matrices are represented using upper-case bold letters, e.g.,  $A$ . Columns of a matrix are represented using its lower-case equivalent vectors, indexed by lower subscripts, e.g.,  $a_1$  denotes the first column of  $A$ . Similarly, rows of a matrix are denoted by the corresponding subscripted row vectors, e.g.,  $a_1^\top$  is the first row of  $A$ . When manipulating matrices with complex elements, the conjugate (or Hermitian) transpose is denoted by  $A^*$ . If an individual element of a matrix is required, it is represented as a scalar with two comma separated subscript indices so that  $a_{i,j}$  is the element in the  $i^{th}$  row and  $j^{th}$  column of  $A$ .

$\mathbb{R}$	Set of real numbers
$\mathbf{1}$	Column vector of ones
$\mathbf{I}$	Identity matrix
$\mathbf{P}$	Permutation matrix
$\otimes$	Kronecker product
$\nabla$	Gradient operator
$\ \cdot\ _p$	$p$ -norm
$\ \cdot\ _F$	Frobenius norm
$\ \cdot\ _*$	Nuclear (or trace) norm
$\text{vec}(\cdot)$	Vectorization operator
$\text{vec}^{-1}(\cdot)$	Inverse vectorization operator
$\mu(\cdot)$	Mutual coherence
$\text{FFT3}(\cdot)$	Forward 3D fast Fourier transform
$\text{IFFT3}(\cdot)$	Inverse 3D fast Fourier transform
$\text{prox}_{\square, \rho}(\cdot)$	Proximity operator of $\square$ with penalty parameter $\rho$
$\text{rank}(\cdot)$	Rank of a matrix
$L$	Lipschitz constant
$n, m$	General-purpose set length
$n_e$	Number of endmembers
$n_x$	Number of pixels along the spatial $x$ -direction
$n_y$	Number of pixels along the spatial $y$ -direction
$n_\lambda$	Number of measured wavelength
$\gamma, \alpha, \beta$	Regularization parameters
$\rho$	ADMM penalty parameter or maximal abundance value
$\mathbf{a}_i \in \mathbb{R}^{n_x n_y}$	$n_x \times n_y$ -pixel abundance vector of material $i$
$\mathbf{e}_i \in \mathbb{R}^{n_\lambda}$	Endmember spectrum of material $i$ with $n_\lambda$ bands
$\mathbf{y} \in \mathbb{R}^m$	CS measurement vector of $m$ measurements
$\boldsymbol{\alpha} \in \mathbb{R}^n$	Vector in a sparse transform
$\boldsymbol{\phi}_i^\top \in \mathbb{R}^n$	$i^{th}$ CS measurement basis

$\psi_i \in \mathbb{R}^n$	$i^{th}$ orthonormal basis vector of length $n$
$A \in \mathbb{R}^{n_x n_y \times n_e}$	Abundance matrix of $n_e$ materials with $n_x \times n_y$ -pixels
$E \in \mathbb{R}^{n_\lambda \times n_e}$	Endmember matrix of $n_e$ materials with $n_\lambda$ bands
$Y \in \mathbb{R}^{m \times n_\lambda}$	Matrix of CS measurements of $n_\lambda$ bands
$\Phi \in \mathbb{R}^{m \times n}$	CS matrix of $m$ measurements basis for a signal of length $n$
$\Psi \in \mathbb{R}^{n \times n}$	Orthonormal sparcifying basis
$\hat{\square}$	Estimate of parameter $\square$

# Chapter 1

## Introduction

*“The sampling theorem is wrong! Not literally wrong, but psychologically wrong. More precisely, it engenders the psychological expectation that we need very large number of samples in situations where we need very few.”*

– David L. Donoho

The well-known Shannon-Nyquist sampling theorem demonstrates that any continuous-time (or space) band-limited signals (e.g., images, videos, audio signals) can be *exactly* recovered from a set of equally spaced samples taken at above the *Nyquist rate* of twice the highest frequency present in the signal. Straightforward and precise, this basic principle, dated as far back as 1915, underpins the theoretical foundation behind almost all digital signal processing applications (e.g., communication systems, imaging devices, audio electronics) and ultimately resulted in the digital revolution maturing our world into the information age. However, a natural tendency followed with such fruitful transition, is the ever growing demand for higher fidelity and resolution of data, and ironically, the celebrated sampling theorem has also become an obstacle faltering the developments.

Meanwhile, multi-channel signal applications are becoming eminent in a variety of scientific fields including audio processing, biomedical science, computer vision, remote sensing, wireless communications etc. The increase in diversity of observations from different points of view brings with it several advantages that many of these areas have long sought for, such

as the ability for source localization and tracking (e.g., in speech processing, neuroimaging, surveillance tracking), source separation (e.g., in spectral imaging, acoustic source separation, medical imaging), and signal interpretation and enhancement (e.g., in active noise control, multimodal image registration, 3D image modelling). But nothing comes without its price, a common ground in all these applications is the necessity of a large number of sensors for acquiring the required information, and along with it entails a tremendously large and increasing flow of data. Needless to say, the size of these data, imposed by non-other than the sampling theorem itself, has become a demanding challenge on the available technologies for both hardware (acquisition hardware, transmission channels, storage devices etc.) and software (processing, analysis, visualisation etc.).

For example, the Large Acoustic Data Array (LOUD) project [1] uses an array of 1020 microphones to capture a large volume of multi-channel audio data (approximately 50 MB of data per second) in an attempt to reduce ambient noise and crosstalk in ubiquitous computing environments. The bandwidth and computational power necessary to process such streaming data has pushed the limits of traditional computer architecture. On the other hand, in clinical studies, multi-channel patient recordings (e.g., electroencephalogram (EEG), heart rate, blood oxygen level, electrocardiogram (EKG) etc.) are constantly monitored for effective patient care and research in epilepsy. A 24-hour recording of these multi-channel recordings for just a single patient can easily generate 5 to 10 GB of data [2], creating significant data management challenges in terms of efficient storage, visualisation and analysis [3, 4]. Similarly, in hyperspectral imaging, hundreds of images are captured from a scene at various wavelengths in order to provide comprehensive material characterisations of the scene [5–7]. To do this, NASA’s Airborne Visible InfraRed Imaging Spectrometer (AVIRIS) captures 224 spectral bands for each hyperspectral image consisting of nearly 1 GB of data<sup>1</sup>. Other areas such as geophysical data analysis, astronomy, and quantum physics just to name a few, are also faced with the problem of transmitting, processing and storing high-dimensional data. Fortunately though, many natural signals typically have a concise representation in certain

---

<sup>1</sup>A standard format of 677 (columns)  $\times$  512 (lines)  $\times$  8 (bytes of quantization accuracy in double precision format)  $\times$  224 (spectral bands) gives approximately 600 MB of data, and up to 1 GB after taking into account of navigation, engineering, and calibration data.

transform bases (e.g., Fourier basis, Wavelet basis, etc.). This means that these signals, when expressed in the appropriate basis, are *sparse* and can therefore be represented accurately by only a small fraction of the data size implied by the Nyquist criterion. In this regard, a plethora of work has been put into exploiting such intra-channel and inter-channel structure in multi-channel signals for dimensionality reduction and data compression, achieving tremendous reduction in data size by orders of magnitudes. For example, JPEG2000 takes advantage of the sparseness of natural images in the Wavelet domain and encodes just those few non-vanishing Wavelet coefficients [8]. While Principal Component Analysis (PCA) has been used extensively for removing inter-channel linear dependencies in high-dimensional data [9]. As a result, for multi-channel signals such as hyperspectral imagery, compression paradigms have been proposed that couple PCA with JPEG2000 for efficient spectral decorrelation and spatial encoding [10, 11].

Despite the success of these compression techniques over the past decades, exploiting underlying structures in signals at post-acquisition seems rather short-lived in the grand scheme of things. While this lowers the burden on transmission and storage, the complexity remains unchanged at the acquisition and processing ends. In the context of hyperspectral imaging, since capturing spectral information outside the visible spectrum requires exotic photodiodes which are very costly, incorporating a large number of them, as dictated by the sampling theorem, to make a high-resolution hyperspectral imager is therefore extremely expensive. As such, while many research have provided promising results in using spectral imaging for various terrestrial-based applications (e.g., quality control [12, 13], forensic analysis [14, 15], medical diagnosis [16, 17], etc.), most existing applications are limited to mapping terrestrial resources where the economic benefits outweigh the cost in acquiring these images.

Such economic restrictions along with the ubiquitous redundancies in the ever-increasing amounts of data raise a very natural question regarding our sampling paradigm: why make such costly effort in acquiring the data when only a tiny fraction of it is kept after yet another computationally expensive compression step?

This thesis examines the various aspects (acquisition, recovery and processing) needed for incorporating the theory of *Compressive Sensing (CS)* – a novel data acquisition paradigm

which overcomes the drawbacks of traditional method – into hyperspectral imaging, aiming at opening up the possibility of simpler, cheaper, and more efficient alternatives to existing hyperspectral systems within the broader context of bridging the gap between hyperspectral imaging and the numerous fields of science that would benefit from it.

An introduction to compressive sensing is presented in this chapter, followed by a review of hyperspectral imaging. A summary of the original contributions within this thesis is then provided along with an outline of the remaining chapters.

## 1.1 Compressive Sensing

Compressive sensing (also known as compressed sensing, or compressive sampling) has emerged out from recent advances in applied mathematics and signal processing as an alternative to the classical Shannon-Nyquist sampling. The novelty of CS lies in combining conventional sampling and compression together in order to acquire (and represent) data at a rate far lower than that imposed by the Nyquist criterion [18–21]. Instead of taking  $n$  periodic point evaluations to discretize the data into a vector  $\mathbf{x} \in \mathbb{R}^n$ , CS takes  $m$  linear measurements,  $y_i$ , as an inner product of  $\mathbf{x}$  with a measurement basis  $\phi_i \in \mathbb{R}^n$  giving the measurement vector  $\mathbf{y} \in \mathbb{R}^m$ :

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} \phi_1^\top \\ \phi_2^\top \\ \vdots \\ \phi_m^\top \end{bmatrix} \mathbf{x} = \Phi \mathbf{x} \quad (1.1)$$

where  $\Phi \in \mathbb{R}^{m \times n}$  is known as the measurement matrix. If  $\mathbf{x}$  can be represented in a certain orthonormal basis  $\Psi = [\psi_1 \dots \psi_n] \in \mathbb{R}^{n \times n}$  as  $\mathbf{x} = \Psi \alpha$ , then (1.1) can also be written as

$$\mathbf{y} = \Phi \mathbf{x} = (\Phi \Psi) \alpha = \Theta \alpha. \quad (1.2)$$

If  $\Phi$  is also orthonormal (i.e.,  $m = n$  with  $\Phi^* \Phi = I$ ), and therefore invertible (and well-conditioned), it is then trivially possible to recover  $\alpha$  (and equivalently,  $\mathbf{x}$ ) from the measurements (1.2) by simply applying  $(\Phi \Psi)^* = \Theta^*$  to the measurement vector  $\mathbf{y}$ .

So far, this is classical, and there is no benefit in doing so, nor is it better than taking point samples of  $\mathbf{x}$  in the first place (i.e., with  $\Phi = \mathbf{I}$ ). What is remarkable from CS results, on the other hand, is that if  $\boldsymbol{\alpha}$  contains no more than  $K$  non-zero coefficients (i.e.,  $\|\boldsymbol{\alpha}\|_0 \leq K \ll n^2$  and we say  $\boldsymbol{\alpha}$  is at most  $K$ -sparse), then it is suffice to only keep a little more than  $K$  elements of  $\mathbf{y}$  (i.e.,  $K < m \ll n$ ) in order to accurately and stably reconstruct  $\boldsymbol{\alpha}$  from (1.2) provided the measurement basis  $\Phi$  and the basis of representation  $\Psi$  exhibit a low *mutual coherence*  $\mu(\Phi\Psi)$ :

$$1 \leq \mu(\Phi\Psi) \stackrel{\text{def}}{=} \sqrt{n} \max_{i,j} |\phi_i^\top \psi_j| \leq \sqrt{n}. \quad (1.3)$$

Complete coherence,  $\mu(\Phi\Psi) = \sqrt{n}$ , is obtained when at least one of  $\phi_i$  is identical to a  $\psi_i$ , for example at the extreme case when  $\Phi^*$  is a subset of  $\Psi$ . While one can easily check that maximal incoherence,  $\mu(\Phi\Psi) = 1$ , can be obtained, for example, when  $\Phi = \mathbf{I}$ , the identity, with  $\Psi$  as the Fourier basis or the Noiselet basis [22]. One striking fact is that if  $\Phi$  is a random matrix with each element drawn Independently and Identically Distributed (i.i.d.) from a sub-Gaussian<sup>3</sup> distribution [23] (e.g., Gaussian, Bernoulli, etc.), then it will exhibit a very low coherence with any fixed basis  $\Psi$  with  $\mu(\Phi\Psi) \leq \sqrt{c \log n} \ll \sqrt{n}$  for some small constant  $c$  depending on the distribution used [24, 25]. This suggests that if sampling with incoherent basis is ideal, then random matrices are universally good CS measurement matrices irrespective of  $\Psi$ .

In essence, the notion of incoherence specifies that signals having a sparse representation in a defined basis must be spread out in the basis in which they are acquired. Figure 1.1 illustrates this point: if  $\Phi^*$  is a subset of  $\Psi$  (completely coherent) as in Figure 1.1(a), each  $\phi_i$  only reveals how much of *one* basis element  $\psi_i$  is present in the signal, therefore excluding the  $i^{th}$  measurement is taking a very risky gamble that  $\alpha_i = 0$ , and so one would end up requiring all  $n$  measurements. On the other extreme, if  $\Phi$  and  $\Psi$  are completely uncorrelated as in Figure 1.1(b), then each measurement basis  $\phi_i$  would tell us a little bit of information about every component of  $\boldsymbol{\alpha}$  and consequently one would, in principle, need far fewer measurements

<sup>2</sup>The  $l_0$  “norm”,  $\|\mathbf{x}\|_0$ , is the number of non-zero coefficients of the vector  $\mathbf{x}$

<sup>3</sup>A random variable  $X$  is said to be sub-Gaussian if its moment-generating function is bounded by that of a zero-mean Gaussian, i.e.,  $\mathbb{E}\{\exp(Xt)\} \leq \exp(c^2 t^2 / 2)$  for all  $t \in \mathbb{R}$  and some positive constant  $c$ .  $\mathbb{E}\{\exp(Xt)\}$  is the moment-generating function of  $X$  since differentiating  $\mathbb{E}\{\exp(Xt)\} = 1 + t\mathbb{E}\{X\} + (t^2/2!)\mathbb{E}\{X^2\} + (t^3/3!)\mathbb{E}\{X^3\} + \dots$   $i$  times w.r.t.  $t$  and letting  $t \rightarrow 0$  results in the  $i^{th}$  moment of  $X$  about the origin.

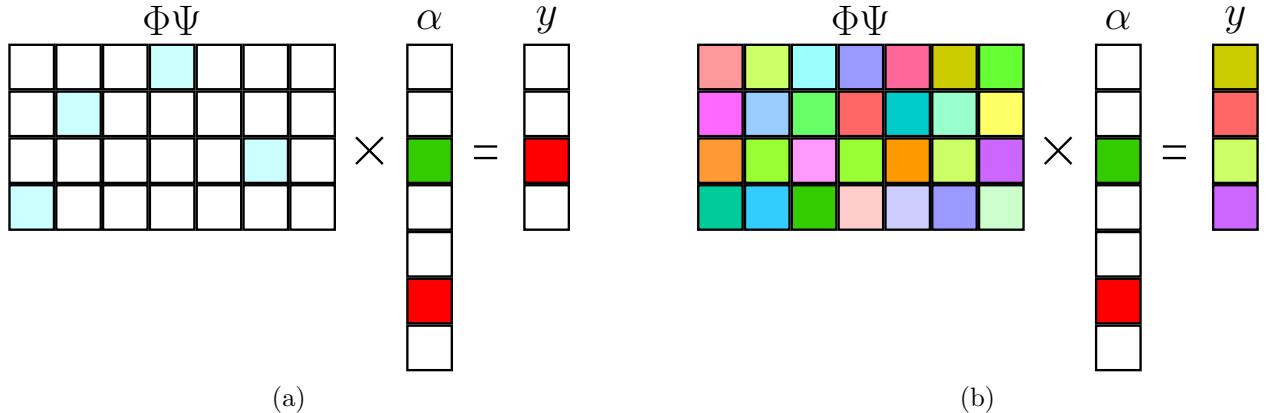


Figure 1.1: Pictorial representation of linear measurements  $\mathbf{y} = \Theta\boldsymbol{\alpha}$ . (a) Undersampling by using measurement basis that is completely coherent with the sparsifying basis means the measurements are likely to contain incomplete information about the sparse signal. (b) Maximal incoherence on the other hand means while no individual measurement convey anything certain about a particular element, they however contain a little bit of information about the entire signal, and so there is no one special measurement without which reconstruction is impossible.

to describe  $\boldsymbol{\alpha}$  completely.

Having obtained the measurements, all that remains is to “decompress”  $\mathbf{y}$  from (1.2) and recover  $\mathbf{x}$ . Since (1.2) is a vastly underdetermined system of linear equations, it admits infinitely many solutions that are contained in the subspace  $S = \boldsymbol{\alpha} + \mathcal{N}(\Phi\Psi) \in \mathbb{R}^{n-m}$  and thus one must impose additional constraints on the solutions for there to be any hope in recovering the desired one. For this precise reason, CS relies on the notion of sparsity in the signals of interest to isolate them out of the subspace  $S$ . To that end, the recovery process consists of a search for the signal  $\mathbf{x}$  with the fewest non-zero coefficients  $\boldsymbol{\alpha}$  when expressed in  $\Psi$  that yields the measurements  $\mathbf{y}$ . An intuitive way of doing so, is to find a solution in  $S$  that has the minimum  $l_0$  norm:

$$\hat{\mathbf{x}} = \Psi\hat{\boldsymbol{\alpha}} : \hat{\boldsymbol{\alpha}} = \arg \min_{\mathbf{z}} \|\mathbf{z}\|_0 \quad \text{subject to} \quad \mathbf{y} = \Theta\mathbf{z}. \quad (1.4)$$

It is easy to see that if  $\boldsymbol{\alpha}$  is  $K$ -sparse in  $\Psi$ , and  $\Theta$  is injective (i.e., one-to-one) on all  $2K$ -sparse vectors, then the minimizer of (1.4) must be  $\boldsymbol{\alpha}$ : Since  $\boldsymbol{\alpha}$  is feasible, the solution  $\mathbf{z}$  to (1.4) must be at most  $K$ -sparse which in turn implies that  $\mathbf{z} - \boldsymbol{\alpha}$  is at most  $2K$ -sparse. Since both  $\boldsymbol{\alpha}$  and  $\mathbf{z}$  are feasible, one has  $\Theta(\mathbf{z} - \boldsymbol{\alpha}) = \mathbf{0}$ , but  $\Theta$  is one-to-one on all  $2K$ -sparse

vectors and so the only conclusion is  $\mathbf{z} = \boldsymbol{\alpha}$ .

The observation outlined above also entails a lower bound on the minimal number of measurement necessary for there to be any hope in the recovery of  $\boldsymbol{\alpha}$ : to guarantee the injectivity of  $\Theta$ , its nullspace  $\mathcal{N}(\Theta)$  must not contain vectors that are at most  $2K$ -sparse, but this is only possible if the smallest number of dependent columns<sup>4</sup> in  $\Theta$  is greater than  $2K$ . Since the smallest number of dependent columns for  $\Theta$  (more generally, for any matrix) is bounded within the interval  $[2, m + 1]$ , it follows that the number of measurements must, necessarily, be  $m \geq 2K$ .

Unfortunately though, solving (1.4) directly is both numerically unstable and NP-hard, requiring an exhaustive search through  $\sum_{i=0}^K \binom{n}{i}$  possible locations of the non-zero entries in  $\boldsymbol{\alpha}$  which is simply not computationally feasible. With such difficulties, it is tempting to, instead, fall back to the classical least mean squares approach, based on minimizing the  $l_2$  norm:

$$\hat{\mathbf{x}} = \Psi\hat{\boldsymbol{\alpha}} : \hat{\boldsymbol{\alpha}} = \arg \min_{\mathbf{z}} \|\mathbf{z}\|_2 \quad \text{subject to} \quad \mathbf{y} = \Theta\mathbf{z}. \quad (1.5)$$

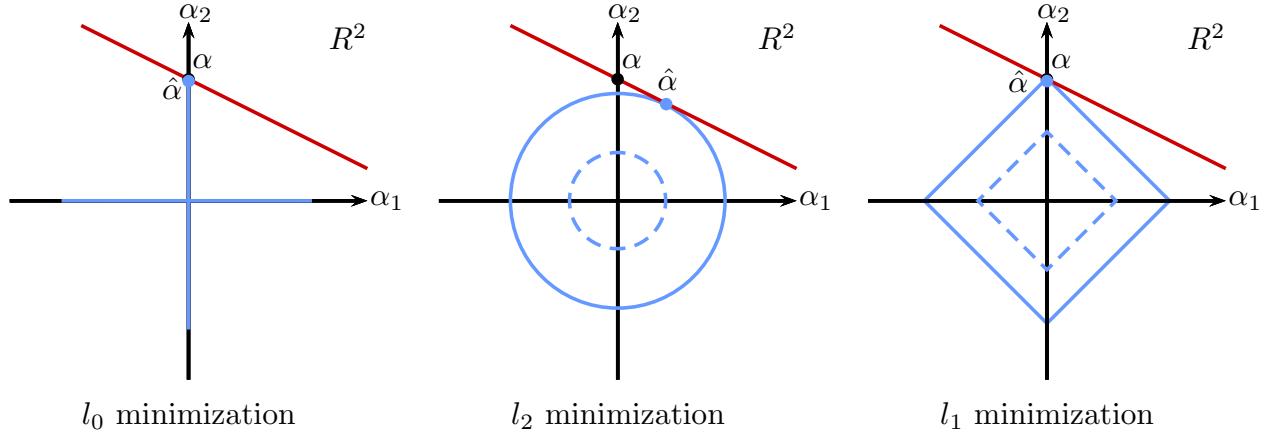
However, the minimization (1.5) will almost never give a solution that is sparse as the geometry of the  $l_2$ -norm ball does not lend itself well to detecting sparsity. To see this, suppose  $\boldsymbol{\alpha} = [\alpha_1 \ \alpha_2]^\top \in \mathbb{R}^2$  has only one non-zero entry  $\alpha_2$ , and only one CS measurement is taken. In this scenario, the solution space  $S \in \mathbb{R}^1$  is simply a straight line passing through  $[0 \ \alpha_2]^\top$ . Minimizing the  $l_p$  norms of the solution is equivalent of inflating the corresponding norm ball from the origin until the closest intersection with a vector in  $S$ , the vector with the minimum corresponding norm. Figure 1.2 compares such geometries for the  $l_0$ ,  $l_1$ , and  $l_2$  norms, and clearly the minimum  $l_2$  norm solution, while has the least energy, is rarely sparse.

One may then wish to consider  $l_1$  minimization problem which seems to work well based on its geometry as depicted in Figure 1.2:

$$\hat{\mathbf{x}} = \Psi\hat{\boldsymbol{\alpha}} : \hat{\boldsymbol{\alpha}} = \arg \min_{\mathbf{z}} \|\mathbf{z}\|_1 \quad \text{subject to} \quad \mathbf{y} = \Theta\mathbf{z}. \quad (1.6)$$

---

<sup>4</sup>Also known as the *spark* of a matrix. Formally,  $\text{spark}(\Phi) = \min_{\mathbf{x} \neq 0} \|\mathbf{x}\|_0$  s.t.  $\Phi\mathbf{x} = \mathbf{0}$

Figure 1.2: Geometric representation of  $l_0$ ,  $l_2$  and  $l_1$  minimizations.

where  $\|\mathbf{x}\|_1 = \sum_{i=1}^n |\mathbf{x}_i|$  is the  $l_1$  norm of a vector. It turns out if  $\Phi$  is sufficiently incoherent with  $\Psi$ , then the solution to (1.6) indeed coincide with that of (1.4) [18–21]. Fortunately, the  $l_1$  optimization (1.6), commonly known as *Basis Pursuit (BP)* [26], is computationally feasible and has the practical significance of being able to be solved with traditional linear programming techniques.

Such practical revelation comes, however, not without its cost: the equivalence is only possible at the expense of requiring more than  $2K$  measurements in order to recover sparse signals from (1.6). The following section summarises the representative results in CS that sum up the general idea.

### 1.1.1 Main Results

An early result [27] showed that if  $m$  rows are selected uniformly at random from an orthonormal matrix  $\Phi \in \mathbb{R}^{n \times n}$  with

$$m \geq C\mu^2(\Phi\Psi)K \log(n)$$

for some positive constant  $C$  independent of  $K$  and  $n$ , then all  $K$ -sparse  $\boldsymbol{\alpha}$  can be recovered exactly from (1.6) with overwhelming probability<sup>5</sup>. This early result captures the two main on-going themes in CS: (1) the smaller the coherence between the sensing and representation basis, the fewer measurements are needed for recovery, and (2) the number of measurements

---

<sup>5</sup>It is shown that the probability of success exceeds  $1 - \alpha$  if  $m \geq C\mu^2(\Phi\Psi) \log(n/\alpha)$

needed is directly proportional to the sparsity of the signal, i.e., in the order of  $K \log(n)$ , which may be far less than the signal size apparently demands.

While remarkable, the above result however, is not practically robust for two reasons. First, most signals of interest are not exactly sparse, for example, natural images are never exactly sparse in the Wavelet basis, they are, however, approximately sparse (or *compressible*) in the sense that the sorted magnitudes of the Wavelet coefficients  $\alpha_i$  decay quickly and, therefore, the Euclidean distance between  $\boldsymbol{\alpha}$  and its  $K$ -sparse approximation  $\boldsymbol{\alpha}_K$  (i.e., keeping only the  $K$  largest values of  $\boldsymbol{\alpha}$ ),  $\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_K\|_2$ , is small. Second, any practically measured data will inevitably be corrupted by noise due to environmental factors or as a result of finite precision limit in sensing devices. This means in practice, the measurements in (1.2) are usually contaminated with a noise vector  $\boldsymbol{\eta} \in \mathbb{R}^m$ :

$$\mathbf{y} = \Theta \boldsymbol{\alpha} + \boldsymbol{\eta} \quad (1.7)$$

and so the optimization (1.6) should correspondingly be relaxed to allow for small perturbations:

$$\hat{\mathbf{x}} = \Psi \hat{\boldsymbol{\alpha}} : \hat{\boldsymbol{\alpha}} = \arg \min_{\mathbf{z}} \|\mathbf{z}\|_1 \quad \text{subject to} \quad \|\mathbf{y} - \Theta \mathbf{z}\|_2 \leq \epsilon. \quad (1.8)$$

where  $\epsilon$  is an upper bound on the size of the noise. (1.8) is also known as *Basis Pursuit De-Noising (BPDN)*.

As discussed previously, to guarantee exact recovery of every  $K$ -sparse signal, the matrix  $\Theta$  needs to be injective on all  $2K$ -sparse vectors, however, in order to deal with these non-idealities, a stronger condition must be enforced. The standard tool in CS proven to be sufficient (but not necessary) is the so-called *Restricted Isometry Property (RIP)* on  $\Theta$  introduced in [28]:

**Definition 1.1.** A matrix  $\Theta$  satisfies the RIP of order  $K$  if there exists a constant  $\delta_K \in [0, 1)$  such that

$$(1 - \delta_K) \|\mathbf{x}\|_2^2 \leq \|\Theta \mathbf{x}\|_2^2 \leq (1 + \delta_K) \|\mathbf{x}\|_2^2$$

holds for all  $\mathbf{x}$  that are at most  $K$ -sparse.

The RIP essentially requires every subset of  $K$  or fewer columns of  $\Theta$  be approximately an

orthonormal system (exactly, when  $\delta_K = 0$ ). When  $\Theta$  satisfies the RIP of order  $2K$ , this property has a nice geometric description that  $\Theta$  is injective on all  $K$ -sparse signals and preserves the Euclidean distance between any pair of  $K$ -sparse signals. Indeed, any arbitrary  $2K$ -sparse vector  $\mathbf{x}$  can be decomposed into two unique  $K$ -sparse vectors,  $\mathbf{x} = \mathbf{x}_K - \mathbf{x}'_K$ , and substituting such decomposition into the definition of RIP reveals that the distance between  $\Theta\mathbf{x}_K$  and  $\Theta\mathbf{x}'_K$  is approximately that of  $\mathbf{x}_K$  and  $\mathbf{x}'_K$  and in particular, the lower bound entails that  $\Theta\mathbf{x}_K$  and  $\Theta\mathbf{x}'_K$  are unique (since  $\delta_{2K} \neq 1$ ). The meat of CS is that provided  $\delta_{2K}$  is small enough, the solution to (1.6) coincide with that of (1.4). While there are many variations of such result using the RIP (see e.g., [18, 28–30]), the following is a popular and elegant one:

**Theorem 1.1.** (due to [31]) Suppose that  $\Theta$  satisfies the RIP of order  $2K$  with  $\delta_{2K} < \sqrt{2} - 1$ , then for any arbitrary signal  $\boldsymbol{\alpha}$  and the corrupted measurements (1.7) with  $\|\eta\|_2 \leq \epsilon$ , the solution  $\hat{\boldsymbol{\alpha}}$  to (1.8) satisfies

$$\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\|_2 \leq C_1 \frac{\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_K\|_1}{\sqrt{K}} + C_2 \epsilon$$

for some small constants  $C_1, C_2$  which depend only on  $\delta_{2K}$ .

In words, Theorem 1.1 says that when no noise is present ( $\epsilon = 0$ ), then for an arbitrary signal  $\boldsymbol{\alpha}$ , not necessarily sparse, Basis Pursuit will approximately recover its largest  $K$  coefficients, and if  $\boldsymbol{\alpha}$  is  $K$ -sparse, then exact reconstruction occurs (since  $\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_K\|_1 = 0$ ). While under the presence of noise, Basis Pursuit is stable to perturbations in the measurements giving a perturbation in reconstruction error no worse than the original error,  $\epsilon$ , introduced to the measurements. The requirement  $\delta_{2K} < \sqrt{2} - 1$  can be slightly relaxed [32–34] but has been shown that (1.8) can fail in recovery of certain sparse signals once  $\delta_{2K} \geq 1/\sqrt{2}$  [35].

Note that there are no probabilities involved in the theorem: it is *uniform* in the sense that if one can find a matrix satisfying the RIP then Basis Pursuit is guaranteed to recover the  $K$  largest coefficients of any signal with no probability of failure. However, in order for the theorem to be useful in practice, one needs to find matrices having a small RIP constant with small number of measurements. Unfortunately, there are very few deterministic matrices

that meet such requirement, but a variety of random matrices have been shown to exhibit such property with exponentially high probability. Some of the most frequently used are detailed in the following section.

### 1.1.2 Matrices Satisfying the RIP

#### 1.1.2.1 Sub-Gaussian Matrices

A random variable  $X$  is said to be sub-Gaussian if its moment-generating function<sup>6</sup> is bounded by that of a zero-mean Gaussian, i.e.,  $\mathbb{E}\{\exp(Xt)\} \leq \exp(c^2 t^2/2)$  for all  $t \in \mathbb{R}$  and some positive constant  $c$ . Well-known examples include Gaussian matrices (whose entries are zero-mean Gaussian), Rademacher matrices (whose entries are uniform  $\pm 1$ ), and more generally, any bounded random variable is sub-Gaussian. The following theorem shows that a sub-Gaussian matrix satisfies the RIP with number of measurements nearly linear in sparsity  $K$  (see also, [36, 37]):

**Theorem 1.2.** *(due to [23]) Let  $\Phi$  be an  $m \times n$  sub-Gaussian random matrix, then the normalized matrix  $\frac{1}{\sqrt{m}}\Phi$  satisfies the RIP with probability at least  $1 - 2\exp(-c\delta^2 m)$  for every sparsity level  $1 \leq K \leq n$  and  $\delta_K \in (0, 1)$  provided that the number of measurements  $m$  satisfies*

$$m \geq \frac{C}{\delta^2} K \log\left(\frac{en}{K}\right)$$

where  $c, C$  depends only on constants in the definition of the sub-Gaussian.

Theorem 1.2 says the number of measurements required for partial sub-Gaussian matrices to satisfy the RIP is in the order of  $K \log(n/K)$ , which has been shown to be optimal in the sense that the log-factor must be present and one cannot hope to recover all  $K$ -sparse signals with fewer measurements.

Since signals are usually sparse or compressible in some orthonormal basis  $\Psi \neq \mathbf{I}$ , in which case, one needs the RIP to hold for the matrix  $\Phi\Psi$  instead. It is shown in [37] that provided

---

<sup>6</sup>The moment generating function of the random variable  $X$  is given by  $\mathbb{E}\{\exp(Xt)\} = 1 + t\mathbb{E}\{X\} + (t^2/2!)\mathbb{E}\{X^2\} + (t^3/3!)\mathbb{E}\{X^3\} + \dots$  since differentiating  $\mathbb{E}\{\exp(Xt)\}$   $i$  times w.r.t.  $t$  and letting  $t \rightarrow 0$  results in the  $i^{th}$  moment of  $X$  about the origin.

$\Phi$  is drawn according to Theorem 1.2, this is indeed still the case regardless of the choice of  $\Psi$ .

### 1.1.2.2 Partial Bounded Orthonormal Matrices

A  $m \times n$  partial bounded orthonormal matrix  $\Phi$  is formed by selecting  $m$  rows uniformly at random from an  $n \times n$  orthonormal matrix whose entries are bounded by  $C/\sqrt{n}$  for some constant  $C$  [38, 39]. The discrete Fourier matrix is the well-known example whose entries are bounded by  $1/\sqrt{n}$ . The following theorem shows that matrices coming from such construction satisfies the RIP with number of measurements approximately linear in sparsity  $K$  (see also, [18, 38]):

**Theorem 1.3.** *(due to [39]) Let  $\Phi$  be a  $m \times n$  partial bounded orthonormal matrix, then the matrix  $\frac{1}{\sqrt{m}}\Phi$  satisfies the RIP with probability at least  $1 - n^{-c\log^3(K)}$  for every sparsity level  $1 \leq K \leq n$  and  $\delta_K \in (0, 1)$  provided the number of measurements  $m$  satisfies*

$$m \geq CK \log^3(K) \log(n)$$

where  $c, C$  are some positive constants depending only on constants in the definition of partial bounded orthonormal matrix.

In essence, Theorem 1.3 says one requires the number of measurements in the order of  $K \log^4(n)$  for partial bounded orthonormal matrices to satisfy the RIP. While in comparison with Theorem 1.2, this is much higher, it has been conjectured that this bound can be lowered to the order of  $K \log(n)$  [19, 38].

As with before, when a signal is sparse in some orthonormal basis  $\Psi \neq \mathbf{I}$ , one would require the entries of the matrix  $\Phi\Psi$  to be bounded instead. But this is exactly captured by the definition of mutual coherence (1.3), thus it follows, from Theorem 1.3, that if two orthonormal bases  $\Phi, \Psi \in \mathbb{R}^{n \times n}$  exhibit a low mutual coherence,  $\mu(\Phi\Psi) \leq C$  for some  $C \ll \sqrt{n}$ , then the matrix formed by uniformly selecting  $m$  rows i.i.d. from the matrix  $\frac{1}{\sqrt{m}}\Phi\Psi$  will also satisfy the RIP with high probability.

### 1.1.2.3 Partial Random Toeplitz Matrices

Let  $\{\phi_i\}_{i=1}^{n+m-1}$  be a sequence whose entries are i.i.d. realisations of sub-Gaussian random variable. Then a  $m \times n$  partial random Toeplitz matrix has the form:

$$\Phi = \begin{bmatrix} \phi_n & \phi_{n-1} & \dots & \phi_2 & \phi_1 \\ \phi_{n+1} & \phi_n & \dots & \phi_3 & \phi_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \phi_{n+m-1} & \phi_{n+m-2} & \dots & \phi_{m+1} & \phi_m \end{bmatrix}$$

The following theorem shows that matrices from such construction satisfies the RIP with number of measurements nearly quadratic in sparsity  $K$  (see also, [39, 40]):

**Theorem 1.4.** *(Due to [41]) Let  $\Phi$  be a partial random Toeplitz matrix, then the matrix  $\frac{1}{\sqrt{m}}\Phi$  satisfies the RIP with probability at least  $1 - \exp(-cm/K^2)$  for every sparsity level  $1 \leq K \leq n$  and  $\delta_K \in (0, 1)$  provided the number of measurements  $m$  satisfies*

$$m \geq CK^2 \log(n)$$

where  $c, C$  are some constants depending only on  $\delta_K$  and the definition of the sub-Gaussian.

While Theorem 1.4 might seem somewhat discouraging in terms of the number of measurements needed compared to Theorem 1.2 and Theorem 1.3, the structure of  $\Phi$  arises naturally in many practical applications, such as radar imaging [42], Fourier optics [43, 44], wireless channel estimation [41], etc. and so such matrices are popular CS measurement matrices despite the additional number of measurements.

## 1.2 Hyperspectral Imaging

When an Electromagnetic (EM) wave and matter interact, photons are generally absorbed, reflected, or transmitted by the material. Depending on the wavelength,  $\lambda$ , of the incident wave, the material will absorb or re-emit the photons in differing proportions, determined completely by the physical properties of the material (such as the chemical compo-

sition, structure, density, etc.). This selective interaction of light by the material gives the reflected photon stream a wavelength-dependent profile described by its *reflectance* function: a dimensionless ratio of the reflected radiant flux  $\Phi_r(\lambda)$  to the incident flux  $\Phi_i(\lambda)$ ,  $R(\lambda) = \Phi_r(\lambda)/\Phi_i(\lambda)$ . By measuring this response over a range of known wavelengths, it is possible to infer information and differentiate between materials to within submicroscopic scale [45–48]. The ability to garner information to such a degree of intricacy through non-invasive and non-destructive measurements ultimately formed the basis for *spectroscopy*, the science of characterizing material properties by its interactions with EM radiations.

While colour (trichromatic) images are inherently a form of spectroscopy, their primary purpose, however, is to portray the hues apparent in a scene as perceived by a human observer. The emphasis on these systems is therefore the qualitative properties they produce, i.e., their capability to match the response of the human visual system. The wealth of spectral information available in the scene is thus traded for aesthetic appeal to the human eye as additional spectral information indiscernible to the human eye cannot be perceived and is therefore unnecessary. While such a decision is rational, restricting the system to create images solely for human comprehension significantly undermines the power of imaging technology in quantitative fields. For example, distinguishing materials based on colour alone is difficult, given the way in which the human visual system integrates together the responses from a broad region of the visible spectrum for establishing colour vision. Many completely different materials may appear to have similar colouring to a human observer (see Figure 1.3(a)). Additionally, the differences between these materials may only be revealed outside of the observable spectrum. Therefore by obtaining additional measurements over a different portion of the EM spectrum, each would provide another possible point-of-difference, and so as the number of measured spectral components increases, the confidence with which materials can be discerned is necessarily improved [49] (see Figure 1.3(b)).

Hyperspectral imaging is a spectroscopic technique that does not limit the acquisition of spectral content in the way that traditional imaging systems do. As opposed to trichromatic systems where the captured image contains only three spectral components (usually corresponding broadly to the red, green, and blue portions of the visible spectrum), each pixel

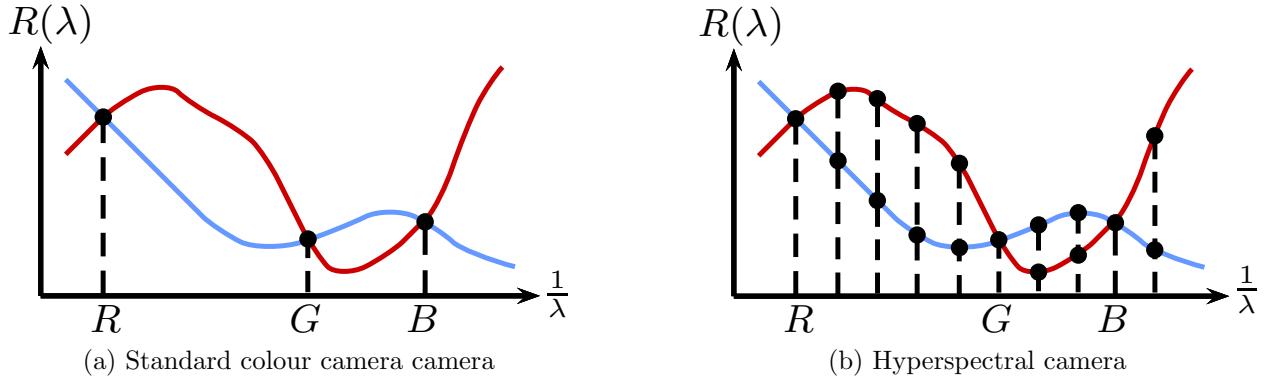


Figure 1.3: Comparison of the measurements obtained from a standard colour camera to that of a hyperspectral camera. Note that the filter responses for acquiring the indicated wavelengths are assumed ideal and are represented as dotted lines. Two material spectra may appear identical from the measurements of a trichromatic camera as in (a) but is completely separable via a hyperspectral camera as in (b).

of the collected hyperspectral image contains a densely sampled portion, typically spanning the visible (VIS,  $400\text{nm} \leq \lambda \leq 750\text{nm}$ ) and near-infrared (NIR,  $750\text{nm} \leq \lambda \leq 1100\text{nm}$ ) section, of the EM spectrum. It therefore combines spatial information available in regular images with detailed spectral information for each pixel. The richness of information entailed within a hyperspectral image therefore enables one to relate the spatial locations with material presence: each pixel in the image contains a complete set of spectral measurements that uniquely attributes to the underlying material<sup>7</sup>. The ability to classify and characterize large areas of interest effectively and reliably has thus motivated the initial developments in this kind of technology (and related analysis methods) [6].

While the development of hyperspectral imaging originated in the field of remote sensing where large areas of planetary resources (e.g., agriculture and minerals) needs to be monitored quickly and efficiently [6], from the discussion above, one can see the capability of such technology is clearly applicable to a vast range of scientific fields [50]. As a result, attention has increasingly turned to the development of terrestrial-based applications, such as quality control, medical diagnosis, and forensic analysis [12–17].

---

<sup>7</sup>The explanation here assumed that, for simplicity, only a single material is associated with each pixel of the image. In practice the pixels usually contain a mixture of disparate materials, see Section 1.2.2.

### 1.2.1 Spectral Image Capture

In order to appreciate how hyperspectral images are captured, it is appropriate to first review more traditional panchromatic imaging systems. When a scene is illuminated by a light source containing a wide range of energies (i.e., wavelengths) with a known spectral power distribution  $S(\lambda)$ , a certain fraction of energy (i.e., photons) from the source is absorbed by the materials within the scene, while the rest is re-emitted (transmitted or reflected). As described earlier, the amount of energy reflected is characterized by the reflectance  $R(\lambda)$  and constitutes a unique spectral signature for each material. The amount of light received by an imaging sensor is therefore, mathematically, the product  $S(\lambda)R(\lambda)$ . Each sensor element then integrates the received photons across all wavelengths, weighted by the relative sensitivity of the sensor to the different wavelengths of light, converting it to a digital value  $I$  representing the intensity value of a pixel within the image:

$$I = \int_{\lambda=0}^{\infty} S(\lambda)R(\lambda)C(\lambda)d\lambda$$

where  $C(\lambda)$  is the relative sensitivity of the sensor and is dependent upon its type (e.g., Charge-coupled Device (CCD), Complementary Metal-oxide-semiconductor (CMOS), etc.). By spatially arranging multiple sensor elements in a two-dimensional grid, the resulting measurements form a gray-scale intensity map which mimics the intensity perceived by a human observer of the very same scene: a greyscale or panchromatic image.

#### 1.2.1.1 Trichromatic Image Capture

Extending this to a trichromatic image (such as those produced by modern commercial colour cameras), a trichromatic imaging system produces three (instead of one) intensity maps, each corresponds to the intensity of a particular region of the EM spectrum. Since the intention of such images is, usually, to portray the hues apparent in the scene as perceived by that of the human visual system, the three maps thus usually correspond to the red, green, and blue (RGB) portion of the visible spectrum that the three types of photoreceptor cells (i.e., cones) in the human eye are sensitive to for establishing colour vision. Each pixel within the

image now consists of three measurements  $I_r$ ,  $I_g$ ,  $I_b$ , each can be obtained by fitting colour filters  $H_r(\lambda)$ ,  $H_g(\lambda)$ ,  $H_b(\lambda)$  that match the spectral sensitivity of the cones over the sensor elements to transmit the corresponding portions of the visible spectrum:

$$I_{\{r,g,b\}} = \int_{\lambda=0}^{\infty} S(\lambda)R(\lambda)H_{\{r,g,b\}}(\lambda)C(\lambda)d\lambda.$$

In practice however, since only one filter type can be fitted on top of a sensor element at a time, each element can only record a single colour value. In order to obtain all three colour values simultaneously, a Colour Filter Array (CFA) consisting of the three filter types is typically used. The filters of the CFA are spatially arranged strategically (typically in a Bayer mosaic [51]), so that while each sensor element can only measure a particular colour, the values for the other two can be interpolated *a posteriori* from neighbouring pixels. The compromise for such convenience (i.e., speed), is therefore the loss of spatial resolution (i.e., quality). However, as the primary goal of trichromatic images is to merely recreate the qualitative details of the imaged scene with sufficient accuracy, such trade-off is clearly favourable over the quantitative details that could otherwise be obtained from a more elaborate and prolonged procedure.

### 1.2.1.2 Hyperspectral Image Capture

More generally, provided the imaging sensors and filters with appropriate sensitivities are available, it is clear that one can image any region of light, over as many wavelengths as one desires, and not necessarily constrain the acquisition by our ability, or lack thereof, to perceive them qualitatively. In the spirit of this, hyperspectral imaging systems further the spectral contents in that of a trichromatic image by acquiring bands of intensity values across the VIS and NIR spectrum, so that each pixel  $I$  contains a set of contiguous spectral measurements, each centred at various wavelengths  $\lambda_0 \in [400\text{nm}, 1100\text{nm}]$ :

$$I[\lambda_0] = \int_{\lambda=0}^{\infty} S(\lambda)R(\lambda)H(\lambda, \lambda_0)C(\lambda)d\lambda \quad (1.9)$$

where  $H(\lambda, \lambda_0)$  is the filter response for the particular portion of the EM spectrum to be imaged, centred at wavelength  $\lambda_0$ . The exact number of discrete bands depends on the system implementation which is usually application driven, and can vary from tens [52, 53] to hundreds [54–56] of bands.

While a trichromatic image can be captured at once using the CFA, the number of spectral bands dictated by a hyperspectral image makes it impractical to capture the entire data simultaneously. Such inconvenience is an unfortunate artefact of sampling a three-dimensional signal using a two-dimensional imaging sensor. Therefore, regardless of the acquisition methods, a hyperspectral image must be acquired over time or captured at the expense of other dimensions. As such, most hyperspectral systems generally sequentially scan the third dimension, trading temporal resolution for either spatial or spectral resolution.

In spectral scanning (i.e., focal plane) systems, successive 2D spatial images of the same scene are captured at different wavelengths over time. A specific wavelength is isolated by placing an optical bandpass filter with the desired response  $H(\lambda, \lambda_0)$  in front of the imaging sensor. By sequentially cycling through the selectable wavelengths  $\lambda_0$  either mechanically by using a rotating filter wheel, or electronically via an electronically tuneable filter [57] (e.g., a Liquid Crystal Tuneable Filter), the hyperspectral image is formed by concatenating the obtained images together ordered by  $\lambda_0$ .

Alternatively, instead of recording both spatial dimensions at once and sweeping through the wavelengths, spatial scanning systems record a spatial and spectral dimension simultaneously whilst sweeping through the remaining spatial dimension temporally. This is typically achieved by utilizing a dispersive optic (e.g., a diffraction grating or prism) to separate the received light into contiguous spectral components [49]. A typical device of this kind is the push-broom imager [58] that scans, spatially, one line of the scene at a time while the dispersive element separates the single line into discrete wavelength bands. The captured image is therefore spatial in one dimension and spectral in the other and the entire hyperspectral image is formed by sweeping the imager across the second spatial dimension over time.

### 1.2.2 Spectral Mixing

Our descriptions thus far have assumed that (1) spatially, only a single material is associated with each pixel in the hyperspectral image, and (2) spectrally, the measured (or inferred) reflectance function uniquely identifies a single material. In practice, however, these assumptions rarely hold due to the consequence of discretizing an otherwise continuous real-world phenomenon. Imaging sensors inherently collect discrete samples as modelled by (1.9) where each spectral measurement is an integration over a range of wavelengths for which the response is non-zero. Such a discretization process means that it is entirely possible for materials of different underlying reflectances to result in the same measurement output, and the materials are said to be *metamers*. Accordingly, metamers cannot be distinguishable based purely on their spectral measurements, and the inability to do so is known as *metameric failure*.

To the contrary, since the observable spatial extent of a sensor element covers a finite area in the scene, determined by its Field of View (FOV), more than one distinct material may be directly visible to the sensor. The resulting spectral measurements for the corresponding pixel is then, necessarily, some composite of the individual reflectance spectra associated with the material constituents within the area imaged. When this occurs, the pixel is said to be *mixed* and this phenomenon is termed *spectral mixing*. Because the information contained within a mixed pixel is no longer directly attributable to any one material, this necessitates additional non-trivial processing and analysis techniques for establishing the composition of the scene from a hyperspectral image. This process, of determining which materials are present together with their relative contribution to each and every pixel of the hyperspectral image, is known as *spectral unmixing*. Depending on whether the interaction between the materials and the incident light is considered at a macroscopic or microscopic scale, the type of mixing can be classified, respectively, as *linear* or *non-linear* and the related unmixing techniques also differ. Figure 1.4 illustrates the two spectral mixing categories.

On a macroscopic level, it is reasonable to assume that the area observable within the FOV of a sensor element is segregated into a few distinct regions of  $n_e$  different “pure” mate-

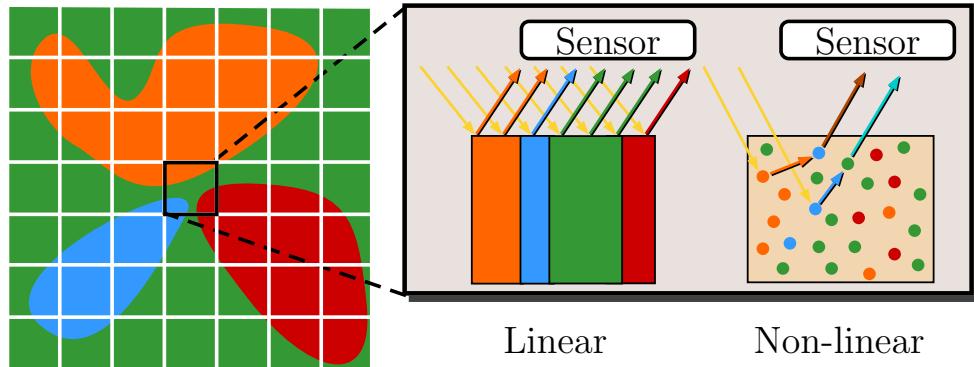


Figure 1.4: Multiple materials often contribute to the signal at a single pixel due to the finite observable extent of an imaging sensor, resulting in a mixed pixel as highlighted. The mixing behaviour can be modelled as either linear or non-linear. In linear mixing the incident illumination is scattered by a single material before reaching the sensor. In non-linear mixing, light is scattered by multiple materials before reaching the sensor.

rial constituents, each occupying a non-overlapping portion of the surface. The reflectance spectra of these  $n_e$  material constituents are referred to as *endmembers*, and their relative fractional contributions (corresponding to the occupied proportion of the area) to the given pixel are called *abundances*. This assumption implies that each ray of incident radiation has only interacted with, and therefore has been altered by, a single material before arriving at the detector [7, 59] (illustrated in Figure 1.4). If the spectral properties within each  $n_e$  distinct regions are relatively constant, it logically follows that the measured EM spectrum of a given pixel  $\mathbf{x} \in \mathbb{R}^{n_\lambda}$  (measured over  $n_\lambda$  discrete wavelengths) is a linear combination of the endmembers  $\mathbf{e}_i \in \mathbb{R}^{n_\lambda}$  within the observable area, weighted by the fractional abundances  $a_i \in \mathbb{R}$  of the materials comprising the area. The mixing is therefore said to be *linear*, and  $\mathbf{x}$  can be modelled according to the so-called Linear Mixing Model (LMM):

$$\mathbf{x} = \sum_{i=1}^{n_e} a_i \mathbf{e}_i + \boldsymbol{\eta} = \mathbf{E}\mathbf{a} + \boldsymbol{\eta} \quad \text{subject to} \quad \begin{aligned} \sum_{i=1}^{n_e} a_i &= 1 \\ 0 \leq a_i, e_i &(\leq 1) \quad \forall i \in \{1, \dots, n_e\}. \end{aligned} \quad (1.10)$$

where  $\mathbf{E} = [\mathbf{e}_1 \dots \mathbf{e}_{n_e}] \in \mathbb{R}^{n_\lambda \times n_e}$  is a matrix containing a column-wise concatenation of the endmembers,  $\mathbf{a} = [a_1 \dots a_{n_e}]^\top \in \mathbb{R}^{n_e}$  is the abundance vector for the pixel, and  $\boldsymbol{\eta} \in \mathbb{R}^{n_\lambda}$  models the measurement noise. The two constraints associated with the LMM are commonly referred to as, respectively, the Sum-to-Unity (STU) (or additivity) constraint, and

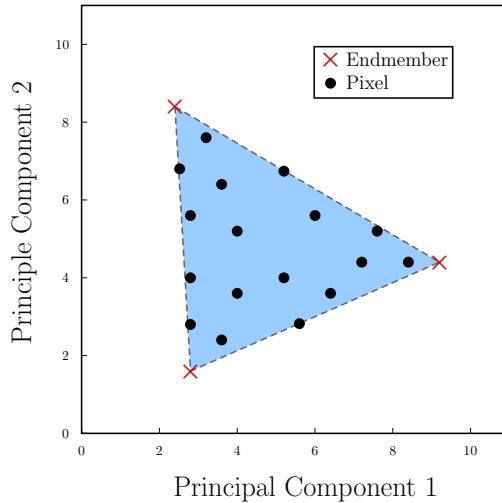


Figure 1.5: Geometric representation of the linear mixing model. Each pixel (black dots) resides within the simplex (blue area) formed by the endmembers (red crosses). The relative position of the pixel in the simplex corresponds to its abundance vector.

the non-negativity constraint. The STU constraint ensures that all endmember contributions are accounted for, while the non-negativity restricts (1) the abundances to depict physical proportions and (2) the endmembers to correspond to real-world reflectance spectra that do not absorb (i.e.,  $R(\lambda) \geq 0$ ), or reflect (i.e.,  $R(\lambda) \leq 1$ ), more energy than was incident. Geometrically, the combined effect of the STU and non-negativity constraint means that each pixel will reside within an  $(n_e - 1)$ -dimensional simplex as shown in Figure 1.5 whose vertices corresponds to the endmembers, reflecting the fact that each pixel is a convex combination of the endmembers. The LMM is suitable in situations where the occurrence of spectral mixing is, predominantly, due to the lack of sufficient resolution resulting in the presence of multiple materials within the FOV of the sensor elements. If given sufficient resolution, each sensor element would only receive energy scattered by a single material, all pixels would be “pure” and consequently no mixing would occur.

On the other hand, if the received photons have been scattered multiple times between multiple material constituents in the scene before reaching a sensor element, the mixing is termed non-linear (illustrated in Figure 1.4). These circumstances are usually apparent when looking from a molecular level, where the material in question is further decomposed to be an intimate mixture of different components. Measurements from secondary (or higher) reflections where light is reflected off more than one object before reaching the sensor is

also an example of non-linear mixing [7]. In these cases, the photons received by the sensor are viewed as having interacted with more than one material and the measured spectrum for the given pixel is no longer a linear mixture of the endmembers. Methods aiming at dealing with non-linear mixing are comparatively more complex and difficult than their linear counterparts [59], and require domain-specific models. Therefore, while all hyperspectral imagery can be considered non-linearly mixed to a certain extent, the LMM (1.10) is, for most practical scenarios, widely accepted in the literature as the most tractable alternative. Apart from its mathematical simplicity, leading to a plethora of unmixing techniques (see e.g., [7, 59] and the references therein) the use of the LMM is generally justified by the intentionally low spatial resolution employed in most hyperspectral applications, confining the mixing to a macroscopic level.

### 1.3 Direction and Original Contributions

Despite the considerable redundancies in digital data leading to efficient post-acquisition compression schemes, from the preceding overview, the philosophy behind all of our current technologies for digital data acquisition remains unchanged: in lieu of *a priori* requirements, existing systems fall victim to the psychological expectation of requiring very large number of samples in situations that, in actuality, may only need very few. Hyperspectral imaging, with its large inflow of data, is all but immune to this psychological downfall. The problem is made more prevalent by the underlying principle behind spectroscopy, that, an increase in spectral sampling improves the likelihood of collecting suitable features for differentiating arbitrary materials. While not all hyperspectral applications require a large number of spectral bands for effective operation, in the absence of prior constraints, the small number of bands needed to sufficiently characterize a particular scenario may be substantially inadequate for another. As such, hyperspectral systems typically oversample the spectral details in a scene, so that not only can the additional information accommodate any acquisition variabilities, but more importantly, provides flexibility for the increasing interests in extending the use of hyperspectral data to the various fields of science. The advantages, however, comes at the

cost of a data deluge, leading to a substantial increase in acquisition, storage, communication, and processing overheads. While these difficulties hinder the accessibility of spectral imaging systems, there is nevertheless great promise that hyperspectral data will provide the richness of information for solving a wide range of problems.

This thesis represents a contribution to the effort to bring hyperspectral techniques into mainstream application by examining the various aspects needed for incorporating the theory of compressive sensing into hyperspectral imaging. In particular, each of the remaining chapters focuses on, respectively, the acquisition, recovery, and unmixing aspect of a compressive hyperspectral imaging system. The structure and original contributions within the following chapters are as follow:

A detailed survey of existing compressive hyperspectral acquisition concepts is provided in Chapter 2. Following this survey, a discussion into the limitations and difficulties of existing systems culminates in the exploration of an alternative compressive hyperspectral imaging strategy. This new strategy exploits existing measurement techniques to produce a hybrid reconstruction method that allows hyperspectral compressive sensing to be achieved more efficiently using a combination of 2D measurements obtained from different acquisition modalities. A reconstruction algorithm tailored specifically to this new method is also presented.

Chapter 3 details an investigation into CS recovery models aimed at modelling data structures inherent to hyperspectral imagery. A review of prominent models used in literature leads to an investigation of two convex optimization recovery schemes that accounts for both spatial and spectral sparsity exhibited in typical hyperspectral data. Both methods are shown to successfully reduce the number of CS measurements necessary for accurate recovery of the underlying data and yield promising results over existing joint spatio-spectral models.

Chapter 4 investigates means of performing spectral unmixing of hyperspectral data directly from their CS measurements. A blind compressive unmixing technique based on the LMM is presented which forgoes the assumption of dominant pure pixel presence that previous techniques relied on. The developed technique incorporates multiple priors pertinent to the LMM for successful joint recovery of both the endmembers present and their associated

abundances.

Possible areas and ideas for future work directly resulting from this thesis is discussed in Chapter 5 and finally, Chapter 6 concludes the thesis.

### 1.3.1 Summary of Contributions

A list of the original contributions made within this thesis is as follows:

Chapter 2

- A novel HSI acquisition method using multimodal CS measurements (Section 2.4)
- An accompanying recovery algorithm, dubbed MuRecAL, based on the alternating direction method of multipliers (Section 2.5)
- Detailed performance analyses of the proposed multimodal framework (Section 2.6)

Chapter 3

- Two recovery models for improved reconstruction of HSI from CS measurements using compounded spatio-spectral sparsity priors (Section 3.2)
- Detailed performance analyses of the proposed decoding models (Section 3.3)

Chapter 4

- A novel blind compressive HSI linear unmixing technique that directly identifies the endmembers and abundances from CS measurements of HSI with highly mixed pixels (Section 4.2)
- Detailed performance analyses of the proposed CS-HSI unmixing technique (Section 4.2)

### 1.3.2 Publications

Existing publications based on the work presented in this thesis are:

- Lee, W. Y.-L., Eason, D., and Andrews, M., “Compressive Hyperspectral Image Sensing Restoration via Joint TV-L1 Regularisation”, Electronics New Zealand Conference (ENZCon), 2012.

- Eason, D., **Lee, W. Y.-L.**, and Andrews, M., “Compressively Sensed Hyperspectral Image Recovery using Total Variation Minimisation by Approximation”, Electronics New Zealand Conference (ENZCon), 2012.
- Lee, W. Y.-L., and Andrews, M., “Blind spectral unmixing for compressive hyperspectral imaging of highly mixed data”, IEEE International Conference on Image Processing (ICIP), 2014.
- Lee, W. Y.-L., and Andrews, M., “Multimodal Compressive Hyperspectral Imaging”, Submitted for publication.



# Chapter 2

## Compressive Hyperspectral Image Acquisition

The previous chapter discussed the need for a shift in the data acquisition paradigm to accommodate the ever-growing demand in quality of digital representations, especially for high-dimensional datasets such as hyperspectral imagery. The reasons being that hyperspectral images are very costly to acquire due to the expensive optical filters and photodiodes needed to capture light outside the visible spectrum, and the tremendous amount of data to be captured poses a serious burden on system implementation especially in embedded systems where memory storage is very limited. With compressive sensing introduced as a promising alternative, hyperspectral imaging can therefore benefit greatly from such a paradigm since the acquisition burden is significantly reduced, overcoming the aforementioned limitations, while the recovery process, although highly complex, can be delayed to the back-end where resources are more accessible.

While optimal sensing matrices should be random in theory, unfortunately it is rarely feasible to have imaging systems that implement such measurements due to the lack of either available hardware that can emulate random matrices efficiently or memory-efficient ways of simulating or computing these matrices during the recovery for the given size of the underlying data. Hyperspectral imaging systems that utilises compressive sensing can therefore only emulate matrices such as random sub-selection of the rows of the Walsh, Hadamard or

Fourier orthogonal basis in favour of their low storage and computational costs despite being less than ideal. Furthermore, as with the conventional hyperspectral imaging techniques, the three-dimensional nature of the hyperspectral data poses restrictions on the ways compressive measurements can be made since existing imaging sensors can only acquire at most two of the three dimensions simultaneously. The remaining data must therefore be acquired over time or captured at the expense of the other dimensions. This means that, in practice, it is extremely difficult to obtain compressive measurements that are truly compressive across the entire hyperspectral volume.

This chapter explores a different approach to the problem of hyperspectral compressive sensing. Because hyperspectral image acquisition is possible in a variety of modalities (e.g., focal plane vs. push-broom sensor, see Section 1.2.1.2), instead of simply extending CS methods to higher dimensions, where the sensing methods are currently impractical or unknown, we exploit these measurement techniques to produce a hybrid reconstruction method that allows high dimensional compressive sensing (in this case, 3D hyperspectral images) to be achieved much more efficiently using multiple 2D measurements. A survey of existing compressive hyperspectral acquisition techniques is firstly presented. A discussion of the problems and difficulties in these methods then culminates in the proposal and analysis of a novel compressive hyperspectral imaging strategy along with the proposed reconstruction algorithm and experimental results.

## 2.1 Notations

In order to describe the compressive hyperspectral image acquisition process, suppose the  $i^{th}$  band of an  $n_x \times n_y$ -pixel hyperspectral image is represented by the vector  $\mathbf{x}_i \in \mathbb{R}^{n_x n_y}$ , that is,  $\mathbf{x}_i$  is a column-wise concatenation of the pixels in the  $i^{th}$  band of the hyperspectral image. A hyperspectral image containing  $n_\lambda$  bands can then be represented as a matrix  $\mathbf{X} \in \mathbb{R}^{n_x n_y \times n_\lambda}$ :

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_{\lambda} \end{bmatrix}.$$

Let us define the linear operator  $\mathcal{A} : \mathbb{R}^{n_x n_y \times n_\lambda} \rightarrow \mathbb{R}^m$ . Since compressive sensing theory deals with linear sampling schemes, any compressive hyperspectral image acquisition process can therefore be modelled by:

$$\mathbf{y} = \mathcal{A}(\mathbf{X}) + \boldsymbol{\eta} \quad (2.1)$$

where  $\mathbf{y} \in \mathbb{R}^m$  is the vector containing  $m$  linear measurements of  $\mathbf{X}$ , corrupted by the noise vector  $\boldsymbol{\eta} \in \mathbb{R}^m$  inherent to the acquisition process. Alternatively, one can also write (2.1) equivalently as the more familiar matrix-vector product:

$$\mathbf{y} = \mathbf{A} \operatorname{vec}(\mathbf{X}) + \boldsymbol{\eta} \quad (2.2)$$

where  $\operatorname{vec}(\cdot)$  performs a column-wise stacking on its argument:

$$\operatorname{vec}(\mathbf{X}) = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_{n_\lambda} \end{bmatrix}$$

Here, the linear operator  $\mathcal{A}$  is represented as the matrix  $\mathbf{A} \in \mathbb{R}^{m \times n_x n_y n_\lambda}$  so that each element in  $\mathbf{y}$  corresponds to an inner product between the hyperspectral image  $\mathbf{X}$ , and the corresponding row of  $\mathbf{A}$ .

## 2.2 Survey of CS Hyperspectral Imaging Techniques

As discussed earlier, since currently available imaging sensors can only capture two out of the three dimensions of the hyperspectral imagery simultaneously, the remaining dimension must therefore be acquired over time. This means the compressive measurements are usually obtained in a partitioned manner, where the measurements only associate to a subset entry of the hyperspectral data to be captured, such as the image of one spectral band, or the spectral contents along one spatial dimension. The matrix  $\mathbf{A}$  in (2.2) is therefore usually structured. Existing designs for CS acquisition systems of hyperspectral images can be

broadly categorised into systems whose measurements are compressive in (1) the spatial directions only, (2) the spectral direction only, and, less commonly, (3) both spatial and spectral directions. In the following sections, we detail some of the proposed CS architectures representative of each category and the characterisation of their corresponding measurement matrix  $\mathbf{A}$  in (2.2).

### 2.2.1 The Single-Pixel Hyperspectral Camera

Perhaps the most widely used architecture for compressive imaging is the well-known Single-Pixel Camera (SPC) [60]. Figure 2.1 depicts the schematic overview of the SPC system for acquisition of a 2D grayscale image. The incoming light-field is first focused by a lens onto a Digital Micro-mirror Device (DMD). The DMD consists of many programmable micro-mirrors that can be positioned to either reflect the incoming light towards or away from the second lens. Each mirror is dedicated to reflecting a certain pixel in the image and the number of mirrors therefore defines the resolution of the acquired image. The reflected light is focused onto a photodiode and the measurement is therefore a superposition of the values of the pixels that are reflected towards the second lens. Finally, the measurements are quantized by an A/D converter and transmitted for reconstruction. In essence, the DMD performs an optical inner product between a measurement pattern  $\phi_i \in \mathbb{R}^{n_x n_y}$  on the DMD and the image being acquired  $\mathbf{x} \in \mathbb{R}^{n_x n_y}$  to give a single measurement in the form of  $\tilde{y}_i = \phi^\top \mathbf{x}$ . The entries of  $\phi_i$  are either 1 or 0 corresponding to the mirrors that reflect the incident light towards or away from the second lens respectively. The process is then repeated  $\tilde{m}$  times using different measurement patterns resulting in a complete CS measurement vector

$$\tilde{\mathbf{y}} = \begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \vdots \\ \tilde{y}_{\tilde{m}} \end{bmatrix} = \begin{bmatrix} \phi_1^\top \\ \phi_2^\top \\ \vdots \\ \phi_{\tilde{m}}^\top \end{bmatrix} \mathbf{x}$$

Since the DMD is programmable, apart from the Rademacher patterns, other measurement patterns that satisfies the RIP condition (see Section 1.1.2) may also be realised including

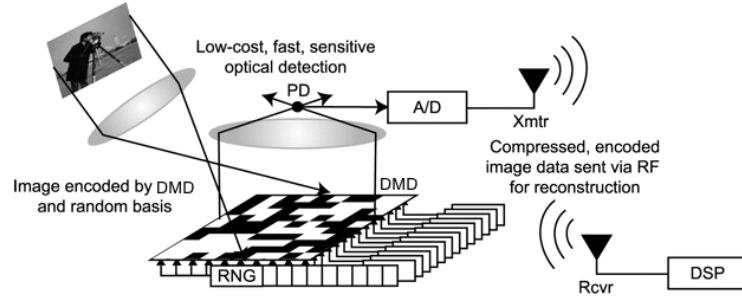


Figure 2.1: The schematic view of the single-pixel camera [60].

the Bernoulli patterns, Walsh-Hadamard orthogonal basis, and an approximation of Gaussian measurement vectors [61].

The SPC architecture can easily be extended to capture hyperspectral images by simply replacing the single photodiode in Figure 2.1 with a spectrometer [62]. This produces  $n_\lambda$  measurement values for each of the measurement patterns  $\{\phi_i\}_{i=1}^{\tilde{m}}$ , one for each spectral band  $\{x_j\}_{j=1}^{n_\lambda}$ , that is,

$$\begin{bmatrix} \tilde{y}_1 & \tilde{y}_2 & \dots & \tilde{y}_{n_\lambda} \end{bmatrix} = \begin{bmatrix} \phi_1^\top \\ \phi_2^\top \\ \vdots \\ \phi_{\tilde{m}}^\top \end{bmatrix} \begin{bmatrix} x_1 & x_2 & \dots & x_{n_\lambda} \end{bmatrix} = \Phi \mathbf{X}$$

In practice, the spectrometer is usually implemented via a spectral dispersion after the second lens, followed by a linear array of photodiodes. Effectively, each spectral band is sampled separately by the same measurement matrix  $\Phi$  and consequently, the acquisition is compressive along the two spatial directions only. The entire acquisition process can be modelled by (2.2) as:

$$\underbrace{\begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \vdots \\ \tilde{y}_{n_\lambda} \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} \Phi & & & \\ & \Phi & & \\ & & \ddots & \\ & & & \Phi \end{bmatrix}}_{\mathbf{A} \in \mathbb{R}^{m \times n_x n_y n_\lambda}} \text{vec}(\mathbf{X}) + \underbrace{\begin{bmatrix} \tilde{\eta}_1 \\ \tilde{\eta}_2 \\ \vdots \\ \tilde{\eta}_{n_\lambda} \end{bmatrix}}_{\boldsymbol{\eta}} \quad (2.3)$$

with  $m = \tilde{m} n_\lambda$ .

## 2.2.2 Compressive Pushbroom/Whiskbroom Hyperspectral Imagers

While the Single-Pixel Hyperspectral Camera (SPHC) acquires measurements that are compressive across the spatial directions, several designs have also been proposed to make compressive measurements along the spectral direction. For example, Figure 2.2 shows a schematic overview of the pushbroom compressive hyperspectral imager proposed in [63]. Like the SPHC, a DMD is employed to perform the optical compression, however, instead of passing the entire spatial image onto the DMD, one row of the image is passed through a split aperture at a given time to a diffractive element (e.g., a prism) that spreads the spectrum of each pixel in the row onto each column of the DMD. The number of rows on the DMD therefore defines the spectral resolution  $n_\lambda$  of the hyperspectral image while the number of columns determines its horizontal spatial resolution  $n_x$ . The reflections from each column of the DMD are then focused onto  $n_x$  corresponding photodiodes via a cylindrical lens to obtain  $n_x$  measurements that are compressive along the spectral direction, one for each spatial pixel in the current row of the image. Each column can therefore use the same pattern  $\phi_i \in \mathbb{R}^{n_\lambda}$  (although different patterns can be substituted) as they are associated with individual voxels. At the  $j$ th row of the image, we can model the acquisition as

$$\begin{bmatrix} (\tilde{y}_i)_{n_x(j-1)+1} \\ (\tilde{y}_i)_{n_x(j-1)+2} \\ \vdots \\ (\tilde{y}_i)_{n_x(j-1)+n_x} \end{bmatrix} = \begin{bmatrix} \phi_i^\top & & & \\ & \phi_i^\top & & \\ & & \ddots & \\ & & & \phi_i^\top \end{bmatrix} \tilde{\mathbf{P}}_j \text{vec}(\mathbf{X})$$

where  $\tilde{\mathbf{P}}_j \in \mathbb{R}^{n_x n_\lambda \times n_x n_y n_\lambda}$  is a partial permutation matrix that rearranges the entries of  $\text{vec}(\mathbf{X})$  so that  $\phi_i^\top$  acts on each of the voxels in the  $j$ th row of the image, that is, every consecutive  $n_\lambda$  elements of  $\tilde{\mathbf{P}}_j \text{vec}(\mathbf{X})$  corresponds to a voxel in the  $j$ th row. The pattern  $\phi_i$  on the DMD is then changed  $\tilde{m}$  times to obtain  $\tilde{m}$  CS measurements for each individual voxel. Mathematically, this is equivalent to applying the same measurement matrix  $\Phi \in \mathbb{R}^{\tilde{m} \times n_\lambda}$  to

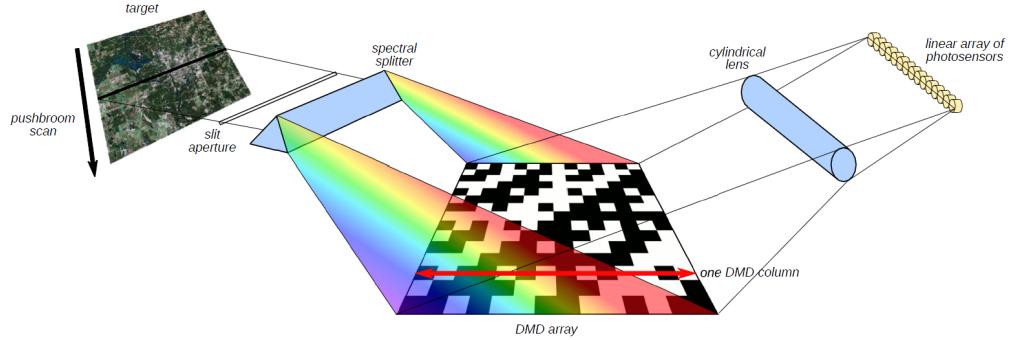


Figure 2.2: A compressive whiskbroom hyperspectral sensor [63].

each voxel and can be written as (after re-ordering the elements)

$$\begin{bmatrix} (\tilde{\mathbf{y}}_i)_{n_x(j-1)+1} \\ (\tilde{\mathbf{y}}_i)_{n_x(j-1)+2} \\ \vdots \\ (\tilde{\mathbf{y}}_i)_{n_x(j-1)+n_x} \end{bmatrix} = \begin{bmatrix} \Phi & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \Phi \end{bmatrix} \tilde{\mathbf{P}}_j \text{vec}(\mathbf{X})$$

The sensor then advances to the next line where the whole process is repeated for all rows  $j = 1, \dots, n_y$ . This entire acquisition can be modelled in a form equivalently to (2.2) as:

$$\underbrace{\begin{bmatrix} \tilde{\mathbf{y}}_1 \\ \tilde{\mathbf{y}}_2 \\ \vdots \\ \tilde{\mathbf{y}}_{n_x n_y} \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} \Phi & & & \\ & \Phi & & \\ & & \ddots & \\ & & & \Phi \end{bmatrix}}_{\mathbf{A} \in \mathbb{R}^{m \times n_x n_y n_\lambda}} \begin{bmatrix} \tilde{\mathbf{P}}_1 \\ \tilde{\mathbf{P}}_2 \\ \vdots \\ \tilde{\mathbf{P}}_{n_y} \end{bmatrix} \text{vec}(\mathbf{X}) + \underbrace{\begin{bmatrix} \tilde{\boldsymbol{\eta}}_1 \\ \tilde{\boldsymbol{\eta}}_2 \\ \vdots \\ \tilde{\boldsymbol{\eta}}_{n_x n_y} \end{bmatrix}}_{\boldsymbol{\eta}} \quad (2.4)$$

where  $m = \tilde{m} n_x n_y$ .

Alternatively, in [64] the authors suggested replacing the cylindrical lens in Figure 2.2 with a regular converging lens and use a single photodiode for sensing as in the SPC. This would result in measurements that are compressive across a 2D spatial-spectral slice of the hyper-

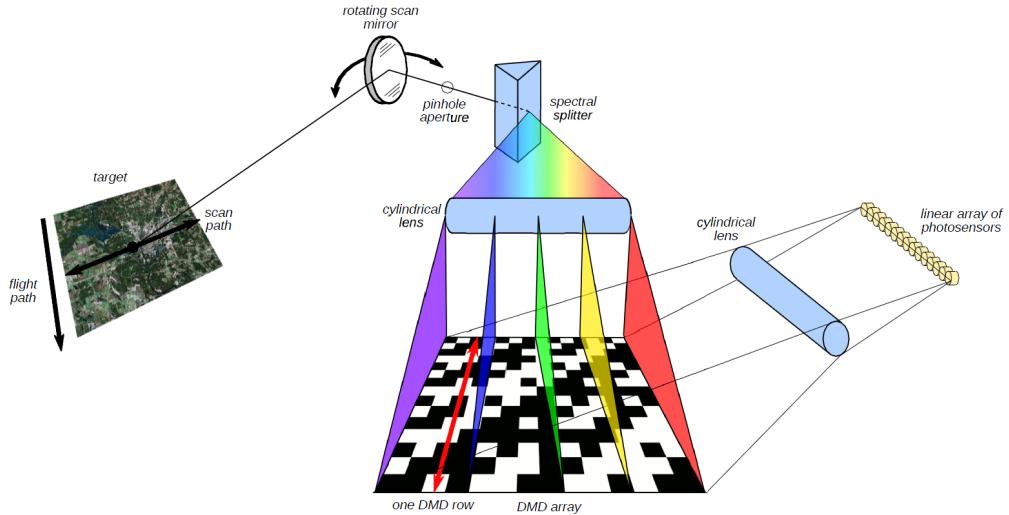


Figure 2.3: A compressive whiskbroom hyperspectral sensor [63].

spectral data and can also be modelled in a form very similar to before:

$$\underbrace{\begin{bmatrix} \tilde{\mathbf{y}}_1 \\ \tilde{\mathbf{y}}_2 \\ \vdots \\ \tilde{\mathbf{y}}_{n_x n_y} \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} \Phi' & & & \\ & \Phi' & & \\ & & \ddots & \\ & & & \Phi' \end{bmatrix}}_{\mathbf{A} \in \mathbb{R}^{m \times n_x n_y n_\lambda}} \underbrace{\begin{bmatrix} \tilde{\mathbf{P}}_1 \\ \tilde{\mathbf{P}}_2 \\ \vdots \\ \tilde{\mathbf{P}}_{n_y} \end{bmatrix}}_{\text{vec}(\mathbf{X})} + \underbrace{\begin{bmatrix} \tilde{\boldsymbol{\eta}}_1 \\ \tilde{\boldsymbol{\eta}}_2 \\ \vdots \\ \tilde{\boldsymbol{\eta}}_{n_x n_y} \end{bmatrix}}_{\boldsymbol{\eta}} \quad (2.5)$$

with  $m = \tilde{m} n_y$ , and  $\Phi' \in \mathbb{R}^{\tilde{m} \times n_x n_y n_\lambda}$  where each row of  $\Phi'$  corresponds to a pattern for the entire DMD array and not just the columns.

Another popular hyperspectral imaging paradigm is the whiskbroom scan, which captures a single voxel at a time, sweeping across the scan line perpendicular to the motion of the sensor. A rotating mirror is employed to facilitate the sweep. As shown in Figure 2.3, [63] has extended this setup to compressive acquisition by passing the light emitted from the pinhole aperture through a spectral splitter, thereby dispersing the spectral contents across the DMD as parallel rays via a cylindrical lens. Since each row of the DMD array corresponds to a different spectral band, a different measurement pattern  $\{\phi_i\}_{i=1}^{\tilde{m}} \in \mathbb{R}^{n_\lambda}$  is placed in each column of the DMD, allowing simultaneous acquisition of  $\tilde{m}$  compressive measurements for the particular voxel. This is then repeated for every voxel in the image, and the entire process can be characterised by the exact same model as (2.4) after re-ordering the elements.

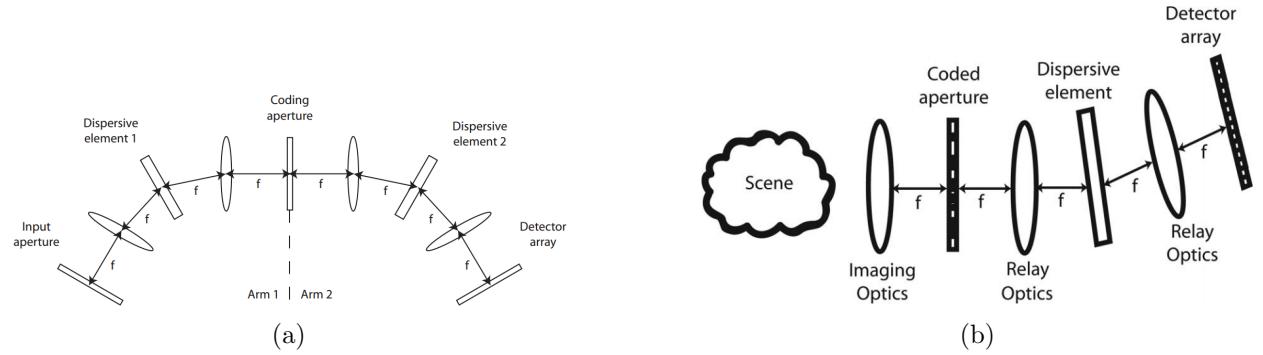


Figure 2.4: Schematic overview of the Coded Aperture Snapshot Spectral Imager (CASSI). (a) The Dual-Disperser configuration (DD-CASSI) [65], and (b) the Single-Disperser configuration (SD-CASSI) [66].

### 2.2.3 Coded Aperture Snapshot Spectral Imager

As an alternative to the architectures in the previous sections, which use linear arrays of photo-sensors to acquire the compressive measurements, the Coded Aperture Snapshot Spectral Imager (CASSI) captures compressive measurements of hyperspectral data using a 2D array of photo-sensors [65, 66]. Figure 2.4(a) shows the schematic overview of the Dual-Disperser Coded Aperture Snapshot Spectral Imager (DD-CASSI) architecture. The light-field of a scene is passed through two sequentially dispersive elements (e.g., Amici prisms) arranged in opposition so that the dispersion in the second element cancels the dispersion introduced by the first. A coded aperture composed of  $n_x \times n_y$  blocks of randomly distributed 0 and 1 patterns is placed in between the dispersive elements to effect the compressive measurement. Since the hyperspectral data is spectrally sheared before passing through the coded pattern, it is therefore modulated at various spatial and spectral locations. The second dispersive element then undoes the dispersion and the  $n_x \times n_y$  array of photo-sensors then sums up the modulated voxels at each spatial location. Effectively, each voxel is modulated by a different measurement pattern  $(\phi_i)_j \in \mathbb{R}^{n_\lambda}$ , for  $j = 1, \dots, n_x n_y$  and the single snapshot

results in one compressive measurement for each of the voxels:

$$\begin{bmatrix} (\tilde{y}_i)_1 \\ (\tilde{y}_i)_2 \\ \vdots \\ (\tilde{y}_i)_{n_x n_y} \end{bmatrix} = \begin{bmatrix} (\boldsymbol{\phi}_i)_1^\top & & & \\ & (\boldsymbol{\phi}_i)_2^\top & & \\ & & \ddots & \\ & & & (\boldsymbol{\phi}_i)_{n_x n_y}^\top \end{bmatrix} \text{vec}(\mathbf{X}^\top)$$

The coded aperture mask is then replaced  $\tilde{m}$  times and  $\tilde{m}$  snapshots are obtained. Note that instead of changing the masks physically, it can also be implemented using the DMD as proposed in [67]. Since the signal at each photo-sensor is a sum of the corresponding modulated voxels, the measurements are only compressive along the spectral direction, and can be modelled in a similar way to (2.4) but with different measurement matrices  $\Phi$  for each voxel:

$$\underbrace{\begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \vdots \\ \tilde{y}_{n_x n_y} \end{bmatrix}}_y = \underbrace{\begin{bmatrix} \Phi_1 & & & \\ & \Phi_2 & & \\ & & \ddots & \\ & & & \Phi_{n_x n_y} \end{bmatrix}}_A \begin{bmatrix} \tilde{\mathbf{P}}_1 \\ \tilde{\mathbf{P}}_2 \\ \vdots \\ \tilde{\mathbf{P}}_{n_y} \end{bmatrix} \text{vec}(\mathbf{X}) + \underbrace{\begin{bmatrix} \tilde{\boldsymbol{\eta}}_1 \\ \tilde{\boldsymbol{\eta}}_2 \\ \vdots \\ \tilde{\boldsymbol{\eta}}_{n_x n_y} \end{bmatrix}}_\eta \quad (2.6)$$

The Single-Disperser Coded Aperture Snapshot Spectral Imager (SD-CASSI) shown in Figure 2.4(b) on the other hand, makes measurements that are compressive both spatially and spectrally. In contrast to the DD-CASSI, the light-field from the scene is first modulated by the coded aperture before being dispersed spectrally. Instead of having a second dispersive element to cancel this dispersion, the overlapping images at distinct wavelengths due to the dispersion are directly acquired by each photo-sensor. The measurements at each detector from the single snapshot are therefore a combination of different modulated voxels of the

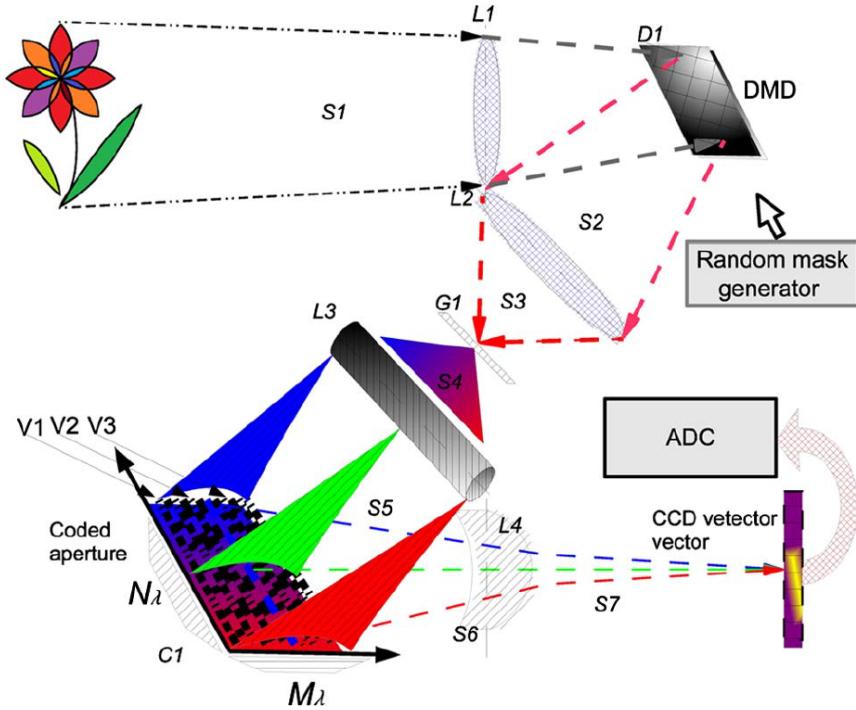


Figure 2.5: The schematic diagram of the CHISSL system [68].

scene. Mathematically, this process can simply be written as:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \underbrace{\begin{bmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_{n_\lambda} \end{bmatrix}}_A \text{vec}(\mathbf{X}) + \boldsymbol{\eta} \quad (2.7)$$

where  $\Phi \in \mathbb{R}^{m \times n_x n_y n_\lambda}$ .

## 2.2.4 Compressive Hyperspectral Imaging by Separable Spatial and Spectral Operators

More recently, the author of [68] proposed an acquisition scheme shown in Figure 2.5 that allows compressive measurements along both spatial and spectral directions. The design is equivalent to a cascade of the SPC (Section 2.2.1) followed by the compressive whiskbroom imager (Section 2.2.2). For each measurement pattern  $\{\phi_i\}_{i=1}^{\tilde{m}_1}$  on the DMD (D1), the spatially modulated signal is focused onto a diffractive or dispersive element (G1), dispersing the

spatially modulated output into  $n_\lambda$  spectral spots. At this point, each spectral spot contains the modulated spatial data for each wavelength, encoded using the same pattern  $\phi_i$ :

$$\begin{bmatrix} (\tilde{y}_i)_1 & (\tilde{y}_i)_2 & \dots & (\tilde{y}_i)_{n_\lambda} \end{bmatrix} = \phi_i^\top \mathbf{X}.$$

By means of a cylindrical lens (L3), these spectral spots are spread along  $n_\lambda$  parallel horizontal rays onto a coded aperture (C1) consisting of  $\tilde{m}_2$  vertical lines. Each of the  $\tilde{m}_2$  vertical lines is encoded with a different pattern  $\{\tilde{\phi}_j\}_{j=1}^{\tilde{m}_2} \in \mathbb{R}^{n_\lambda}$ , which modulates the  $n_\lambda$  spectral spots, performing compression along the spectral direction:

$$\begin{bmatrix} (\tilde{y}_i)_1 & (\tilde{y}_i)_2 & \dots & (\tilde{y}_i)_{\tilde{m}_2} \end{bmatrix} = \begin{bmatrix} (\tilde{y}_i)_1 & (\tilde{y}_i)_2 & \dots & (\tilde{y}_i)_{n_\lambda} \end{bmatrix} \begin{bmatrix} \tilde{\phi}_1 & \tilde{\phi}_2 & \dots & \tilde{\phi}_{\tilde{m}_2} \end{bmatrix}$$

The  $\tilde{m}_2$  spectrally modulated samples are then summed by means of another cylindrical lens (L4) and collected by the appropriate photo-sensors in a linear array. The DMD (D1) then changes its pattern for  $1 \leq i \leq \tilde{m}_1$  resulting in a total of  $m = \tilde{m}_1 \tilde{m}_2$  measurements

$$\underbrace{\begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \dots & \mathbf{y}_{\tilde{m}_2} \end{bmatrix}}_{\Phi_1 \in \mathbb{R}^{\tilde{m}_1 \times n_x n_y}} = \begin{bmatrix} \phi_1^\top \\ \phi_2^\top \\ \vdots \\ \phi_{\tilde{m}_1}^\top \end{bmatrix} \mathbf{X} \underbrace{\begin{bmatrix} \tilde{\phi}_1 & \tilde{\phi}_2 & \dots & \tilde{\phi}_{\tilde{m}_2} \end{bmatrix}}_{\Phi_2 \in \mathbb{R}^{n_\lambda \times \tilde{m}_2}}$$

which can be equivalently expressed in the form of (2.2) as

$$\underbrace{\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_{\tilde{m}_2} \end{bmatrix}}_{\mathbf{y}} = \underbrace{(\Phi_2^\top \otimes \Phi_1)}_{\mathbf{A} \in \mathbb{R}^{m \times n_x n_y n_\lambda}} \text{vec}(\mathbf{X}) + \underbrace{\begin{bmatrix} \tilde{\eta}_1 \\ \tilde{\eta}_2 \\ \vdots \\ \tilde{\eta}_{\tilde{m}_2} \end{bmatrix}}_{\boldsymbol{\eta}} \quad (2.8)$$

where  $\otimes$  denotes the matrix Kronecker product.

## 2.3 Discussion of Limitations and Difficulties

One problem that arises from the partitioning of measurements witnessed in the existing acquisition setups is that sparsity models that exploit structures beyond the measured dimension(s) are enforcing what is not actually observed during the acquisition process: *the measurements are independent of adjacent channels and only contain information within the dimension(s) in which they are obtained.* For example, the first  $\tilde{m}$  elements of  $\mathbf{y}$  obtained using the single-pixel hyperspectral imager (Section 2.2.1) only correspond to measurements taken from the first spectral band, while the second  $\tilde{m}$  elements only contain information from the second spectral band and so on. These measurements therefore only reflect *local* information of a specific band. Similarly, each group of  $\tilde{m}$  measurements obtained from the pushbroom, whiskbroom (Section 2.2.2) and DD-CASSI (Section 2.2.3) only reflect information of a specific voxel. The correlations between adjacent channels are therefore inherently lacking in these measurements and the CS guarantees associated with models that exploit sparsity beyond the dimension(s) in which the measurements are made may not necessarily hold since these measurements do not reflect *global* information of the underlying structure needed to capture the required sparsity structure [37]. Take the extreme example in which the signal in which the signal  $\mathbf{X}$  contains identical rows (i.e.,  $\mathbf{X}$  is rank one) corresponding to an ideal scene consisting of entirely one material. Suppose we use the pushbroom design to acquire  $m = n_x n_y \tilde{m}$  compressive measurements as (2.4),  $\tilde{m}$  measurements for each voxel. Since a unique sensing matrix  $\Phi$  is used for all voxels, this would result in  $n_x n_y$  copies of the same  $\tilde{m}$  measurements and exploiting any inter-channel correlations beyond the individual voxels is therefore pointless because the measurements are highly redundant and do not include any new information across channels.

To overcome the aforementioned problem, one could use different sensing matrices on each channel instead so that the measurements of different channels carry diverse information for recovery. In Section 2.6 we experimentally show that by using such sensing strategy does indeed lead to improvements when exploiting sparsity beyond the measured dimension(s) compared to using the same sensing matrix for all channels. The measurements, however,

still do not contain information along the dimension in which they are partitioned. Furthermore, in order to emulate such sensing strategy, the acquisition hardware must generate independent patterns for each set of measurements, and this may not be suitable for designs that capture simultaneous measurements across multiple channels for a given pattern (such as the single-pixel hyperspectral imager) since the acquisition time would be significantly extended.

Measurements that are compressive in both spatial and spectral directions on the other hand are difficult to realise in hardware due to the delicate optical setup required to perform the actual measurements as seen in the SD-CASSI (Section 2.2.3) and the separable spatial-spectral imager (Section 2.2.4). The physical setup in SD-CASSI is accompanied by commensurate losses in spatial resolution due to the dispersive shearing, and inferior recovery performance as a result of severe spatial and spectral ambiguities in the measurements with the spatial multiplexing being localised as opposed to the global multiplexing that CS theory needs [66]. Furthermore, the CASSI architecture requires an array of photo-sensors of the same size as the spatial dimensions of the image, in contrast to the much simpler linear array employed in setups that makes compressive measurements along one (spatial or spectral) direction.

The limitations and difficulties observed from both ends of the spectrum raises a natural question of whether it is possible to obtain compressive measurements that contain information about all dimensions of the hyperspectral data but without a significant increase in the acquisition complexity. That is, can we get the best of both worlds?

## 2.4 Multi-modal Compressive Hyperspectral Imaging

Perhaps the most natural and practical way to achieve this is by combining multiple sets of partitioned compressive measurements that are made along different dimensions, that is, a fusion of the sets of measurements taken from different perspectives, commonly referred to as *multimodal imaging*. In this regard, one may simply utilize a combination of the previously mentioned setups to obtain the necessary measurements in practice. Such a strategy means that, for the same number of measurements, the measurements themselves would contain a

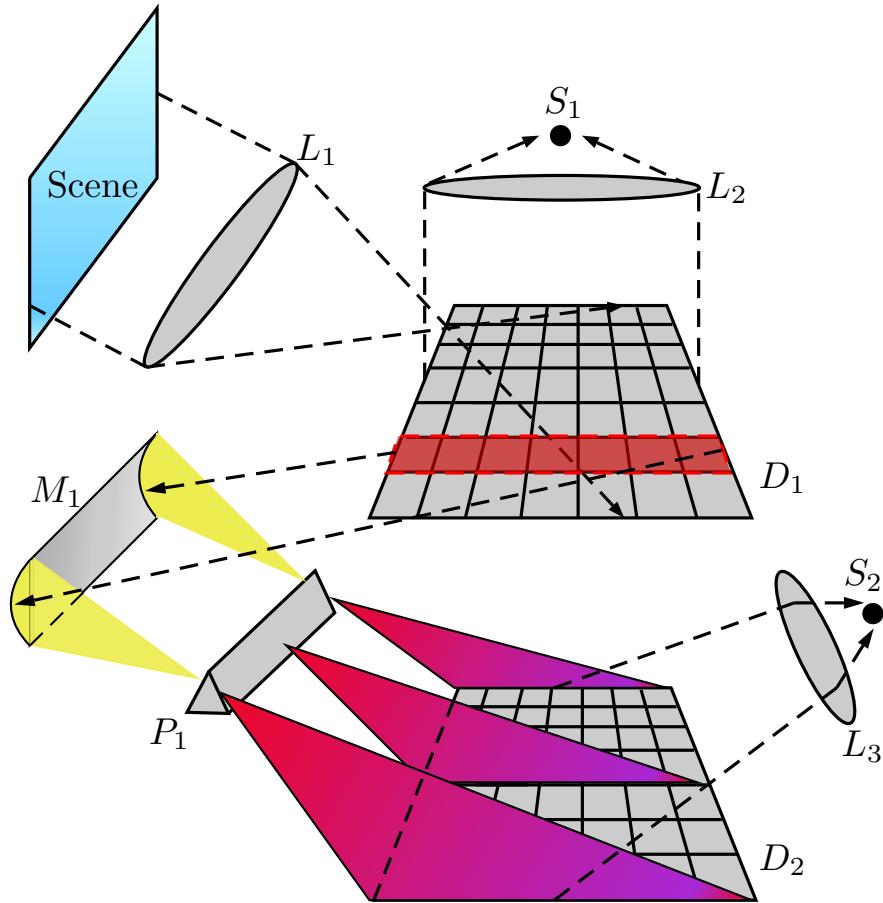


Figure 2.6: Schematic overview of a multi-modal acquisition system.

mixture of information from *all* dimensions of the hyperspectral data, thus capturing existing sparsity structure from all dimensions. On the other hand, Figure 2.6 also demonstrates a plausible system that allows acquisition of measurements which are compressive along the spatial-spatial and spatial-spectral slices of the hyperspectral data. The description of the system is as follows: the incoming light-field from the scene is first focused by the lens  $L_1$  onto the DMD  $D_1$ . The DMD  $D_1$  is positioned such that mirrors in the “on” position reflect the incident light towards a spectrometer  $S_1$  via lens  $L_2$  thus obtaining compressive measurements in the spatial plane as in the single-pixel hyperspectral imager (Section 2.2.1). The mirrors in the “off” position, on the other hand, direct the rays towards a curved mirror  $M_1$ , which concentrates the incoming light onto a diffractive element  $P_1$ . In this configuration, the DMD  $D_1$  also acts as a pushbroom scanner whereby each row (or column) of  $D_1$  is placed in the “off” position sequentially to pass through a single line of the scene. The diffractive element  $P_1$  then spreads the spectrum of the line onto a second DMD  $D_2$

to effect compressive measurements in the one of the two spatial-spectral planes. These measurements are captured by the photo-sensor  $S_2$  through the final collector lens  $L_3$ .

### 2.4.1 Problem Formulation

Let  $x, y$  represent the spatial directions of the hyperspectral image and  $\lambda$  be its spectral direction, then let the quantity of any variable with a subscripted pair of  $x, y, \lambda$  be associated with the corresponding plane. For example,  $\Phi_{xy} \in \mathbb{R}^{m_{xy} \times n_x n_y}$  refers to a measurement matrix that operates on the entries in a particular slice of the  $x$ - $y$  plane of  $\mathbf{X}$ .

As seen from Section 2.2 (see also, [69–72] for examples), the measurement process in many existing compressive hyperspectral imaging setups, consists of distributing a unique measurement matrix  $\Phi_{xy}$  to each spatial slice of the hyperspectral image so that the resulting measurement  $\mathbf{y}_{xy}$  is a concatenation of measurements obtained from each spatial  $x$ - $y$  plane of  $\mathbf{X}$  (with  $m = m_{xy}n_\lambda$ ). In these cases the measurement matrix  $\mathbf{A}$  has the form

$$\mathbf{A}_{xy} = \begin{bmatrix} \Phi_{xy} & & & \\ & \Phi_{xy} & & \\ & & \ddots & \\ & & & \Phi_{xy} \end{bmatrix} = (\mathbf{I}_{n_\lambda} \otimes \Phi_{xy}) \quad (2.9)$$

where  $\mathbf{I}$  represents the identity matrix with the subscripted dimension. Of course, an independent matrix per band can also be used in theory, with  $\mathbf{A}$  taking the form

$$\mathbf{A}_{xy} = \begin{bmatrix} (\Phi_{xy})_1 & & & \\ & (\Phi_{xy})_2 & & \\ & & \ddots & \\ & & & (\Phi_{xy})_{n_\lambda} \end{bmatrix} \quad (2.10)$$

however, as mentioned previously, the acquisition hardware must therefore generate independent patterns for each set of measurements, which may significantly prolong the acquisition time.

Similarly, obtaining measurements in the spatial-spectral ( $y\text{-}\lambda$  and  $x\text{-}\lambda$ ) planes can be modelled as distributing measurement patterns  $\Phi_{y\lambda} \in \mathbb{R}^{m_{y\lambda} \times n_y n_\lambda}$  and  $\Phi_{x\lambda} \in \mathbb{R}^{m_{x\lambda} \times n_x n_\lambda}$  respectively along the corresponding planes of  $\mathbf{X}$ . The three sets of measurements from these different perspectives can therefore be written as

$$\mathbf{y}_{xy} = \mathbf{A}_{xy} \text{vec}(\mathbf{X}) + \boldsymbol{\eta}_{xy} \quad (2.11)$$

$$\mathbf{y}_{y\lambda} = \mathbf{A}_{y\lambda} \mathbf{P}_{y\lambda} \text{vec}(\mathbf{X}) + \boldsymbol{\eta}_{y\lambda} \quad (2.12)$$

$$\mathbf{y}_{x\lambda} = \mathbf{A}_{x\lambda} \mathbf{P}_{x\lambda} \text{vec}(\mathbf{X}) + \boldsymbol{\eta}_{x\lambda} \quad (2.13)$$

where  $\mathbf{y}_{xy}, \boldsymbol{\eta}_{xy} \in \mathbb{R}^{m_{xy} n_\lambda}$ ,  $\mathbf{y}_{y\lambda}, \boldsymbol{\eta}_{y\lambda} \in \mathbb{R}^{m_{y\lambda} n_x}$ ,  $\mathbf{y}_{x\lambda}, \boldsymbol{\eta}_{x\lambda} \in \mathbb{R}^{m_{x\lambda} n_y}$ , and  $\mathbf{P}_{y\lambda}, \mathbf{P}_{x\lambda} \in \mathbb{R}^{n_x n_y n_\lambda \times n_x n_y n_\lambda}$  are permutation matrices that rearrange the entries of  $\text{vec}(\mathbf{X})$  so that each column of the resulting unvectorized matrices  $\text{vec}^{-1}(\mathbf{P}_{y\lambda} \text{vec}(\mathbf{X}))$  and  $\text{vec}^{-1}(\mathbf{P}_{x\lambda} \text{vec}(\mathbf{X}))$  correspond to a slice of the hyperspectral image along the subscripted planes. While it is possible, as seen from existing literature, to recover the hyperspectral image  $\mathbf{X}$  directly from (2.11), (2.12), or (2.13) alone, in order to benefit from the inter-channel correlations, we take an alternative approach and seek to combine the measurement sets (2.11)-(2.13) together for joint recovery from multiple modalities. That is, we consider the measurements

$$\underbrace{\begin{bmatrix} \mathbf{y}_{xy} \\ \mathbf{y}_{y\lambda} \\ \mathbf{y}_{x\lambda} \end{bmatrix}}_y = \underbrace{\begin{bmatrix} \mathbf{A}_{xy} & & \\ & \mathbf{A}_{y\lambda} & \\ & & \mathbf{A}_{x\lambda} \end{bmatrix}}_A \underbrace{\begin{bmatrix} \mathbf{P}_{xy} \\ \mathbf{P}_{y\lambda} \\ \mathbf{P}_{x\lambda} \end{bmatrix}}_{\mathbf{P}} \text{vec}(\mathbf{X}) + \underbrace{\begin{bmatrix} \boldsymbol{\eta}_{xy} \\ \boldsymbol{\eta}_{y\lambda} \\ \boldsymbol{\eta}_{x\lambda} \end{bmatrix}}_\eta \quad (2.14)$$

where  $\mathbf{P}_{xy}$  is simply the identity  $\mathbf{I}_{n_x n_y n_\lambda}$ . The contention is that, for the same number of measurements, recovery models that use (2.14) would lead to significant improvements in the reconstruction quality over using (2.11), (2.12), or (2.13) alone because (2.14) contains a mixture of information from all dimensions and thus captures more structural information about the hyperspectral image. In essence, (2.14) is a generalization of the measurement scheme (2.11)-(2.13), where they are simply special cases of (2.14) with the corresponding  $\mathbf{A}$  being non-zero.

## 2.5 Multi-modal CoHSI Reconstruction Algorithm

In order to evaluate and compare the performance of the proposed multi-modal acquisition framework to the standard uni-modal setups we discuss here our algorithmic framework based on the Alternating Direction Method of Multipliers (ADMM) for solving three general convex minimization models in literature for Hyperspectral Image (HSI) recovery incorporated with the proposed multimodal measurements. The three models we considered are listed below, details of these recovery schemes are discussed in Chapter 3.

- (i) 2D-Total Variation (TV) norm minimization where the HSI is recovered from its CS measurements by finding a solution that minimizes the TV norm of each individual spectral band of the hyperspectral image [73]:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \sum_{i=1}^{n_\lambda} \|\mathbf{x}_i\|_{TV} \quad \text{s.t.} \quad \begin{aligned} \|\mathbf{y}_{xy} - \mathbf{A}_{xy} \text{vec}(\mathbf{X})\|_2 &\leq \epsilon_{xy} \\ \|\mathbf{y}_{y\lambda} - \mathbf{A}_{y\lambda} \mathbf{P}_{y\lambda} \text{vec}(\mathbf{X})\|_2 &\leq \epsilon_{y\lambda} \\ \|\mathbf{y}_{x\lambda} - \mathbf{A}_{x\lambda} \mathbf{P}_{x\lambda} \text{vec}(\mathbf{X})\|_2 &\leq \epsilon_{x\lambda} \end{aligned} \quad (2.15)$$

- (ii) 3D-Wavelet minimization where the HSI is recovered by finding a solution with a sparse representation in a 3D Wavelet basis [68,74]. This is achieved by minimizing the  $l_1$ -norm of transformed coefficients of the HSI:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \|\Psi_{3D}^\top \text{vec}(\mathbf{X})\|_1 \quad \text{s.t.} \quad \begin{aligned} \|\mathbf{y}_{xy} - \mathbf{A}_{xy} \text{vec}(\mathbf{X})\|_2 &\leq \epsilon_{xy} \\ \|\mathbf{y}_{y\lambda} - \mathbf{A}_{y\lambda} \mathbf{P}_{y\lambda} \text{vec}(\mathbf{X})\|_2 &\leq \epsilon_{y\lambda} \\ \|\mathbf{y}_{x\lambda} - \mathbf{A}_{x\lambda} \mathbf{P}_{x\lambda} \text{vec}(\mathbf{X})\|_2 &\leq \epsilon_{x\lambda} \end{aligned} \quad (2.16)$$

where  $\Psi_{3D} = (\Psi_{1D} \otimes \Psi_{2D})$ , is the 3D Wavelet basis constructed from the matrix Kronecker product of the 1D spectral and the 2D spatial Wavelet bases.

- (iii) Joint TV-Nuclear norm minimization where the HSI is reconstructed by using (i) reg-

ularized with the *nuclear norm*<sup>1</sup> of the HSI [75]:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \sum_{i=1}^{n_\lambda} \|\mathbf{x}_i\|_{TV} + \gamma \|\mathbf{X}\|_* \quad \text{s.t.} \quad \begin{aligned} \|\mathbf{y}_{xy} - \mathbf{A}_{xy} \text{vec}(\mathbf{X})\|_2 &\leq \epsilon_{xy} \\ \|\mathbf{y}_{y\lambda} - \mathbf{A}_{y\lambda} \mathbf{P}_{y\lambda} \text{vec}(\mathbf{X})\|_2 &\leq \epsilon_{y\lambda} \\ \|\mathbf{y}_{x\lambda} - \mathbf{A}_{x\lambda} \mathbf{P}_{x\lambda} \text{vec}(\mathbf{X})\|_2 &\leq \epsilon_{x\lambda} \end{aligned} \quad (2.17)$$

where  $\gamma$  is a scalar parameter controlling the balance between the two norms.

### 2.5.1 Alternating Direction Method of Multiplier

While there are numerous methods for solving optimization problems involving non-differentiable convex objectives paired with non-linear constraints such as (2.15)-(2.17), the ADMM has received considerable attention recently for its ability to solve problems of such kind in a *decentralized* manner [76]. As we will see in the following section, this advantage is exploited in our formulation whereby enabling the bulk of the recovery process to be carried out in parallel - a quality highly desirable in hyperspectral processing due to the complexity and size of the datasets.

In a nutshell, ADMM is a variant of the family of algorithms known as the *augmented Lagrangian methods* [77] (also known as the Method of Multiplier (MM)) that solves problems often in the form of:

$$\hat{\boldsymbol{\nu}} = \arg \min_{\boldsymbol{\nu}} f(\boldsymbol{\nu}) + g(\mathbf{G}\boldsymbol{\nu}) \quad (2.18)$$

where  $\mathbf{G} \in \mathbb{R}^{d \times n}$ , and  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  are lower semi-continuous functions. (2.18) can be equivalently written, via a change of variable, as the constrained problem:

$$\hat{\boldsymbol{\nu}} = \arg \min_{\boldsymbol{\nu}, \mathbf{z}} f(\boldsymbol{\nu}) + g(\mathbf{B}\mathbf{z}) \quad \text{subject to} \quad \mathbf{B}\mathbf{z} = \mathbf{G}\boldsymbol{\nu} \quad (2.19)$$

where  $\mathbf{B} \in \mathbb{R}^{d \times p}$ . Despite this seemingly trivial change, the MM now solves (2.19) by forming the so-called *augmented Lagrangian* of (2.19):

---

<sup>1</sup>The nuclear norm of a matrix is the  $l_1$ -norm of its singular values. Minimizing the nuclear norm therefore promotes sparsity in the singular values and consequently results in matrices with a low-rank representation.

**Algorithm 2.1** The ADMM Algorithm

- 
- 1: Set  $k = 0$ ,  $\mathbf{u}_0$ ,  $\mathbf{z}_0$ , and choose  $\rho > 0$
  - 2: **repeat**
  - 3:    $\mathbf{z}_{k+1} = \arg \min_{\mathbf{z}} g(\mathbf{Bz}) + \frac{\rho}{2} \|\mathbf{G}\nu_k - \mathbf{Bz} + \mathbf{u}_k\|_2^2$
  - 4:    $\nu_{k+1} = \arg \min_{\nu} f(\nu) + \frac{\rho}{2} \|\mathbf{G}\nu - \mathbf{Bz}_{k+1} + \mathbf{u}_k\|_2^2$
  - 5:    $\mathbf{u}_{k+1} = \mathbf{u}_k + (\mathbf{G}\nu_{k+1} - \mathbf{Bz}_{k+1})$
  - 6:    $k = k + 1$
  - 7: **until** stopping criterion is satisfied
- 

$$L_\rho(\boldsymbol{\nu}, \mathbf{z}, \mathbf{w}) \stackrel{\text{def}}{=} f(\boldsymbol{\nu}) + g(\mathbf{Bz}) + \mathbf{w}^\top (\mathbf{G}\boldsymbol{\nu} - \mathbf{Bz}) + \frac{\rho}{2} \|\mathbf{G}\boldsymbol{\nu} - \mathbf{Bz}\|_2^2$$

where  $\mathbf{w}$  is the associated Lagrange multiplier and  $\rho > 0$  is a scalar constant. Note that  $L_\rho$  is usually written in a more compact form (via completing the squares):

$$L_\rho(\boldsymbol{\nu}, \mathbf{z}, \mathbf{u}) \stackrel{\text{def}}{=} f(\boldsymbol{\nu}) + g(\mathbf{Bz}) + \frac{\rho}{2} \|\mathbf{G}\boldsymbol{\nu} - \mathbf{Bz} + \mathbf{u}\|_2^2 - \frac{\rho}{2} \|\mathbf{u}\|_2^2$$

with the scaled Lagrange multiplier  $\mathbf{u} = \mathbf{w}/\rho$ . The MM finds the solution to (2.19) by iterating between minimizing  $L_\rho(\boldsymbol{\nu}, \mathbf{z}, \mathbf{u})$  with respect to  $(\boldsymbol{\nu}, \mathbf{z})$  while keeping  $\mathbf{u}$  fixed, and updating  $\mathbf{u}$  for the given  $\boldsymbol{\nu}$  and  $\mathbf{z}$  until the designated stopping criterion is satisfied. In general, however, the joint minimization with respect to  $(\boldsymbol{\nu}, \mathbf{z})$  in MM is not trivial (since it involves a non-separable quadratic term) and the so-called ADMM, as its name suggests, thus approximates this joint minimization by alternately minimizing with respect to  $\boldsymbol{\nu}$  and  $\mathbf{z}$  just once per iteration while keeping the others fixed. The resulting algorithm is presented in Algorithm 2.1.

It has been shown (see, e.g., [76, 78]) that while Algorithm 2.1 is the MM with an inexact minimization with respect to  $(\boldsymbol{\nu}, \mathbf{z})$ , it does in fact converge to the solution of (2.19) (and thus (2.18)) and is therefore distinct from the MM. The following theorem demonstrates the convergence of ADMM:

**Theorem 2.1.** (due to [78]) Consider the problem (2.18) where  $\mathbf{G}$  has full column rank and  $f, g$  are closed, proper, convex functions. Let  $\boldsymbol{\nu}_0 \in \mathbb{R}^n$ ,  $\mathbf{z}_0 \in \mathbb{R}^p$ ,  $\mathbf{u}_0 \in \mathbb{R}^d$  and  $\rho > 0$  be

arbitrary, and  $\{\mu_k \geq 0\}_{k=0}^{\infty}$ ,  $\{\eta_k \geq 0\}_{k=0}^{\infty}$  be two sequences such that

$$\sum_{k=0}^{\infty} \mu_k < \infty \quad \text{and} \quad \sum_{k=0}^{\infty} \eta_k < \infty.$$

Suppose the sequences  $\{\boldsymbol{\nu}_k\}_{k=0}^{\infty}$ ,  $\{\boldsymbol{z}_k\}_{k=0}^{\infty}$ , and  $\{\boldsymbol{u}_k\}_{k=0}^{\infty}$  satisfy

$$\begin{aligned} \|\boldsymbol{\nu}_{k+1} - \arg \min_{\boldsymbol{\nu}} f(\boldsymbol{\nu}) + \frac{\rho}{2} \|\boldsymbol{G}\boldsymbol{\nu} - \boldsymbol{B}\boldsymbol{z}_k + \boldsymbol{u}_k\|_2^2\| \leq \mu_k \\ \|\boldsymbol{z}_{k+1} - \arg \min_{\boldsymbol{z}} g(\boldsymbol{B}\boldsymbol{z}) + \frac{\rho}{2} \|\boldsymbol{G}\boldsymbol{\nu}_{k+1} - \boldsymbol{B}\boldsymbol{z} + \boldsymbol{u}_k\|_2^2\| \leq \mu_k \\ \boldsymbol{u}_{k+1} = \boldsymbol{u}_k + (\boldsymbol{G}\boldsymbol{\nu}_{k+1} - \boldsymbol{B}\boldsymbol{z}_{k+1}) \end{aligned}$$

Then if (2.18) has a solution, say  $\boldsymbol{\nu}^*$ , then  $\{\boldsymbol{\nu}_k\}_{k=0}^{\infty}$  converges to  $\boldsymbol{\nu}^*$ . If (2.18) does not have a solution, then at least one of the sequences  $\{\boldsymbol{\nu}_k\}_{k=0}^{\infty}$  or  $\{\boldsymbol{u}_k\}_{k=0}^{\infty}$  diverges.

The implications of Theorem 2.1 to ADMM are as follows: When Step 3 and 4 of ADMM are solved exactly, the generated sequences  $\{\boldsymbol{\nu}_k\}$ ,  $\{\boldsymbol{z}_k\}$ , and  $\{\boldsymbol{u}_k\}$  satisfy the conditions in Theorem 2.1 strictly with  $\mu_k = \eta_k = 0$  and ADMM therefore converges if  $\boldsymbol{G}$  has full rank. However, Step 3 and 4 of ADMM need not necessarily be solved exactly, convergence is still guaranteed as long as the sequences of errors are absolutely summable. Thus in situations where Step 3 or 4 can only be solved approximately, e.g., using iterative methods, convergence is not compromised.

### 2.5.2 Multi-modal Recovery via Augmented Lagrangian

We propose a simple algorithm based on the ADMM for solving the problems (2.16)-(2.18) which all admit the general form of:

$$\begin{aligned} \|\boldsymbol{y}_{xy} - \boldsymbol{A}_{xy} \boldsymbol{P}_{xy} \text{vec}(\boldsymbol{X})\|_2 &\leq \epsilon_{xy} \\ \hat{\boldsymbol{X}} = \arg \min_{\boldsymbol{X}} \varphi(\boldsymbol{H} \text{vec}(\boldsymbol{X})) \quad \text{subject to} \quad \|\boldsymbol{y}_{y\lambda} - \boldsymbol{A}_{y\lambda} \boldsymbol{P}_{y\lambda} \text{vec}(\boldsymbol{X})\|_2 &\leq \epsilon_{y\lambda} \\ \|\boldsymbol{y}_{x\lambda} - \boldsymbol{A}_{x\lambda} \boldsymbol{P}_{x\lambda} \text{vec}(\boldsymbol{X})\|_2 &\leq \epsilon_{x\lambda} \end{aligned} \tag{2.20}$$

with different convex objectives  $\varphi(\boldsymbol{H} \text{vec}(\boldsymbol{X})) : \mathbb{R}^{n_x n_y n_\lambda} \rightarrow \mathbb{R}$ , where  $\boldsymbol{P}_{xy}$  is simply the identity  $\boldsymbol{I}_{n_x n_y n_\lambda}$ . In short, by writing (2.20) as a sum of lower semi-continuous convex functions

in the form of (2.18) and applying ADMM with an appropriate choice for the matrices  $\mathbf{B}$  and  $\mathbf{G}$ , we reduce ADMM into a series of iterative updates that consist of computing the *proximity operator* of all functions and *averaging* their results in order to obtain the solution to (2.20). As a function of  $\boldsymbol{\delta}$ , the proximity operator of a function  $g$  with penalty  $\rho$  is defined as

$$\text{prox}_{g,\rho}(\boldsymbol{\delta}) = \arg \min_{\mathbf{z}} g(\mathbf{z}) + \frac{\rho}{2} \|\mathbf{z} - \boldsymbol{\delta}\|_2^2 \quad (2.21)$$

and for a variety of functions  $g$ , their corresponding proximity operators admit analytical solutions, or can be solved iteratively [79].

Following a similar approach introduced in [80], we first re-write (2.20) equivalently as an unconstrained problem in the form of

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \varphi(\mathbf{H} \text{vec}(\mathbf{X})) + i_{B_{xy}}(\text{vec}(\mathbf{X})) + i_{B_{y\lambda}}(\text{vec}(\mathbf{X})) + i_{B_{x\lambda}}(\text{vec}(\mathbf{X})) \quad (2.22)$$

where  $B_S = \{\boldsymbol{\nu} \in \mathbb{R}^{n_x n_y n_\lambda} : \|\mathbf{y}_S - \mathbf{A}_S \mathbf{P}_S \boldsymbol{\nu}\|_2 \leq \epsilon_S\}$  and  $i_{B_S}$  denotes the indicator function of the convex set  $B_S$  defined as

$$i_{B_S}(\boldsymbol{\nu}) = \begin{cases} 0 & \text{if } \boldsymbol{\nu} \in B_S \\ +\infty & \text{otherwise} \end{cases}.$$

Comparing (2.22) to (2.18) we create a correspondence using the mapping:

$$\begin{aligned} f(\boldsymbol{\nu}) &= 0 \\ \mathbf{G} &= [\mathbf{I}_{n_x n_y n_\lambda}^T \quad \mathbf{I}_{n_x n_y n_\lambda}^T \quad \mathbf{I}_{n_x n_y n_\lambda}^T \quad \mathbf{I}_{n_x n_y n_\lambda}^T \quad \mathbf{0}^T]^\top \quad (2.23) \\ g(\mathbf{G}\boldsymbol{\nu}) &= \varphi(\mathbf{H}\boldsymbol{\nu}) + i_{B_{xy}}(\boldsymbol{\nu}) + i_{B_{y\lambda}}(\boldsymbol{\nu}) + i_{B_{x\lambda}}(\boldsymbol{\nu}) \end{aligned}$$

where  $\boldsymbol{\nu} = \text{vec}(\mathbf{X})$ . Note that the  $\mathbf{0}$  matrix in  $\mathbf{G}$  is associated with the asymmetric constraint detailed below. With the mapping (2.23), we then re-write (2.22) into the form of (2.19) via

the change of variables:

$$\hat{\boldsymbol{\nu}} = \arg \min_{\boldsymbol{\nu}, \mathbf{z}} \underbrace{0}_{f(\boldsymbol{\nu})} + \underbrace{\varphi(\mathbf{z}^{(d)}) + i_{B_{xy}}(\mathbf{z}^{(2)}) + i_{B_{y\lambda}}(\mathbf{z}^{(3)}) + i_{B_{x\lambda}}(\mathbf{z}^{(4)})}_{g(\mathbf{B}\mathbf{z})}$$

subject to

$$\underbrace{\begin{bmatrix} \mathbf{I} & & & \\ & \mathbf{I} & & \\ & & \mathbf{I} & \\ & & & \mathbf{I} \\ -\mathbf{H} & & & \mathbf{I} \end{bmatrix}}_B \underbrace{\begin{bmatrix} \mathbf{z}^{(1)} \\ \mathbf{z}^{(2)} \\ \mathbf{z}^{(3)} \\ \mathbf{z}^{(4)} \\ \mathbf{z}^{(d)} \end{bmatrix}}_z = \underbrace{\begin{bmatrix} \mathbf{I}_{n_x n_y n_\lambda} \\ \mathbf{I}_{n_x n_y n_\lambda} \\ \mathbf{I}_{n_x n_y n_\lambda} \\ \mathbf{I}_{n_x n_y n_\lambda} \\ \mathbf{0} \end{bmatrix}}_G \boldsymbol{\nu} \quad (2.24)$$

Here instead of setting  $\mathbf{z}^{(1)} = \mathbf{H}\boldsymbol{\nu}$ , the asymmetric constraint  $\mathbf{z}^{(d)} = \mathbf{H}\mathbf{z}^{(1)}$ , although seemingly excessive and unnecessary, actually decouples the optimization (2.20) into a simple framework for incorporating a different number of modalities. We now simply apply the ADMM as given in Algorithm 2.1 to (2.24) with the Lagrange multipliers  $\mathbf{u} = [\mathbf{u}^{(1)\top} \ \mathbf{u}^{(2)\top} \ \mathbf{u}^{(3)\top} \ \mathbf{u}^{(4)\top} \ \mathbf{u}^{(d)\top}]^\top$ . To that end, notice with the specific mappings in (2.24), Step 3 of Algorithm 2.1 reduces to simply an average over the auxiliary variables  $\mathbf{z}$  and  $\mathbf{u}$ :

$$\begin{aligned} \boldsymbol{\nu}_{k+1} &= \arg \min_{\boldsymbol{\nu}} \underbrace{f(\boldsymbol{\nu})}_{0} + \frac{\rho}{2} \|\mathbf{G}\boldsymbol{\nu} - \mathbf{B}\mathbf{z}_k + \mathbf{u}_k\|_2^2 \\ &= (\mathbf{G}^\top \mathbf{G})^{-1} \mathbf{G}^\top (\mathbf{B}\mathbf{z}_k - \mathbf{u}_k) \\ \Rightarrow \boldsymbol{\nu}_{k+1} &= \frac{1}{4} \sum_{j=1}^4 (\mathbf{z}_k^{(j)} - \mathbf{u}_k^{(j)}) \end{aligned} \quad (2.25)$$

On the other hand, the minimization with respect to  $\mathbf{z}$  (Step 4 of Algorithm 2.1) is now decoupled into four separate sub-problems, each with respect to  $(\mathbf{z}^{(1)}, \mathbf{z}^{(d)})$ ,  $\mathbf{z}^{(2)}$ ,  $\mathbf{z}^{(3)}$ , and  $\mathbf{z}^{(4)}$ . Specifically, we have

$$\begin{aligned}
z_{k+1} &= \arg \min g(\mathbf{B}z) + \frac{\rho}{2} \|\mathbf{G}\boldsymbol{\nu}_{k+1} - \mathbf{B}z + \mathbf{u}_k\|_2^2 \\
\Rightarrow \begin{bmatrix} z_{k+1}^{(1)} \\ z_{k+1}^{(2)} \\ z_{k+1}^{(3)} \\ z_{k+1}^{(4)} \\ z_{k+1}^{(d)} \end{bmatrix} &= \arg \min_{z^{(1)}, \dots, z^{(d)}} g(\mathbf{B}z) + \frac{\rho}{2} \left\| \begin{bmatrix} \boldsymbol{\nu}_{k+1} \\ \boldsymbol{\nu}_{k+1} \\ \boldsymbol{\nu}_{k+1} \\ \boldsymbol{\nu}_{k+1} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} z^{(1)} \\ z^{(2)} \\ z^{(3)} \\ z^{(4)} \\ z^{(d)} - \mathbf{H}z^{(1)} \end{bmatrix} + \begin{bmatrix} \mathbf{u}_k^{(1)} \\ \mathbf{u}_k^{(2)} \\ \mathbf{u}_k^{(3)} \\ \mathbf{u}_k^{(4)} \\ \mathbf{u}_k^{(d)} \end{bmatrix} \right\|_2^2 \\
&= \arg \min_{z^{(1)}, \dots, z^{(d)}} g(\mathbf{B}z) + \frac{\rho}{2} \|\mathbf{H}z^{(1)} - z^{(d)} + \mathbf{u}_k^{(d)}\|_2^2 + \sum_{j=1}^4 \frac{\rho}{2} \|\boldsymbol{\nu}_{k+1} - z^{(j)} + \mathbf{u}_k^{(j)}\|_2^2
\end{aligned}$$

and noting  $g(\mathbf{B}z)$  is a sum of functions of individual  $\{z^{(j)}\}_{j=2,3,4,d}$  we have:

$$\begin{aligned}
(z_{k+1}^{(1)}, z_{k+1}^{(d)}) &= \arg \min_{z^{(1)}, z^{(5)}} g_d(z^{(d)}) + \frac{\rho}{2} \|z^{(1)} - (\mathbf{u}_k^{(1)} + \boldsymbol{\nu}_{k+1})\|_2^2 + \frac{\rho}{2} \|z^{(d)} - (\mathbf{u}_k^{(d)} + \mathbf{H}z^{(1)})\|_2^2 \\
z_{k+1}^{(j)} &= \arg \min_{z^{(j)}} g_j(z^{(j)}) + \frac{\rho}{2} \|z^{(j)} - (\mathbf{u}_k^{(j)} + \boldsymbol{\nu}_{k+1})\|_2^2, \quad \forall j = \{2, 3, 4\}
\end{aligned} \tag{2.26}$$

where  $g_d = \varphi$ ,  $g_2 = i_{B_{xy}}$ ,  $g_3 = i_{B_{y\lambda}}$ , and  $g_4 = i_{B_{x\lambda}}$ .

In light of Theorem 2.1, the joint minimization with respect to  $(z^{(1)}, z^{(d)})$  (and more generally each minimizations in (2.26)) need not be solved exactly. Taking advantage of this, we simply cycle over the minimization with respect to  $z^{(1)}$  and  $z^{(d)}$  once at each iteration (i.e., a single Gauss-Seidel pass). In doing so, each individual sub-problem in (2.26) now all simply takes the form of the proximity operator (2.21) of the corresponding function:

$$\begin{aligned}
z_{k+1}^{(j)} &= \arg \min_{z^{(j)}} g_j(z^{(j)}) + \frac{\rho}{2} \|z^{(j)} - \boldsymbol{\delta}\|_2^2 \\
&= \text{prox}_{g_j, \rho}(\boldsymbol{\delta}), \quad \forall j = \{1, 2, 3, 4, d\}
\end{aligned} \tag{2.27}$$

with  $g_1(z) = \frac{\rho}{2} \|z_k^{(d)} - (\mathbf{u}_k^{(d)} + \mathbf{H}z)\|_2^2$  and  $\{g_j\}_{j=2,3,4,d}$  as defined in (2.26).

In particular, the proximity operator of the indicator functions  $\text{prox}_{i_{B_S}, \rho}(\boldsymbol{\delta})$  is simply the orthogonal projection of  $\boldsymbol{\delta}$  onto the ellipse  $B_S$  (recall  $B_S = \{\boldsymbol{\nu} \in \mathbb{R}^{n_x n_y n_\lambda} : \|\mathbf{y}_S - \mathbf{A}_S \mathbf{P}_S \boldsymbol{\nu}\|_2 \leq \epsilon_S\}$ )

and can be shown<sup>2</sup> to have an explicit expression:

$$\text{prox}_{i_{B_S}, \rho}(\boldsymbol{\delta}) = \boldsymbol{\delta} + \begin{cases} 0 & \text{if } \|\mathbf{y}_S - \mathbf{A}_S \mathbf{P}_S \boldsymbol{\delta}\|_2 \leq \epsilon_S \\ \alpha (\mathbf{A}_S \mathbf{P}_S)^* (\mathbf{A}_S \mathbf{A}_S^*)^{-1} (\mathbf{y}_S - \mathbf{A}_S \mathbf{P}_S \boldsymbol{\delta}) & \text{if } \|\mathbf{y}_S - \mathbf{A}_S \mathbf{P}_S \boldsymbol{\delta}\|_2 > \epsilon_S \end{cases}$$

where  $\alpha = 1 - \epsilon_S \|\mathbf{y}_S - \mathbf{A}_S \mathbf{P}_S \boldsymbol{\delta}\|_2^{-1}$  is a scalar constant depending on  $\boldsymbol{\delta}$ . Also, a direct calculation reveals the proximity operator for  $g_1$  has the closed-form

$$\text{prox}_{g_1, \rho}(\boldsymbol{\delta}) = (\mathbf{H}^* \mathbf{H})^{-1} \left( \mathbf{H}^* \left( \mathbf{z}_k^{(d)} - \mathbf{u}_k^{(d)} \right) + \boldsymbol{\delta} \right)$$

and as we will see, for each of the objectives in (2.15)-(2.17), the resulting proximity operators all have analytical expressions.

Since the  $(\mathbf{z}^{(1)}, \mathbf{z}^{(d)}), \mathbf{z}^{(2)}, \mathbf{z}^{(3)}, \mathbf{z}^{(4)}$  (and their corresponding  $\mathbf{u}$ ) updates in (2.26) are independent of one and the other, they can be carried out in parallel and the resulting algorithm therefore follows a *decomposition-coordination* procedure: the original problem (2.20) is decomposed into smaller sub-problems (2.27) in which the solutions can be found in parallel and are coordinated via (2.25) to find the solution to the original problem. The resulting algorithmic framework dubbed Multi-modal Recovery via Augmented Lagrangian (MuRecAL), is shown in Algorithm 2.2.

Note that under the MuRecAL framework, one can easily check that the uni-modal and bi-modal setups can be implemented by simply omitting the updates for the indicator function(s) not in effect and applying the averaging step (Step 3 of Algorithm 2.2) only to the participating auxiliary variables.

---

<sup>2</sup>Observe that if  $\boldsymbol{\delta} \in B_S$  then the minimizer to (2.21) with  $g = i_{B_S}$  is  $\mathbf{z} = \boldsymbol{\delta}$ . On the other hand, if  $\boldsymbol{\delta} \notin B_S$ , then the minimizer corresponds to the direct solution of

$$\hat{\mathbf{z}} = \arg \min_{\mathbf{z}} \|\mathbf{z} - \boldsymbol{\delta}\|_2^2 \quad \text{subject to} \quad \|\mathbf{y}_S - \mathbf{A}_S \mathbf{P}_S \mathbf{z}\|_2 = \epsilon_S.$$

**Algorithm 2.2** Multi-modal Recovery via Augmented Lagrangian (MuRecAL)

---

1: Set  $k = 0$ ,  $\mathbf{u}_k^{(1)}, \dots, \mathbf{u}_k^{(d)}, \mathbf{z}_k^{(1)}, \dots, \mathbf{z}_k^{(4)}$ , and choose  $\rho > 0$   
2: **repeat**  
3:      $\boldsymbol{\nu}_{k+1} = \frac{1}{4} \sum_{j=1}^4 (\mathbf{z}_k^{(j)} - \mathbf{u}_k^{(j)})$   
4:      $\mathbf{z}_{k+1}^{(1)} = \text{prox}_{g_1, \rho}(\mathbf{u}_k^{(1)} + \boldsymbol{\nu}_{k+1})$   
5:      $\mathbf{z}_{k+1}^{(d)} = \text{prox}_{\varphi, \rho}(\mathbf{u}_k^{(d)} + \mathbf{H}\mathbf{z}_{k+1}^{(1)})$   
6:      $\mathbf{u}_{k+1}^{(1)} = (\mathbf{u}_k^{(1)} + \boldsymbol{\nu}_{k+1}) - \mathbf{z}_{k+1}^{(1)}$   
7:      $\mathbf{u}_{k+1}^{(d)} = (\mathbf{u}_k^{(d)} + \mathbf{H}\mathbf{z}_{k+1}^{(1)}) - \mathbf{z}_{k+1}^{(d)}$   
8:      $\mathbf{z}_{k+1}^{(2)} = \text{prox}_{i_{B_{xy}}, \rho}(\mathbf{u}_k^{(2)} + \boldsymbol{\nu}_{k+1})$   
9:      $\mathbf{u}_{k+1}^{(2)} = (\mathbf{u}_k^{(2)} + \boldsymbol{\nu}_{k+1}) - \mathbf{z}_{k+1}^{(2)}$   
10:     $\mathbf{z}_{k+1}^{(3)} = \text{prox}_{i_{B_{y\lambda}}, \rho}(\mathbf{u}_k^{(3)} + \boldsymbol{\nu}_{k+1})$   
11:     $\mathbf{u}_{k+1}^{(3)} = (\mathbf{u}_k^{(3)} + \boldsymbol{\nu}_{k+1}) - \mathbf{z}_{k+1}^{(3)}$   
12:     $\mathbf{z}_{k+1}^{(4)} = \text{prox}_{i_{B_{x\lambda}}, \rho}(\mathbf{u}_k^{(4)} + \boldsymbol{\nu}_{k+1})$   
13:     $\mathbf{u}_{k+1}^{(4)} = (\mathbf{u}_k^{(4)} + \boldsymbol{\nu}_{k+1}) - \mathbf{z}_{k+1}^{(4)}$   
14:     $k = k + 1$   
15: **until**  $\|\mathbf{G}\boldsymbol{\nu}_{k+1} - \mathbf{B}\mathbf{z}_{k+1}\|_2 \leq \mu_{\text{thresh}}$

---

**2.5.2.1 2D-TV Minimization**

To solve (2.15), we first note that the TV can be equivalently re-written as (see Section 3.1 for more details):

$$\sum_{i=1}^{n_\lambda} \|\mathbf{x}_i\|_{TV} = \left\| \underbrace{\begin{bmatrix} \mathbf{I}_{n_\lambda} \otimes \mathbf{D}_x \\ \mathbf{I}_{n_\lambda} \otimes \mathbf{D}_y \end{bmatrix}}_{\mathbf{D}_{xy}} \text{vec}(\mathbf{X}) \right\|_1.$$

where  $\mathbf{D}_x, \mathbf{D}_y \in \mathbb{R}^{n_x n_y \times n_x n_y}$  are the first-order horizontal and vertical difference matrices respectively with defined boundary conditions. Therefore, we set  $\varphi = \|\cdot\|_1$  and  $\mathbf{H} = \mathbf{D}_{xy}$  and apply MuRecAL to give

$$\begin{aligned} \text{prox}_{g_1, \rho}(\boldsymbol{\delta}) &= (\mathbf{D}_{xy}^\top \mathbf{D}_{xy} + \mathbf{I})^{-1} \left( \mathbf{D}_{xy}^\top (\mathbf{z}_k^{(d)} - \mathbf{u}_k^{(d)}) + \boldsymbol{\delta} \right) \\ \text{prox}_{\varphi, \rho}(\boldsymbol{\delta}) &= \arg \min_{\mathbf{z}} \|\mathbf{z}\|_1 + \frac{\rho}{2} \|\mathbf{z} - \boldsymbol{\delta}\|_2^2 \end{aligned}$$

Standard calculations show that  $\forall i \in \{1, \dots, n_x n_y n_\lambda\}$

$$(\text{prox}_{g_1, \rho}(\boldsymbol{\delta}))_i = \begin{cases} \boldsymbol{\delta}_i - \frac{1}{\rho}, & \boldsymbol{\delta}_i > \frac{1}{\rho} \\ 0, & \boldsymbol{\delta}_i \in [-\frac{1}{\rho}, \frac{1}{\rho}] \\ \boldsymbol{\delta}_i + \frac{1}{\rho}, & \boldsymbol{\delta}_i < -\frac{1}{\rho} \end{cases} \quad (2.28)$$

which is the *soft-thresholding* operator [81] applied to each element of  $\boldsymbol{\delta}$ .

Regarding the computation of  $\text{prox}_{g_1, \rho}(\boldsymbol{\delta})$ , note that under periodic boundary conditions, the first-order difference matrices  $\mathbf{D}_x$  and  $\mathbf{D}_y$  in  $\mathbf{D}_{xy}$  are block-circulant with circulant blocks (each representing a periodic convolution). Since  $(\mathbf{I}_{n_\lambda} \otimes \mathbf{D}_x)$  and  $(\mathbf{I}_{n_\lambda} \otimes \mathbf{D}_y)$  are themselves block-circulant, their triple block-circulant structure can therefore be diagonalised by the 3D Discrete Fourier Transform (DFT) [82] as  $\tilde{\mathbf{F}}^* \tilde{\mathbf{D}}_x \tilde{\mathbf{F}}$  and  $\tilde{\mathbf{F}}^* \tilde{\mathbf{D}}_y \tilde{\mathbf{F}}$  respectively with  $\tilde{\mathbf{F}} = (\mathbf{F} \otimes \mathbf{F} \otimes \mathbf{F})$  where  $\mathbf{F}$  is the 1D discrete Fourier matrix and  $\tilde{\mathbf{D}}_x$ ,  $\tilde{\mathbf{D}}_y$  are diagonal. Thus the inverse in  $\text{prox}_{g_1, \rho}(\boldsymbol{\delta})$  can be expressed as

$$(\mathbf{D}_{xy}^\top \mathbf{D}_{xy} + \mathbf{I}_{n_x n_y n_\lambda})^{-1} = \tilde{\mathbf{F}}^* \left( |\tilde{\mathbf{D}}_x|^2 + |\tilde{\mathbf{D}}_y|^2 + \mathbf{I}_{n_x n_y n_\lambda} \right)^{-1} \tilde{\mathbf{F}}$$

where  $|\cdot|^2$  denotes the squared absolute value performed element-wise to the argument. Noting that the multiplications by  $\tilde{\mathbf{F}}$  and its conjugate transpose correspond to performing the forward and inverse 3D DFT respectively,  $\text{prox}_{g_1, \rho}(\boldsymbol{\delta})$  can therefore be computed efficiently using the 3D Fast Fourier Transform (FFT) algorithm

$$\text{prox}_{g_1, \rho}(\boldsymbol{\delta}) = \text{IFFT3} \left( \frac{\text{FFT3} \left( \mathbf{D}_{xy}^\top \left( \mathbf{z}_k^{(d)} - \mathbf{u}_k^{(d)} \right) + \boldsymbol{\delta} \right)}{\text{diag} \left( |\tilde{\mathbf{D}}_x|^2 + |\tilde{\mathbf{D}}_y|^2 + \mathbf{I}_{n_x n_y n_\lambda} \right)} \right) \quad (2.29)$$

where  $\text{FFT3}(\cdot)$  and  $\text{IFFT3}(\cdot)$  denote the forward and inverse 3D-FFT,  $\text{diag}(\cdot)$  outputs the main diagonal and the division is carried out element-wise. We note the slight abuse of notation in that the 3D-FFT and its inverse are actually performed to the unvectorized argument while the division is performed on the vectorized arguments.

### 2.5.2.2 3D-Wavelet Minimization

By making the correspondence  $\varphi = \|\cdot\|_1$ ,  $\mathbf{H} = \Psi_{3D}^\top$ , we apply MuRecAL to (2.16) to obtain

$$\begin{aligned}\text{prox}_{g_1,\rho}(\boldsymbol{\delta}) &= (\Psi_{3D}\Psi_{3D}^\top + \mathbf{I})^{-1} \left( \Psi_{3D} \left( \mathbf{z}_k^{(d)} - \mathbf{u}_k^{(d)} \right) + \boldsymbol{\delta} \right) \\ \text{prox}_{\varphi,\rho}(\boldsymbol{\delta}) &= \arg \min_{\mathbf{z}} \|\mathbf{z}\|_1 + \frac{\rho}{2} \|\mathbf{z} - \boldsymbol{\delta}\|_2^2\end{aligned}$$

where the proximity operator for  $\varphi$  can be computed as before using (2.28).

When constructed using orthonormal 1D and 2D wavelets (as in [68, 74]),  $\Psi_{3D}$  is also orthonormal:  $\Psi_{3D}\Psi_{3D}^\top = (\Psi_{1D} \otimes \Psi_{2D})(\Psi_{1D} \otimes \Psi_{2D})^\top = (\Psi_{1D}\Psi_{1D}^\top \otimes \Psi_{2D}\Psi_{2D}^\top) = \mathbf{I}$  and so the matrix inversion in  $\text{prox}_{g_1,\rho}(\boldsymbol{\delta})$  reduces to simply a scalar multiplication.

### 2.5.2.3 Joint TV-Nuclear Norm Minimization

Since the objective in (2.17) consists of two terms, namely, the 2D-TV norm  $\|\mathbf{D}_{xy} \text{vec}(\mathbf{X})\|_1$  and the nuclear norm  $\|\mathbf{X}\|_*$ , we let  $\varphi = \|\cdot\|_1$ ,  $\mathbf{H} = \mathbf{D}_{xy}$ , and introduce a new set of auxiliary variables  $\mathbf{z}^{(5)}$  and  $\mathbf{u}^{(5)}$  for the nuclear norm to re-write (2.27) as

$$\begin{aligned}\hat{\boldsymbol{\nu}} &= \arg \min_{\boldsymbol{\nu}, \mathbf{z}} \varphi(\mathbf{z}^{(d)}) + \gamma \|\text{vec}^{-1}(\mathbf{z}^{(5)})\|_* + i_{B_{xy}}(\mathbf{z}^{(2)}) + i_{B_{y\lambda}}(\mathbf{z}^{(3)}) + i_{B_{x\lambda}}(\mathbf{z}^{(4)}) \\ \text{subject to } &\begin{bmatrix} \mathbf{I} & & & & \\ & \mathbf{I} & & & \\ & & \mathbf{I} & & \\ & & & \mathbf{I} & \\ -\mathbf{H} & & & & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{z}^{(1)} \\ \mathbf{z}^{(2)} \\ \mathbf{z}^{(3)} \\ \mathbf{z}^{(4)} \\ \mathbf{z}^{(5)} \\ \mathbf{z}^{(d)} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{n_x n_y n_\lambda} \\ \mathbf{0} \end{bmatrix} \text{vec}(\mathbf{X})\end{aligned}$$

which is simply (2.24) with an additional convex objective and constraint. It is thus easy to show, by applying ADMM given in Algorithm 2.1 as before, that the resulting algorithm is equivalent to MuRecAL for the 2D-TV minimization but with the inclusion of the new

auxiliary variables in the averaging (Step 2 of Algorithm 2.2):

$$\boldsymbol{\nu}_{k+1} = \frac{1}{5} \sum_{j=1}^5 \left( \mathbf{z}_k^{(j)} - \mathbf{u}_k^{(j)} \right)$$

and the additional set of updates to be carried out in parallel:

$$\begin{aligned} \mathbf{z}_{k+1}^{(5)} &= \text{prox}_{\|\cdot\|_*, \frac{\rho}{\gamma}} \left( \text{vec}^{-1} \left( \mathbf{u}_k^{(5)} + \boldsymbol{\nu}_{k+1} \right) \right) \\ \mathbf{u}_{k+1}^{(5)} &= \left( \mathbf{u}_k^{(5)} + \boldsymbol{\nu}_{k+1} \right) - \mathbf{z}_{k+1}^{(5)} \end{aligned}$$

Regarding  $\text{prox}_{\|\cdot\|_*, \frac{\rho}{\gamma}}(\boldsymbol{\delta})$ , if  $\boldsymbol{\delta} = \mathbf{U} \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r) \mathbf{V}^*$  is the economy-sized singular value decomposition of the matrix  $\boldsymbol{\delta}$ , then the proximity operator for the nuclear norm given above is

$$\text{prox}_{\|\cdot\|_*, \frac{\rho}{\gamma}}(\boldsymbol{\delta}) = \text{vec}(\mathbf{U} \text{diag}(\tilde{\sigma}_1, \tilde{\sigma}_2, \dots, \tilde{\sigma}_r) \mathbf{V}^*)$$

where

$$\tilde{\sigma}_i = \begin{cases} \sigma_i - \frac{\gamma}{\rho}, & \sigma_i > \frac{\gamma}{\rho}, \\ 0, & \sigma_i \leq \frac{\gamma}{\rho} \end{cases} \quad \forall i \in \{1, 2, \dots, r\}$$

that is, the soft-thresholding operator applied to the singular values of  $\boldsymbol{\delta}$  [83].

## 2.6 Simulations and Performance Analysis

In this section, we evaluate and compare the performance of the proposed multi-modal acquisition framework to existing uni-modal setups using the various compressive hyperspectral image recovery approaches described in Section 3.1. The goal is to analyse the influence of distributing the measurements across the three independent planes ( $x-y$ ,  $y-\lambda$ , and  $x-\lambda$ ) and jointly reconstruct the hyperspectral image from the three sets of measurements. To that end, for a fixed number of total measurements, we solve (2.15)-(2.17) with measurements (2.14) varied as

- (a)  $\mathbf{A}_{y\lambda}$ ,  $\mathbf{A}_{x\lambda} = \mathbf{0}$  (one mode, using (2.11))
- (b)  $\mathbf{A}_{x\lambda} = \mathbf{0}$  (two modes, using (2.11) and (2.12)), and
- (c)  $\mathbf{A}_{xy}$ ,  $\mathbf{A}_{y\lambda}$ ,  $\mathbf{A}_{x\lambda} \neq \mathbf{0}$  (three modes, using (2.11), (2.12), and (2.13))

The sensing matrices  $\Phi_{xy}$ ,  $\Phi_{y\lambda}$ , and  $\Phi_{x\lambda}$  are constructed using randomly sub-sampled 2D discrete Fourier orthogonal basis (see e.g., [84]) for generating the compressive measurements (2.12)-(2.13). The reasons behind this choice are purely computational: they admit a fast transform and considerably accelerate the recovery process by reducing the computation of the orthogonal projections in Steps 8, 10, and 12 of MuRecAL since the inversion  $(\mathbf{A}_S \mathbf{A}_S^*)^{-1}$  reduces to the identity (see Section 2.5.2 for more details).

The MuRecAL algorithm for all cases was initialized with all variables set to 0 and for all cases the algorithm is terminated when  $\|\mathbf{G}\boldsymbol{\nu}_{k+1} - \mathbf{B}\mathbf{z}_{k+1}\|_2$  falls below  $\mu_{\text{thresh}} = 10^{-4}$ . All experiments were performed under MATLAB v7.13 (R2011b) running on a Desktop PC with an Intel Core2 Quad CPU at 2.6 GHz and 4 GB of RAM.

### 2.6.1 Experimental Results with Synthetic Hyperspectral Data

In this part, the intention is to see to what extent can the multimodality affect the quality of the reconstruction using contrived hyperspectral data that is both smooth and of low rank, i.e., the two influential priors for HSI. To do so, we randomly selected four minerals from the ASTER Spectral Library [54] whose spectral signatures are shown in Figure 2.7(a). For convenience, we selected 128 bands (i.e.,  $n_\lambda = 128$ ) in the range of 0.4 to 2.3  $\mu\text{m}$  resulting in a hyperspectral cube to simplify the construction of the measurement matrices for all three dimensions. These signatures are then superimposed onto an  $128 \times 128$ -pixel head phantom image with each mineral occupying one of the four distinct regions as depicted in Figure 2.7(b). The simulated data  $\mathbf{X}$  is therefore spatially smooth in each column (due to the phantom image) with highly redundant rows (i.e.,  $\text{rank}(\mathbf{X}) = 4$  due to the four endmembers). With such a construction in mind, we evaluate the performance by using the channel-by-channel 2D-TV minimization (2.15) and the joint nuclear-TV norm minimization (2.17) schemes at different under-sampling ratios ( $m/n_x n_y n_\lambda$ ) for single, two, and three modes of measurements using:

- (a) a unique sensing matrix for each partition in each mode as in (2.9), and (b) different

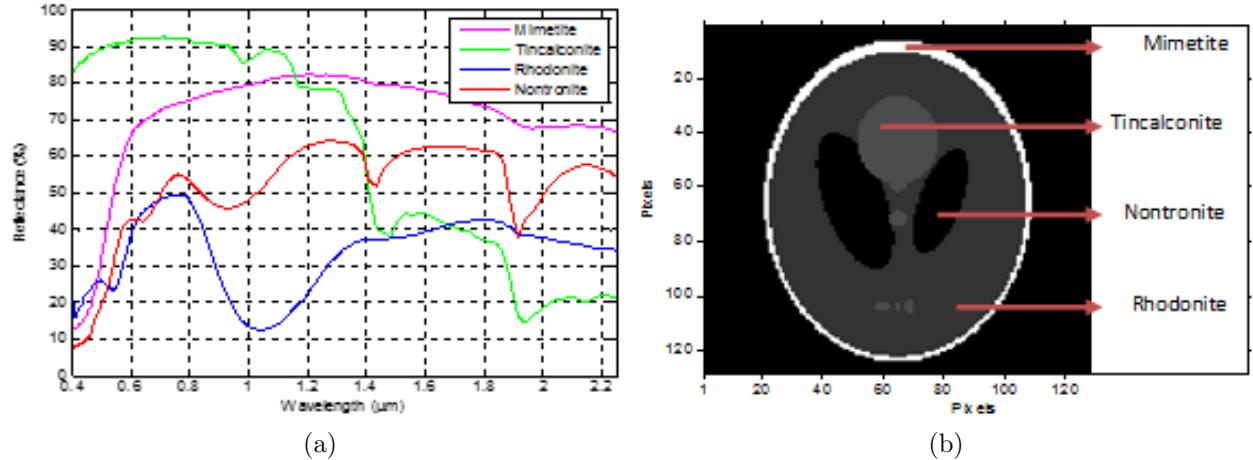


Figure 2.7: Simulated hyperspectral image. (a) The spectral signatures of the four chosen materials; (b) The head phantom image used along the spatial direction with the four distinct regions indicated.

sensing matrices for each partition in each mode as in (2.10). The compressive measurements obtained were also contaminated by white Gaussian noise with a standard deviation  $\sigma$  corresponding to a Signal-to-Noise Ratio (SNR) ( $= 10 \log_{10}(\|\mathbf{A}_S \text{vec}(\mathbf{X})\|_2^2 / \|\boldsymbol{\eta}_S\|_2^2)$ ) of 30 dB. The Signal-to-Reconstruction Error Ratio (SRE) was used to measure the quality of the HSI reconstruction:

$$\text{SRE} \stackrel{\text{def}}{=} 10 \log_{10} \left( \frac{\|\text{vec}(\mathbf{X})\|_2^2}{\|\text{vec}(\mathbf{X} - \hat{\mathbf{X}})\|_2^2} \right) \text{ dB.}$$

Figure 2.8 and Figure 2.10 plots the performance evaluation results, averaged over 5 Monte-Carlo runs, for different sampling rates, sampling modes, and sampling mechanism (i.e., unique sensing matrix vs. different sensing matrices) using 2D-TV and joint nuclear-TV minimization respectively. Figure 2.9 and Figure 2.11 give visual accounts of the reconstructed HSI (of a particular band) from the two methods at a particular under-sampling rate. Note that in order for the total number of measurements to remain constant for each modality at each of the under-sampling ratios, the number of measurements is allocated equally between the modes considered. The regularization parameter  $\gamma$  in (2.17) was set to  $5 \times 10^{-3}$  based on preliminary simulation giving the optimal SRE for the considered case.

One unusual observation seen in Figure 2.8 is that the SREs for all three modalities seem to saturate after an under-sampling ratio of approximately 0.1. This is due to the early

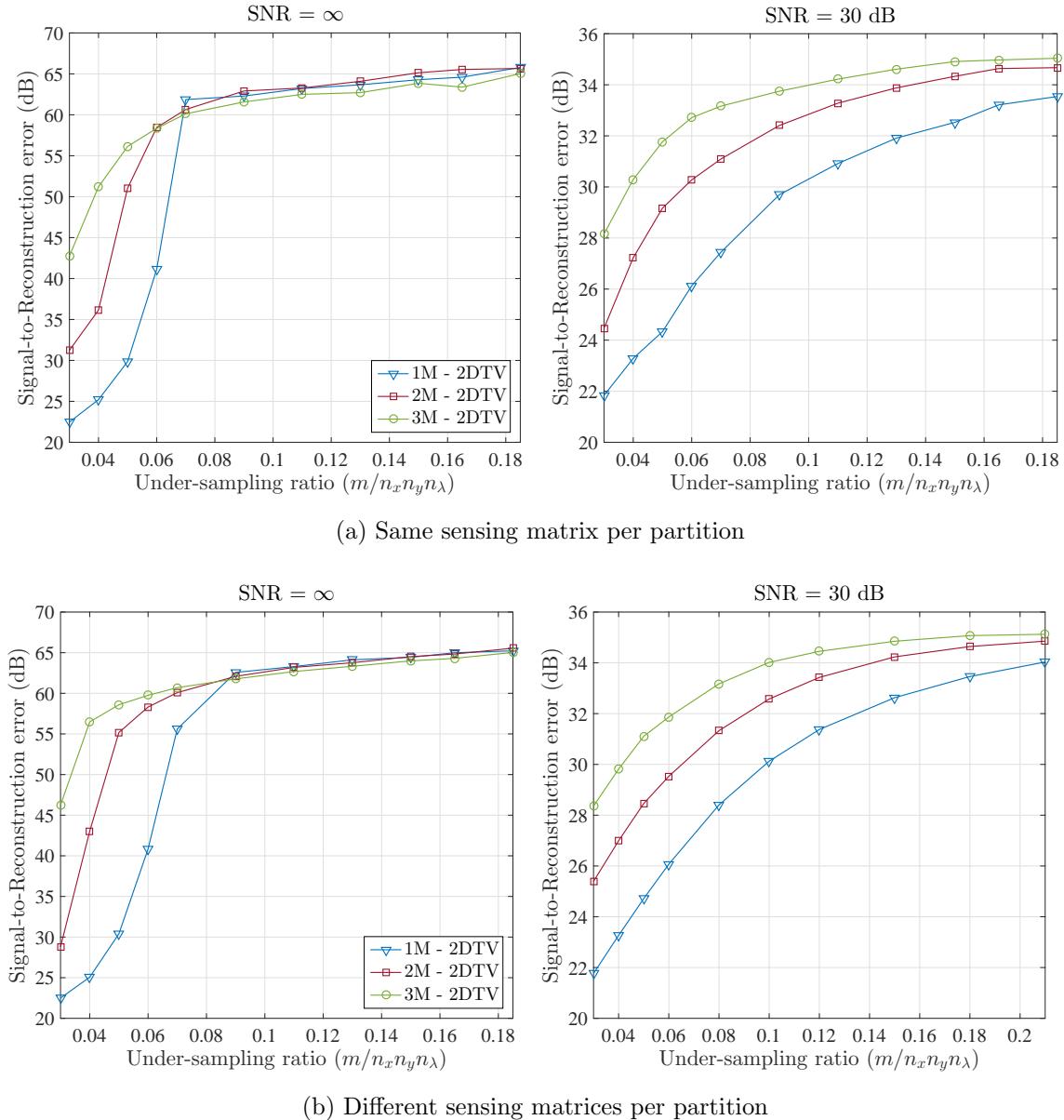


Figure 2.8: Synthetic HSI reconstruction performance from 2D-TV minimization at different sampling rates, SNRs, and sensing matrices with measurements taken from a single mode (1M), two modes (2M), and three modes (3M).

termination criterion used and all solutions were visually indistinguishable from the ground truth after this point which is not surprising considering the extremely smooth nature of the image resulting in exact reconstruction according to the CS theory.

In almost all cases in Figure 2.8, the channel-by-channel recovery benefited significantly from the multimodal measurements and, in all setups, outperforms its uni-modal counterparts. In particular, distributing the measurements across all three planes resulted in the best performance. Specifically, at severely under-sampled rates ( $m/n_x n_y n_\lambda < 0.07$ ) the multimodal

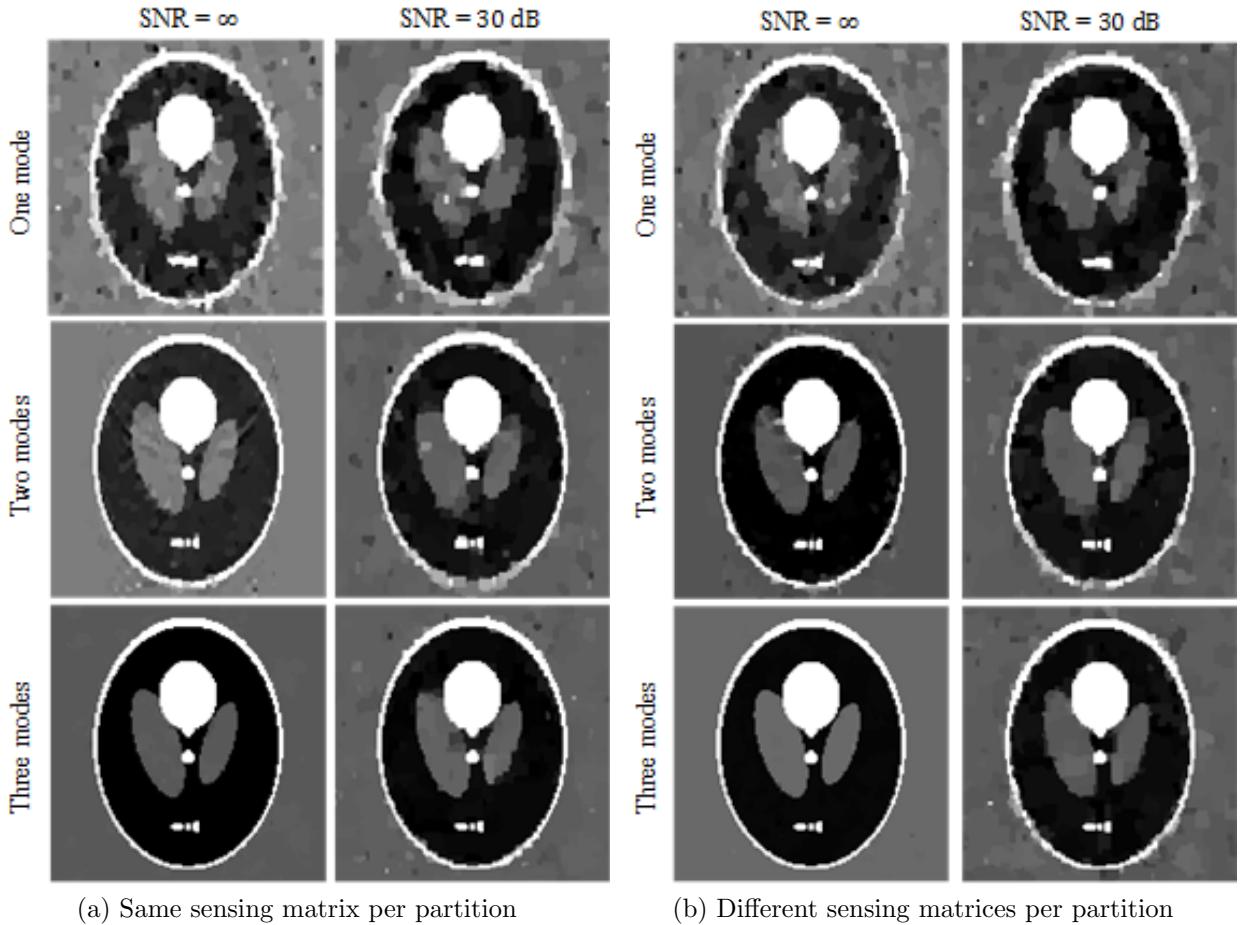


Figure 2.9: Synthetic HSI (at spectral band 65) reconstructed from 2D-TV minimization from various modes of measurements for an under-sampling ratio of 0.03.

measurements resulted in significant enhancement in SRE by 6 to 28 dB compared to the uni-modal measurements (see also Figure 2.9).

Comparing the two different sampling mechanisms (i.e., same matrix vs. different sensing matrices per partition), with one mode of measurements the reconstruction quality of the two are very similar for the channel-by-channel recovery. This should not come as a surprise since the measurements from each partition are used independently in the reconstruction and thus there is no particular benefit in using different sensing matrices because such information is only used within the corresponding partition's recovery. On the other hand, with two and three modes of measurements, by using different sensing matrices per partition has led to a slight overall improvement in SREs due to the information diversity is now shared across each partition during the reconstruction process.

With regard to the joint nuclear-TV recovery shown in Figure 2.10, interestingly using the

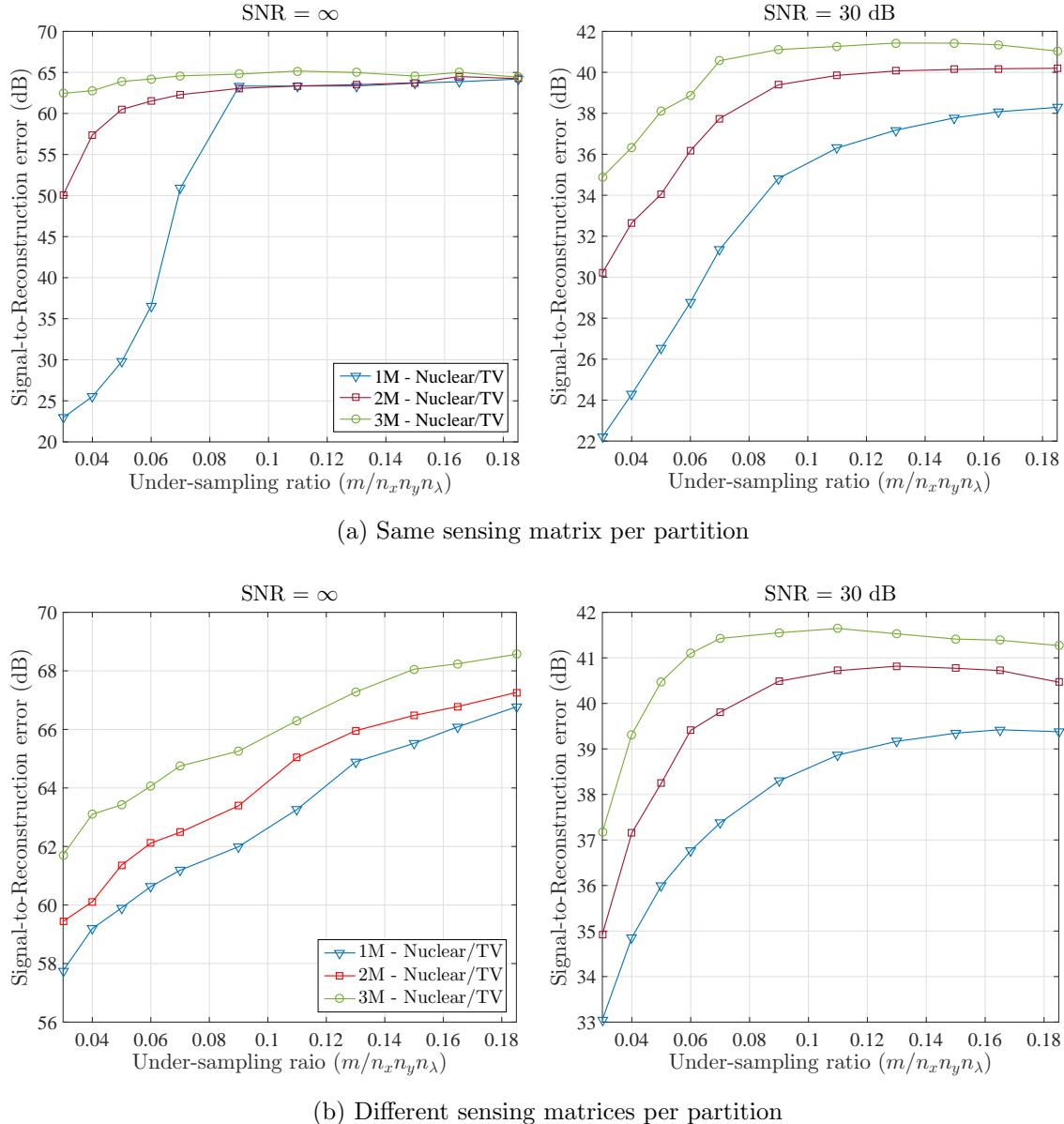


Figure 2.10: Synthetic HSI reconstruction performance from joint Nuclear/TV minimization at different sampling rates, SNRs, and sensing matrices with measurements taken from a single mode (1M), two modes (2M), and three modes (3M).

same sensing matrix per partition with one mode of measurement achieved no better performance than the single mode 2D-TV recovery. Such obscurity is consistent with our discussions in earlier sections: using the same sensing matrix per partition leads to highly redundant measurements and do not carry sufficient information for exploiting the sparsity structure beyond the measured dimensions (i.e., the low-rank structure). By using different matrices per partition instead, we observe in Figure 2.10(b) that the low-rank prior is successfully exploited leading to significant improvement in the SRE owing to the information

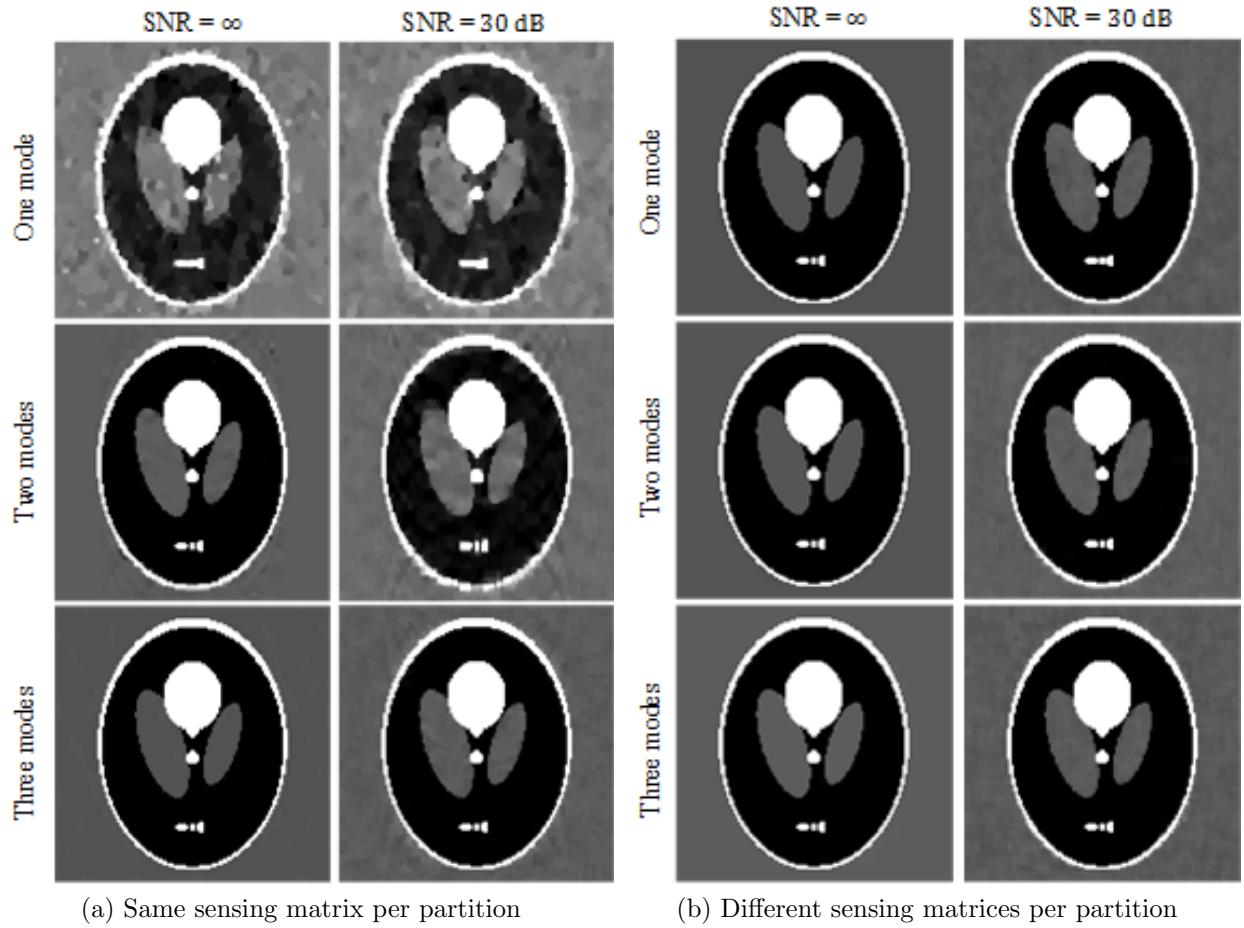


Figure 2.11: Synthetic HSI (at spectral band 65) reconstructed from joint Nuclear/TV minimization from various modes of measurements for an under-sampling ratio of 0.03.

diversity gained.

As with the channel-by-channel recovery, using multimodal measurements has led to appreciable improvements in the joint nuclear-TV reconstruction for all cases by at least 3 dB when using three modes of measurements. Remarkably, the multimodal measurements made under the same sensing matrix sampling scheme significantly enhanced the performance over the single mode measurements (e.g., when the number of CS measurements is approximately 3% of the data size) and becomes comparable to the multimodal measurements made under the different matrices per partition scheme (see Figure 2.11). This suggests that the multimodality is particularly favourable in situations where acquisition hardware can only emulate a unique sensing matrix or benefits from doing so (e.g., due to physical, time, or cost constraints).

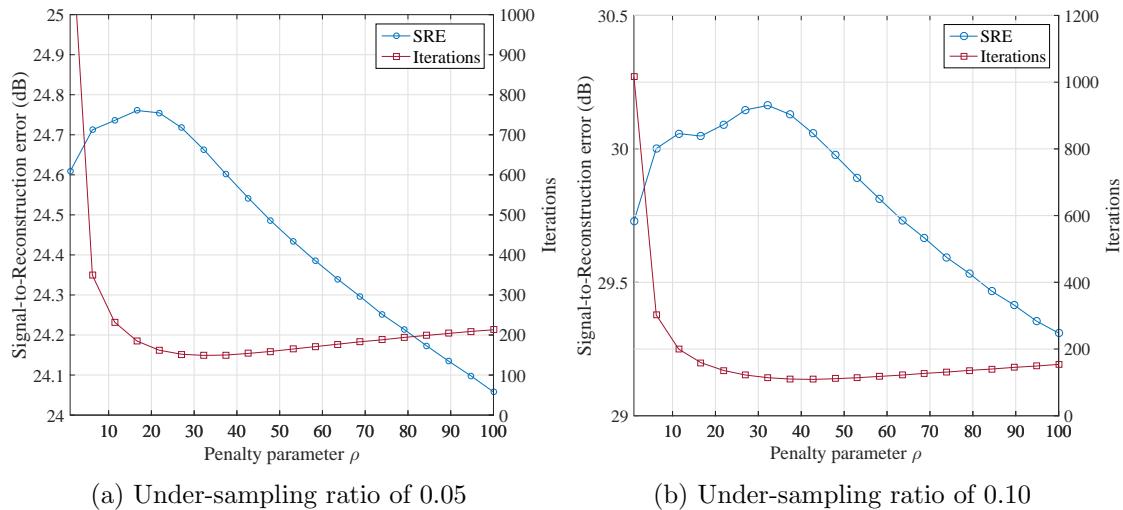


Figure 2.12: Trade-off between the SRE and number of iterations before convergence for different values of  $\rho$ . Results shown are obtained via 2D-TV minimization from single mode measurements with a SNR of 30 dB using synthetic HSI.

### 2.6.1.1 Choice of Penalty Parameter $\rho$

The value of the quadratic penalty parameter  $\rho$  in ADMM was found to have little or no effect on the reconstruction quality and only affects the speed of convergence provided it was within the range  $30 \leq \rho \leq 500$  (which coincides with that reported in [71]) depending on the noise level and under-sampling ratio. An example is given in Figure 2.12. Beyond this range, the reconstruction quality degrades and takes more iterations to converge. This becomes increasingly apparent as the under-sampling ratio drops. Although there are heuristics for the selection and update of  $\rho$  (see, e.g., [76, 77]), we simply select  $\rho$  by the speed of convergence which usually corresponds to an adequate performance.

### 2.6.1.2 Choice of the Noise Parameter $\epsilon$

In our experiments, the elements of the noise matrices  $\boldsymbol{\eta}_S$  in (2.11)-(2.13) were zero-mean Gaussian with standard deviation  $\sigma$  (corresponding to a SNR of 30 dB). While it is tempting to simply set  $\epsilon_S$  in the inequality constraints of the optimization problems (2.15)-(2.17) to be  $\|\boldsymbol{\eta}_S\|_2$  as the ideal choice (so that the probability of  $\|\boldsymbol{\eta}_S\|_2 \leq \epsilon_S$  is exactly one), choosing  $\epsilon$  this way is unfortunately infeasible in practice since the magnitude of the noise needs to be known *a priori*. On the other hand, since  $\|\boldsymbol{\eta}_S\|_2^2$  is distributed as  $\chi$ -squared, therefore

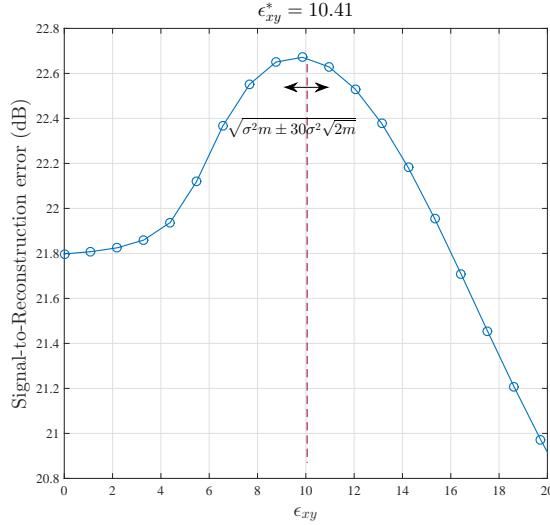


Figure 2.13: Reconstruction SRE vs.  $\epsilon_{xy}$ . Results obtained via 2D-TV minimization from single mode measurements at an under-sampling ratio of 0.05 and SNR of 30 dB.

owing to the concentration inequalities, the probability that  $\|\boldsymbol{\eta}_S\|_2^2$  exceeds its mean (i.e.,  $\sigma^2 m$ ) plus two standard deviations (i.e.,  $\sigma^2 \sqrt{2m}$ ) is small [20]. Based on this observation, we set  $\epsilon_{xy}$ ,  $\epsilon_{y\lambda}$ , and  $\epsilon_{x\lambda}$  in the inequality constraints of the optimization problems (2.15)-(2.17) instead as

$$\begin{aligned}\epsilon_{xy}^* &= \sigma \sqrt{m_{xy} n_\lambda + \mu \sqrt{2m_{xy} n_\lambda}} \\ \epsilon_{y\lambda}^* &= \sigma \sqrt{m_{y\lambda} n_x + \mu \sqrt{2m_{y\lambda} n_x}} \\ \epsilon_{x\lambda}^* &= \sigma \sqrt{m_{x\lambda} n_y + \mu \sqrt{2m_{x\lambda} n_x}}\end{aligned}$$

with  $\mu = 2$ .

Figure 2.13 plots the SRE against  $\epsilon_{xy}$  ranging from 0 to  $2\epsilon_{xy}^*$  using the synthetic data with single-mode measurements under-sampled at 0.05 ( $m = m_{xy} n_\lambda = 104832$ ) and SNR of 30 dB ( $\sigma = 0.032$ ). Note that  $2\epsilon_{xy}^*$  in this case corresponds to almost 700 standard deviations away from the mean of  $\|\boldsymbol{\eta}_{xy}\|_2^2$ . The results suggest that the algorithm's performance is very similar for a modest range of  $\mu$  ( $\pm 30$ ), and may be fine-tuned if required.

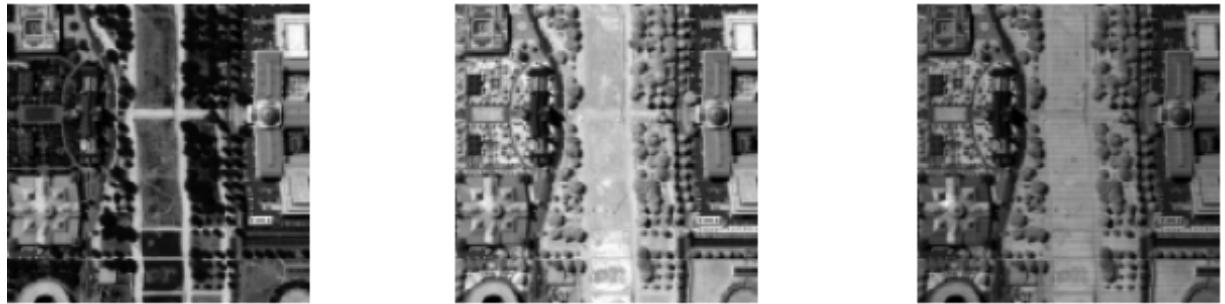


Figure 2.14: A portion of the HYDICE Washington DC Mall HSI displayed (from left to right) using spectral band 50 (658 nm), 63 (794 nm), and 86 (1115 nm).

### 2.6.2 Experimental Results with Real Hyperspectral Data

In this section, we conduct an experiment using real hyperspectral data. Compressive measurements (2.11)-(2.13) were generated from the HYDICE Washington DC Mall HSI<sup>3</sup>, which, after removing atmospheric absorption bands, contains 191 spectral bands in the range of 0.4 to 2.5  $\mu m$ . For convenience, the HSI is cropped to a  $128 \times 128 \times 128$  cube displayed in Figure 2.14, showing three different bands. Note that the intensity of the HSI has been normalized to lie between [0, 1]. We evaluate the performance with the 2D-TV minimization (2.15), the 3D-wavelet minimization (2.16), and the joint nuclear-TV norm minimization (2.17).

Figure 2.15 shows the performance evaluation results using different modalities of measurements that were distributed equally between the modes considered. The penalty and noise parameters were set in the same way as before, and the regularization parameter  $\gamma$  for (2.17) was set to  $3 \times 10^{-4}$  based on preliminary simulations. Figure 2.16, Figure 2.17, and Figure 2.18 shows the reconstructed HSI at a particular under-sampled rate (for the three different bands shown in Figure 2.14).

Comparing Figure 2.15(a), Figure 2.15(b) and Figure 2.15(c), Figure 2.15(d) we observe that with three modes of measurements, the channel-by-channel TV recovery actually outperforms the single mode recovery using the joint spatial-spectral Wavelet minimization (see also Figure 2.16 and Figure 2.17 for a visualization of the reconstructed HSI). This reiterates the idea of capturing structural dependencies within the measurements whereby using

---

<sup>3</sup>Available online at: <https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html>.

multimodal measurements with a recovery model that does not exploit inter-channel sparsity assumption resulted in better performance over single mode measurements with a model that does.

Under the same sensing matrix per partition scheme (Figure 2.16(a), Figure 2.17(a), and Figure 2.18(a)) we can see that while the uni-modal approaches have completely failed to capture the scene, the multi-modal approaches were able to retain most of the important features such as the tree outlines, roof structures, and road layouts. On the other hand, using different sensing matrices per partition, the multimodality of measurements has led to further improvements over its single mode counterparts. For example, in Figure 2.17(b), using three modes of measurements drastically reduced the Wavelet artefacts observed in the single mode case, while Figure 2.15(f) suggests an 1-3 dB improvements in the SRE by using multimodal measurements for the joint nuclear-TV minimization. This evidence positively indicates that *the multimodal measurements capture additional inter-channel signal structures about the HSI that cannot be inferred from existing sparsity models alone.*

## 2.7 Conclusions

*Global* measurements that multiplex most or all of the values of the underlying signal together are required to capture the arbitrary sparsity structure needed for the universality of the choice of sparsity basis model in the recovery process. Unfortunately, for hyperspectral imaging, such measurements require simultaneous multiplexing in the spectral and spatial dimensions, which is a challenge with the current optical and spectral modulators. Most of the existing compressive hyperspectral imagers therefore only capture *local* measurements that depend only on a subset of the entries of the hyperspectral volume being acquired. However, such independent multiplexing limits the structure of the measurements obtained and consequently impacts the performance of sparsity models that exploit structures beyond the measured dimension(s).

Throughout this chapter we have explored a different approach to the problem of hyperspectral compressive sensing acquisition. Instead of simply extending CS methods to higher

dimensions, where the sensing methods are currently impractical or unknown, we take a multi-modal approach that only requires hardware that is currently available. Since hyperspectral image acquisition is possible in a variety of modalities, we exploit these measurement techniques to produce a hybrid reconstruction method that recovers the hyperspectral image from a combination of measurements obtained via the different modalities. An algorithm, MuRecAL, has been developed for evaluating the performance of the proposed multimodal technique.

Experimental results revealed that the multimodal measurements captures additional signal structures and leads to superior reconstruction over single-mode measurements. The multi-modal measurements were also found to successfully facilitate recovery models that exploit sparsity beyond the dimensions captured by each modality that would otherwise fail if single-mode measurements were used. Additional improvements observed from using the proposed multimodal regime with state-of-the-art recovery models suggest that the multimodality also captures unique information about the data that are not inferred by the existing sparsity models alone.

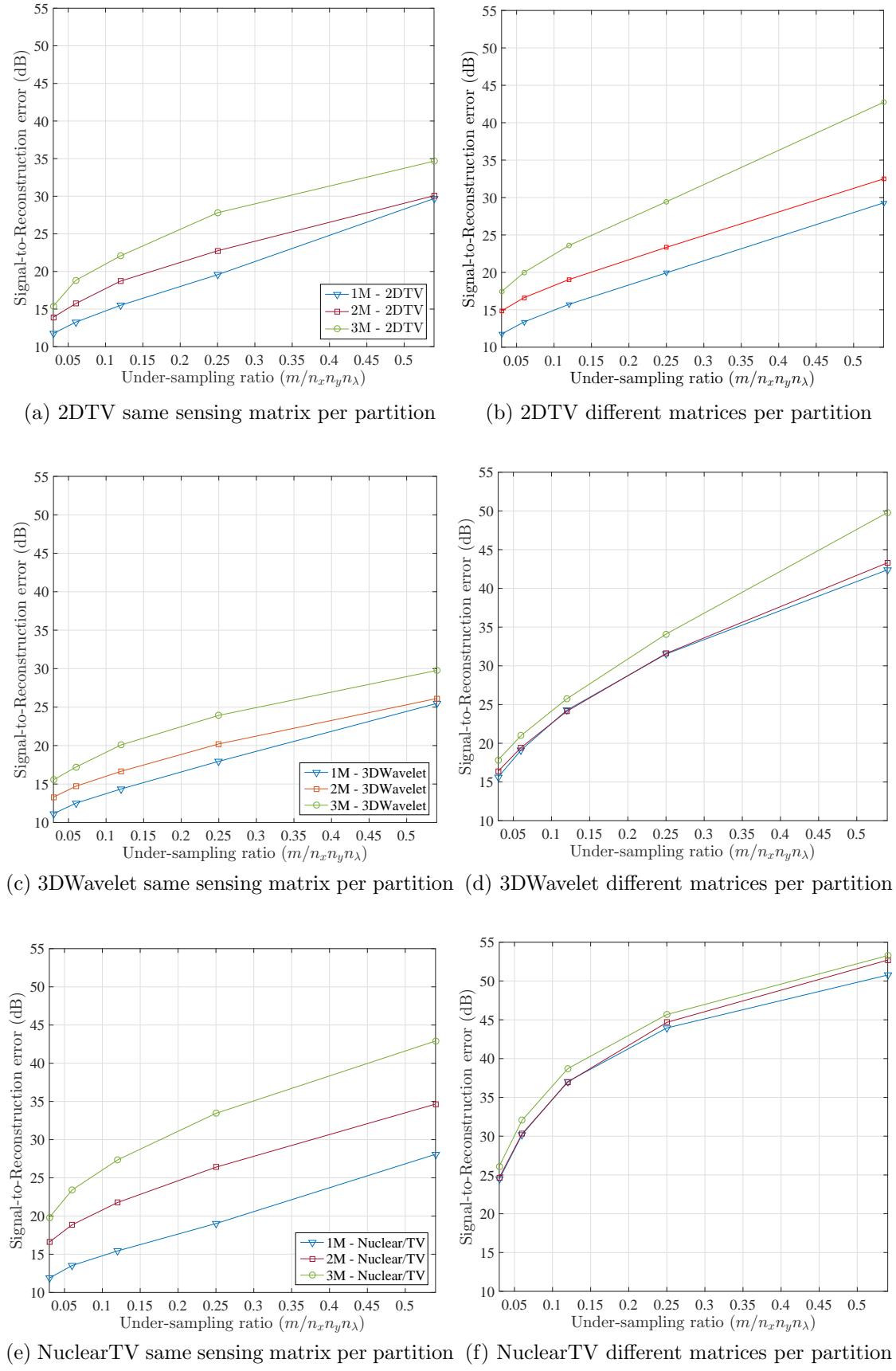


Figure 2.15: HYDICE HSI reconstruction performance from 2D-TV, 3D-Wavelet, and joint Nuclear/TV minimization.

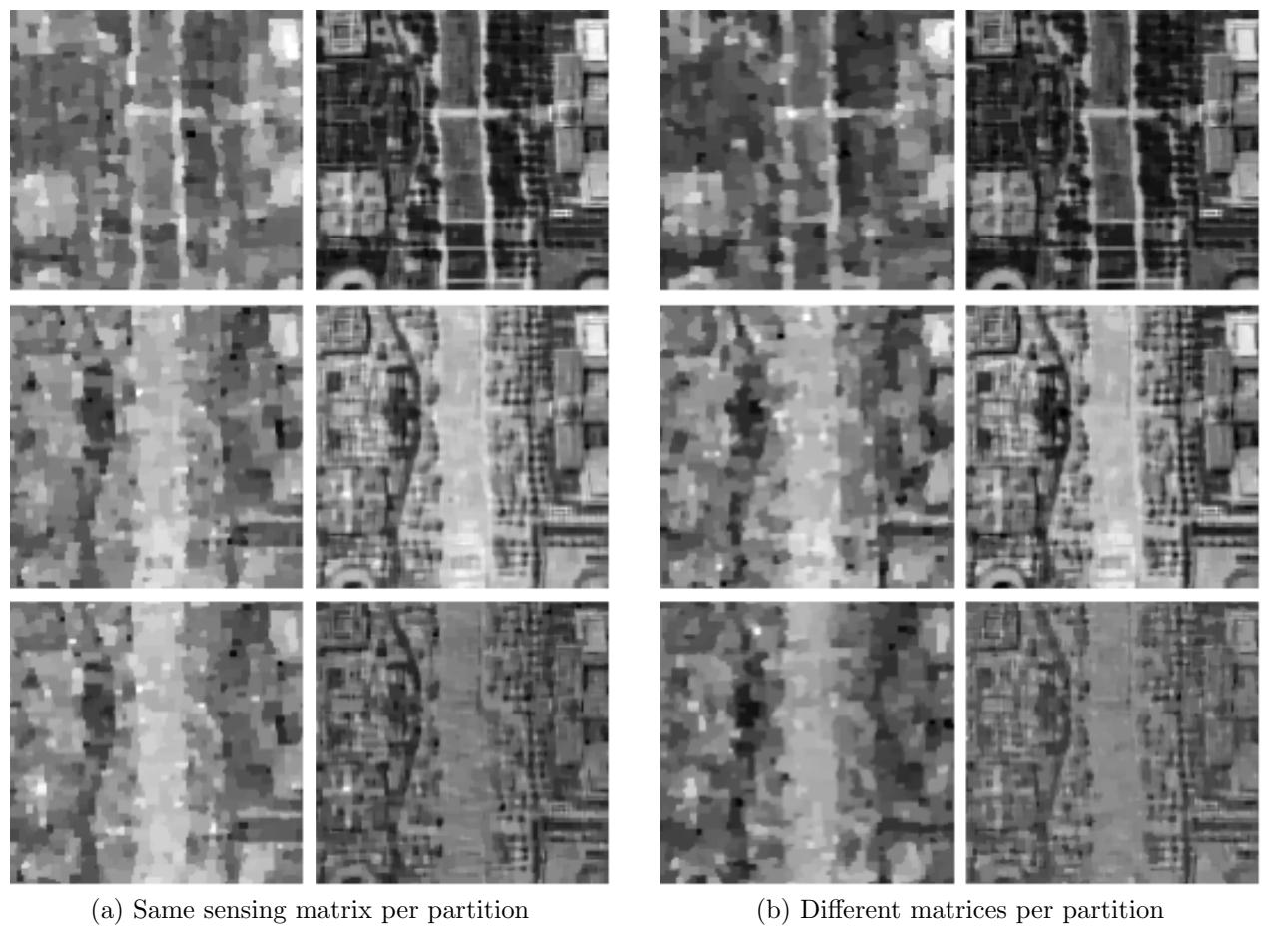


Figure 2.16: HYDICE HSI reconstructed from 2D-TV minimization for an under-sampling ratio of 0.05. Three different bands are displayed from top to bottom. Single mode reconstruction is shown on the left column while three modes is shown on the right.

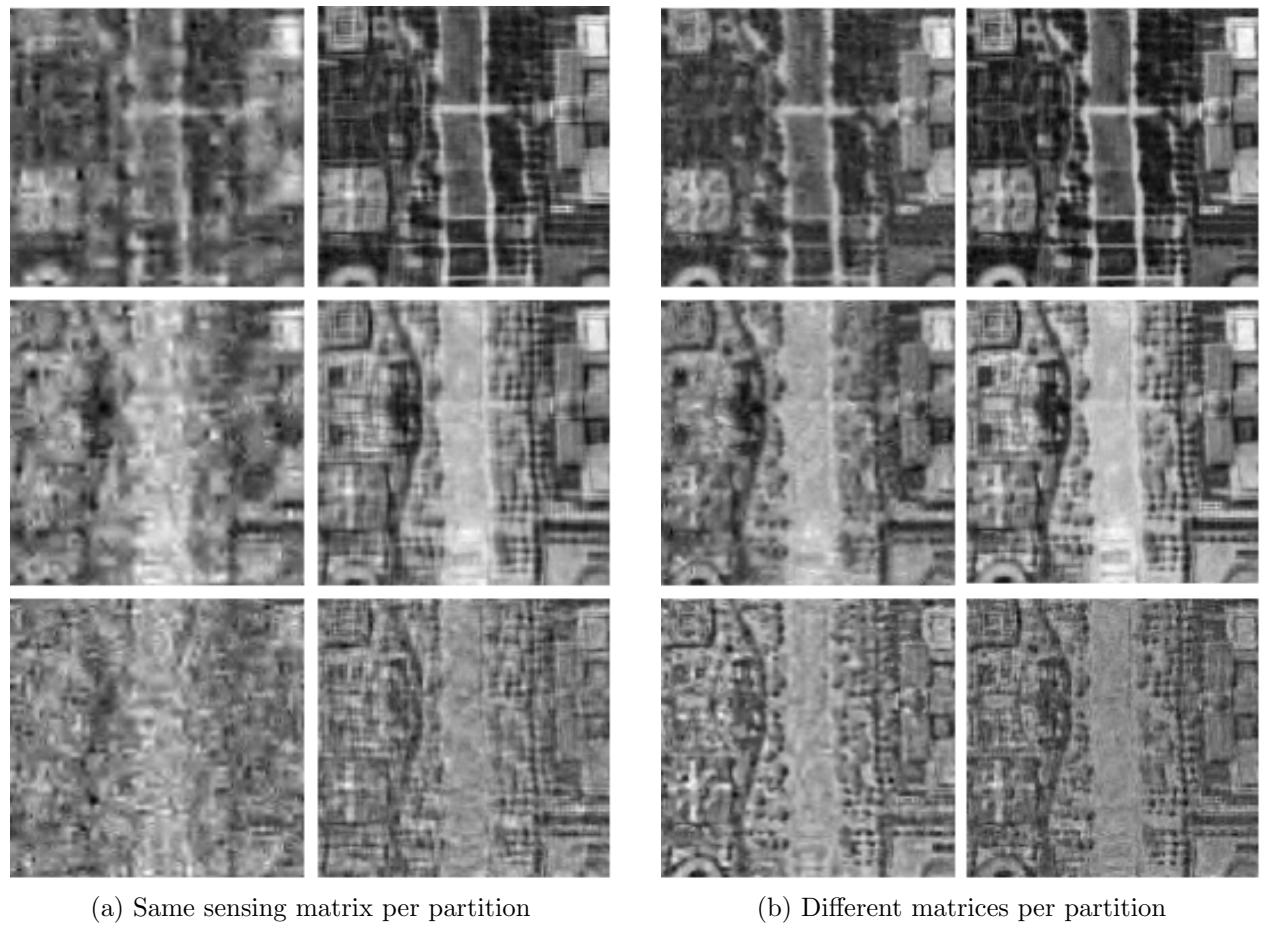
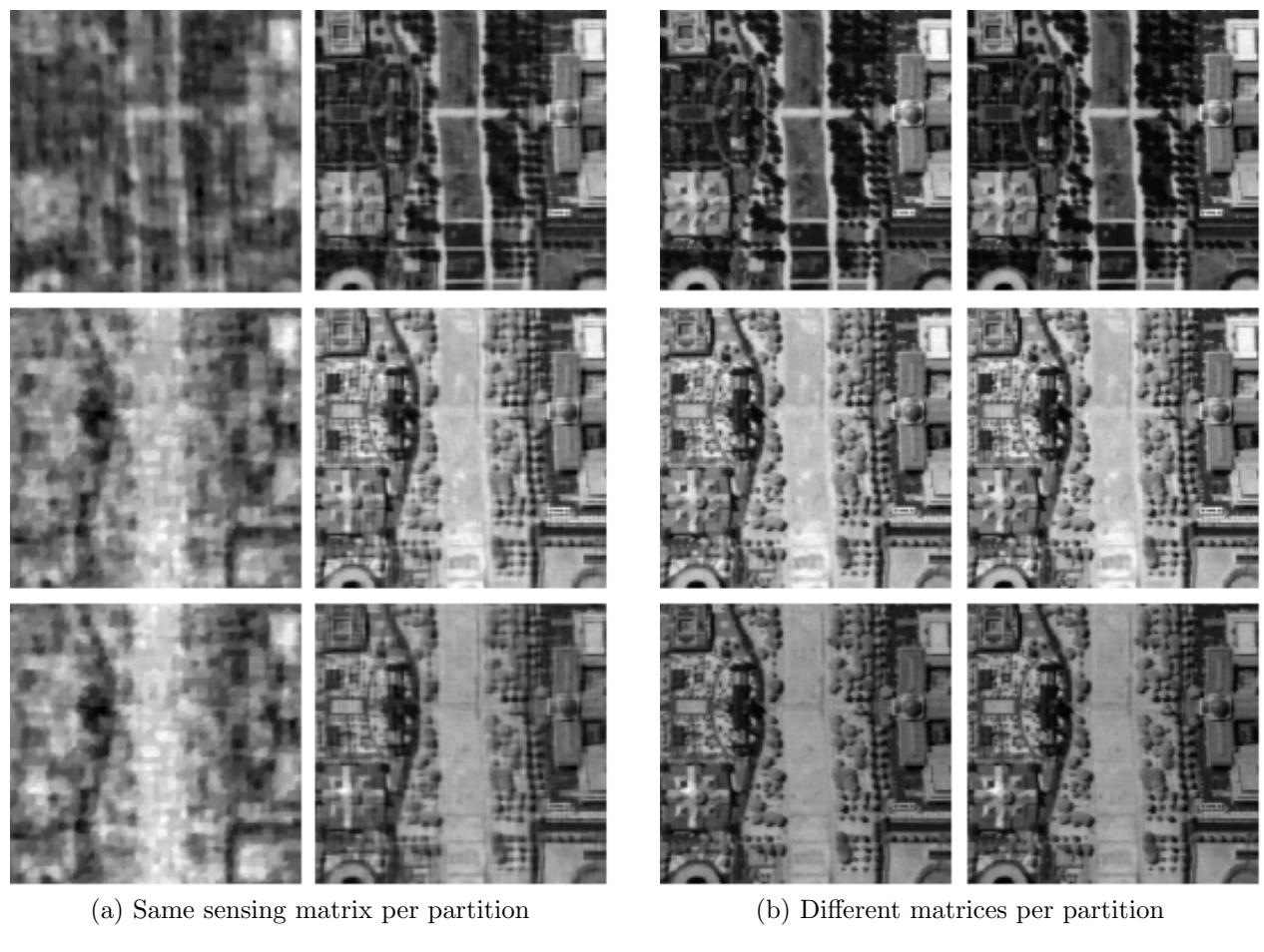


Figure 2.17: HYDICE HSI reconstructed from 3D-Wavelet minimization for an undersampling ratio of 0.05. Three different bands are displayed from top to bottom. Single mode reconstruction is shown on the left column while three modes is shown on the right.



(a) Same sensing matrix per partition

(b) Different matrices per partition

Figure 2.18: HYDICE HSI reconstructed from joint Nuclear/TV minimization for an undersampling ratio of 0.05. Three different bands are displayed from top to bottom. Single mode reconstruction is shown on the left column while three modes is shown on the right.

# Chapter 3

## Compressive Hyperspectral Image Recovery

In the previous chapter, we have witnessed several compressive hyperspectral imaging designs aimed at acquiring spectral images by a relatively few number of linear measurements. Provided the acquisition system (that implements the sensing matrix) is designed in accordance with CS theory, the recorded measurements, although seemingly lacking, contain sufficient information for accurate rendition of the underlying imagery. The practical benefit comes from the asymmetry of the CS sampling paradigm: a significant reduction in the complexity of the acquisition but at the expense of additional recovery procedures that must be carried out by the decoder to recover the imagery. Such trade-off is particularly advantageous in the hyperspectral context since existing implementations are mostly bottlenecked by the tremendous cost (of optical filters and/or photo-sensors) and effort (in storage and/or transmission bandwidth) required at the acquisition, while computational resources are usually readily available at post-acquisition for processing. Thus in hyperspectral systems where the sensor and decoder are not in close proximity (e.g., remote sensing), the sensor side no longer needs to perform compression nor transmit any covariance matrices, instead only a few linear measurements are required of the sensor that can be forwarded directly to the decoder. On the other hand, for systems where storage is at a premium (e.g., medical diagnosis), adopting the CS strategy enables an efficient storage solution by retaining just the CS measurements,

and the HSI,  $\mathbf{X} \in \mathbb{R}^{n_x n_y \times n_\lambda}$ , is only reconstructed as needed.

Since the number of measurements  $m$  is much fewer than the  $n_x n_y n_\lambda$  dimensions the original image apparently demands, the decoder must solve a heavily under-determined system of  $m$  linear equations in  $\mathbf{X}$ . Methods aiming at recovering  $\mathbf{X}$  must then incorporate prior information about the underlying data to limit the otherwise infinite solutions to the problem. As such, most approaches define a solution  $\hat{\mathbf{X}}$  to be the minimizer to the constrained optimization problem:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \phi(\mathbf{X}) \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{A} \text{vec}(\mathbf{X})\|_2 \leq \epsilon \quad (3.1)$$

where  $\phi : \mathbb{R}^{n_x n_y \times n_\lambda} \rightarrow \mathbb{R}$  is a goodness-of-fit measure, indicating the degree to which the solution  $\hat{\mathbf{X}}$  fits the priors. The priors therefore need to accurately and efficiently model the underlying data structures and correlations of the scene, while being as generic as possible. This chapter is an investigation into several recovery schemes aimed at modelling data structures inherent to the HSI.

## 3.1 Existing CoHSI Reconstruction Models

HSI typically possesses a high degree of spectral and spatial correlation, any successful priors (as hinted at the results in the previous chapter) must therefore accurately model one or both of the correlations. As dictated by the CS theory, these correlations must correspond to some form of sparsity in the hyperspectral image in order for successful recovery. This section reviews some of the prominent models used in literature for reconstructing HSI from their CS measurements.

### 3.1.1 Channel-by-Channel Reconstruction

A standard assumption used in processing of 2D images (e.g., compression, de-noising, etc.) is that they typically have a spatially sparse representation in a properly chosen 2D wavelet basis. Such an assumption is naturally applicable to hyperspectral imagery since it consists

of a collection of 2D images at various spectral bands. Let  $\Psi_{2D} \in \mathbb{R}^{n_x n_y \times n_x n_y}$  denote the 2D wavelet basis along the spatial dimensions, then a hyperspectral image  $\mathbf{X}$  can be represented by a collection of wavelet coefficient vectors  $\{\boldsymbol{\alpha}_i\}_{i=1}^{n_\lambda}$ :

$$\text{vec}(\mathbf{X}) = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_{n_\lambda} \end{bmatrix} = \begin{bmatrix} \Psi_{2D} & & & \\ & \Psi_{2D} & & \\ & & \ddots & \\ & & & \Psi_{2D} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \boldsymbol{\alpha}_{n_\lambda} \end{bmatrix} = \underbrace{(\mathbf{I}_{n_\lambda} \otimes \Psi_{2D})}_{\Psi} \text{vec}(\boldsymbol{\Lambda}_{2D})$$

where each  $\boldsymbol{\alpha}_i$  contains the 2D spatial wavelet coefficients of the corresponding spectral band image of  $\mathbf{X}$ . Since  $\boldsymbol{\alpha}_i$  are sparse by assumption, we can therefore recover  $\mathbf{X}$  by searching for a solution with the minimum  $l_1$  norm in each  $\boldsymbol{\alpha}_i$ , which can be expressed as

$$\sum_{i=1}^{n_\lambda} \|\boldsymbol{\alpha}_i\|_1 = \|\text{vec}(\boldsymbol{\Lambda}_{2D})\|_1 = \|\Psi^\top \text{vec}(\mathbf{X})\|_1,$$

via the optimization:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \|\Psi^\top \text{vec}(\mathbf{X})\|_1 \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{A} \text{vec}(\mathbf{X})\|_2 \leq \epsilon. \quad (3.2)$$

Note that this is simply the analysis formulation<sup>1</sup> of the standard BPDN problem (1.8).

Another useful assumption in natural images is their spatially piece-wise smooth structure leading to sparse discrete gradient variations. The discrete TV norm of a 2D image is defined as a summation of the magnitude of the gradients across all image pixels:

$$\|\mathbf{x}\|_{TV} = \sum_{i,j}^{n_x n_y} \left\| \begin{bmatrix} \mathbf{x}_{i,j+1} - \mathbf{x}_{i,j} \\ \mathbf{x}_{i+1,j} - \mathbf{x}_{i,j} \end{bmatrix} \right\|_p$$

where  $\|\cdot\|_p$  refers to either the anisotropic TV ( $p = 1$ ) or the isotropic TV ( $p = 2$ ) corresponding to the  $l_1$ -norm and  $l_2$ -norm of the gradients respectively. While the isotropic TV has been

---

<sup>1</sup>It has been shown that if  $\Psi$  is an orthonormal basis, the synthesis and analysis formulations are equivalent. However, if  $\Psi$  is an over-complete dictionary then the two formulations yield different results [85].

used in the seminal works [86, 87] as an efficient prior for image de-noising, the anisotropic TV has also received considerable attention recently in many imaging inverse problems for its simplicity and effectiveness (see e.g., [88–90]). Minimizing the TV induces sparsity in the discrete gradients of the signal and correspondingly a piece-wise smooth structure in the spatial domain. The anisotropic TV gives an intuitive understanding of this since for  $p = 1$  we can re-write  $\|\mathbf{x}\|_{TV}$  as

$$\begin{aligned} \|\mathbf{x}\|_{TV} &= \sum_{i,j}^{n_x n_y} |\mathbf{x}_{i,j+1} - \mathbf{x}_{i,j}| + |\mathbf{x}_{i+1,j} - \mathbf{x}_{i,j}| = \sum_{i,j}^{n_x n_y} |\mathbf{x}_{i,j+1} - \mathbf{x}_{i,j}| + \sum_{i,j}^{n_x n_y} |\mathbf{x}_{i+1,j} - \mathbf{x}_{i,j}| \\ &= \|\mathbf{D}_x \mathbf{x}\|_1 + \|\mathbf{D}_y \mathbf{x}\|_1 = \left\| \begin{bmatrix} \mathbf{D}_x \\ \mathbf{D}_y \end{bmatrix} \mathbf{x} \right\|_1 \end{aligned}$$

where  $\mathbf{D}_x, \mathbf{D}_y \in \mathbb{R}^{n_x n_y \times n_x n_y}$  are the first-order horizontal and vertical difference matrices respectively with defined boundary conditions. Minimizing the anisotropic TV is therefore equivalent to minimizing the  $l_1$ -norm of the discrete gradients of the signal and induces a signal with a sparse representation in its discrete gradients. A direct extension of this approach to the hyperspectral context is then to minimize the  $l_1$ -norm of the discrete gradients in each spectral band image, which can be expressed conveniently in a combined form

$$\sum_{i=1}^{n_\lambda} \left\| \begin{bmatrix} \mathbf{D}_x \\ \mathbf{D}_y \end{bmatrix} (\mathbf{X} \mathbf{e}_i) \right\|_1 = \left\| \begin{bmatrix} \mathbf{D}_x & & & \\ & \ddots & & \\ & & \mathbf{D}_x & \\ \mathbf{D}_y & & & \\ & \ddots & & \\ & & & \mathbf{D}_y \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_{n_\lambda} \end{bmatrix} \right\|_1 = \left\| \underbrace{\begin{bmatrix} \mathbf{I}_{n_\lambda} \otimes \mathbf{D}_x \\ \mathbf{I}_{n_\lambda} \otimes \mathbf{D}_y \end{bmatrix}}_{\mathbf{D}_{xy}} \text{vec}(\mathbf{X}) \right\|_1.$$

where  $\mathbf{e}_i$  denotes the  $i$ th unit vector in  $\mathbb{R}^{n_\lambda}$ .

In the work of [73], the authors used this approach for reconstructing HSI from the SD-CASSI measurements (Section 2.2.3). The following optimization is therefore commonly considered

as an alternative to (3.2):

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \|\mathbf{D}_{xy} \text{vec}(\mathbf{X})\|_1 \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{A} \text{vec}(\mathbf{X})\|_2 \leq \epsilon. \quad (3.3)$$

Comparing (3.3) to (3.2), the two strategies look closely related with both promoting sparsity in the separate channels of  $\mathbf{X}$  in some known representation. However, the standard RIP-based theoretical guarantee that the wavelet-minimization enjoys does not transfer directly to TV minimization since  $[\mathbf{D}_x^\top \quad \mathbf{D}_y^\top]^\top$  not only fails to be orthonormal but, when viewed as a transform, is not well-conditioned and only recently has provable guarantees been established (see [91, 92] for more details). In spite of this, there is overwhelming empirical evidence suggesting TV minimization gives better image reconstruction results over the  $l_1$ -wavelet coefficient minimization (see e.g., [19, 20] and the references therein).

### 3.1.2 Joint Spatial-Spectral Reconstruction

While the preceding methods only consider sparsity in each individual spatial slice of the hyperspectral image, neglecting any inherent structures along the spectral domain that it may exhibit, the authors of [74] and [68] noted that the voxels of hyperspectral images also have piecewise smooth variations and additionally take this into account by using a 3D wavelet basis as the sparse representation for  $\mathbf{X}$ . If  $\Psi_{1D} \in \mathbb{R}^{n_\lambda \times n_\lambda}$  denotes the 1D wavelet basis along the spectral dimension, then a hyperspectral image  $\mathbf{X}$  can be represented by the matrix  $\Lambda_{3D} \in \mathbb{R}^{n_x n_y \times n_\lambda}$  containing its 3D wavelet coefficients:

$$\mathbf{X} = \Psi_{2D} \Lambda_{3D} \Psi_{1D}^\top.$$

Viewing from the analysis form, the 3D wavelet coefficients  $\Lambda_{3D}$  is simply composed of the 2D spatial wavelet coefficient matrix  $\Lambda_{2D}$  whose rows are further decomposed by a 1D spectral wavelet achieving a joint spatial-spectral representation:

$$\Lambda_{3D} = \underbrace{(\Psi_{2D}^\top \mathbf{X})}_{\Lambda_{2D}} \Psi_{1D}.$$

Equivalently, we can express  $\Lambda_{3D}$  in the more familiar form<sup>2</sup>:

$$\text{vec}(\Lambda_{3D}) = \text{vec}(\Psi_{2D}^\top \mathbf{X} \Psi_{1D}) = (\underbrace{\Psi_{1D} \otimes \Psi_{2D}}_{\Psi_{3D}})^\top \text{vec}(\mathbf{X})$$

where  $\Psi_{3D} = (\Psi_{1D} \otimes \Psi_{2D})$  is the 3D wavelet basis. To obtain  $\mathbf{X}$  that is sparse in such a representation, the following optimization is then considered:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \|\Psi_{3D}^\top \text{vec}(\mathbf{X})\|_1 \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{A} \text{vec}(\mathbf{X})\|_2 \leq \epsilon. \quad (3.4)$$

### 3.1.3 Low-Rank Prior Reconstruction

Hyperspectral images are usually highly correlated along the spectral dimension due to the region of observation containing a limited number of materials with distinct spectral signatures. Indeed, a typical assumption for hyperspectral imagery is that the number of distinct materials,  $n_e$ , in the scene is much fewer than its spatial or spectral resolutions (i.e.,  $n_e \ll n_\lambda \leq n_x n_y$ ) and each voxel of  $\mathbf{X}$  (and hence the entire image) can be decomposed as a linear combination of the spectral signatures corresponding to these unique materials. As discussed in Section 1.2.2, this assumption forms the basis behind the linear mixing model (LMM) and reveals yet another important sparsity prior for HSI: *hyperspectral images have a low-rank structure due to high linear dependency along the spectral dimension.*

The rank of a matrix is mathematically associated with the sparsity in its singular values and this observation led the authors in [75] to recover  $\mathbf{X}$  from CS measurements by minimizing a regularised objective consisting of the *nuclear norm* (or the *trace norm*) of  $\mathbf{X}$  and the TV of individual channel images to simultaneously account for the low-rank and spatially piecewise smooth priors for HSI. This translates to the optimization problem:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \|\mathbf{D}_{xy} \text{vec}(\mathbf{X})\|_1 + \gamma \|\mathbf{X}\|_* \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{A} \text{vec}(\mathbf{X})\|_2 \leq \epsilon \quad (3.5)$$

where  $\gamma$  is a scalar parameter controlling the trade-off between the two priors. A similar

---

<sup>2</sup>For arbitrary matrices  $\mathbf{A}$ ,  $\mathbf{X}$ , and  $\mathbf{B}$  of appropriate dimensions, we have the identity  $\text{vec}(\mathbf{AXB}) = (\mathbf{B}^\top \otimes \mathbf{A}) \text{vec}(\mathbf{X})$ .

formulation using the regularized  $l_1$ -norm of the 2D wavelet coefficients of each channel instead of the TV has also been considered in [93]. Minimizing the nuclear norm  $\|\cdot\|_*$  of a matrix (equivalent to the  $l_1$  norm of its singular values) induces sparsity in its singular values (i.e., promoting a low-rank structure) and has been shown to stably recover low-rank matrices provided the measurement matrix satisfies an extended notion of RIP for matrices [94, 95].

### 3.2 CoHSI Recovery via Joint Spatio-Spectral Prior

The previous section shows that while HSI exhibits redundancies under a wide variety of models, there is no single model that can encourage all of the possible inherent correlations. But because HSI typically possesses a high degree of correlation along multiple dimensions (spatial and spectral), it is clearly favourable to promote solutions that simultaneously capture the spatial redundancies within each individual band as well as the spectral similarities across the bands, i.e., properties that are enforced by multiple priors. Since there is no “simple” model that can encourage both of these characteristics<sup>3</sup>, it is thus desirable to combine multiple models to account for them simultaneously and consider an optimization problem in the form of:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} [\phi_1(\mathbf{X}) + \gamma \phi_2(\mathbf{X})] \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{A} \text{vec}(\mathbf{X})\|_2 \leq \epsilon.$$

where each  $\phi_i : \mathbb{R}^{n_x n_y \times n_\lambda} \rightarrow \mathbb{R}$  is a measure of goodness-of-fit to, for example, one of the two aforementioned characteristics of the underlying HSI, and  $\gamma$  is a scalar parameter that controls the weighting between the two.

The standard two-dimensional TV minimization (3.3) penalizes each channel separately, treating adjacent spectral bands as independent signals. On a band-by-band basis, the recovered imagery may fit accurately within the spatial TV constraint, but as a whole the minimizer may not accurately represent the actual hyperspectral imagery. This is because such a method neglects the intrinsic structure along the spectral dimension: the spectrum at each voxel of the HSI is an aggregation of the spectrum of substances present in the

---

<sup>3</sup>By “simple”, we are referring to the complexity of minimizing the objective in (3.1).

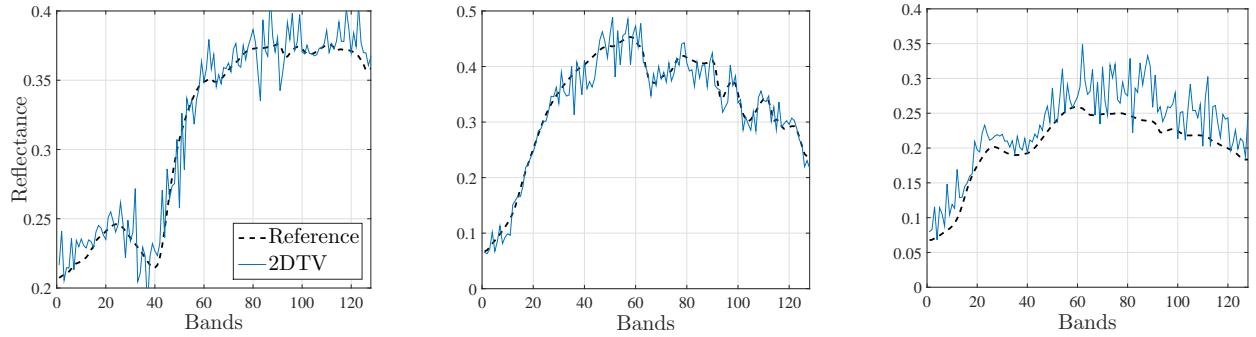


Figure 3.1: Reconstructed voxels from 1.5% CS measurements using Algorithm 3.1 for 2DTV minimization. Voxel locations are (10,120), (64,10), and (120,119) respectively of the Synthetic HSI (see Section 3.3.1).

respective pixels that are usually smooth. Therefore, the solution to (3.3) naturally exhibits a discrepancy in each voxel in the recovered  $\mathbf{X}$  as demonstrated in Figure 3.1. Such discrepancy occurs because *an accurate recovery of the gradient does not necessarily imply an accurate recovery of the image itself*. To see this, suppose the error in the recovered gradient of the  $i$ th band of  $\mathbf{X}$  (i.e.,  $\nabla \mathbf{x}_i - \nabla \hat{\mathbf{x}}_i$ ) has only a single non-zero component of size  $\delta$ , that is, the gradient is recovered perfectly except at one pixel location. This alone means every pixel in  $\hat{\mathbf{x}}_i$  could differ from that in  $\mathbf{x}_i$  by the amount  $\delta$ . If such error occurs in every band  $1 \leq i \leq n_\lambda$  with no restriction, one can only expect a lack of structure in the recovered voxels. Based on this observation, to minimize such discrepancy one could incorporate an additional penalty that promotes smoothness along the spectral dimension. A simple method to achieve this is to minimize the objective:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \left[ \sum_{i=1}^{n_\lambda} \|\mathbf{X} \mathbf{e}_i\|_{TV} + \gamma \sum_{i=1}^{n_x n_y} \|\Psi(\mathbf{X}^\top \boldsymbol{\epsilon}_i)\|_1 \right] \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{A} \text{vec}(\mathbf{X})\|_2 \leq \epsilon. \quad (3.6)$$

where  $\boldsymbol{\epsilon}_i$  denotes the  $i$ th unit vector in  $\mathbb{R}^{n_x n_y}$  and  $\Psi \in \mathbb{R}^{n_\lambda \times n_\lambda}$  is some sort of sparsifying transform that induces smoothness such as the Discrete Cosine Transform (DCT). Minimizing the  $l_1$ -norm of the DCT coefficients of each voxel will enforce smoothness, which makes sense, at least when the spectral signatures of the underlying materials encompassing the scene are smooth. While it has been empirically observed that most of the material spectra in libraries are generally smooth and continuous, we do note that such an assumption may not necessarily hold in certain applications. For example, the earth's atmospheric transmittance can induce

sharp discontinuities in the measured spectra found from remotely sensed hyperspectral data. The recovery of these localized spectral features would require the use of application-specific models and are therefore outside the scope of this work, which is to provide a general framework for the broader hyperspectral context. Note that the additional regularizer in the objective of (3.6) can be equivalently written as the matrix-vector product, similar to the 2DTV:

$$\underbrace{\sum_{i=1}^{n_\lambda} \|\mathbf{X}\mathbf{e}_i\|_{TV}}_{\|\mathbf{D}_{xy} \text{vec}(\mathbf{X})\|_1} + \gamma \underbrace{\sum_{i=1}^{n_x n_y} \|\Psi(\mathbf{X}^\top \boldsymbol{\epsilon}_i)\|_1}_{\|\Psi_\lambda \text{vec}(\mathbf{X})\|_1} = \|\mathbf{D}_{xy} \text{vec}(\mathbf{X})\|_1 + \gamma \|\Psi_\lambda \text{vec}(\mathbf{X})\|_1.$$

where  $\Psi_\lambda \in \mathbb{R}^{n_x n_y n_\lambda \times n_x n_y n_\lambda}$  is a matrix that computes the discrete cosine transform of each voxel (along the spectral dimension).

In order to solve (3.6), we rewrite the minimization as the unconstrained problem:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \|\mathbf{D}_{xy} \text{vec}(\mathbf{X})\|_1 + \gamma \|\Psi_\lambda \text{vec}(\mathbf{X})\|_1 + i_{\mathcal{C}}(\text{vec}(\mathbf{X})) \quad (3.7)$$

where  $i_{\mathcal{C}}$  is the indicator function for the convex set  $\mathcal{C} = \{\boldsymbol{\nu} \in \mathbb{R}^{n_x n_y n_\lambda} : \|\mathbf{y} - \mathbf{A}\boldsymbol{\nu}\|_2 \leq \epsilon\}$ .

We then simply apply ADMM as outlined in Section 2.5.1 to (3.7) by expressing it in the form of (2.19) (via a change of variable) as

$$\hat{\boldsymbol{\nu}} = \arg \min_{\boldsymbol{\nu}, \mathbf{z}} \underbrace{0}_{f(\boldsymbol{\nu})} + \underbrace{\|\mathbf{z}^{(1)}\|_1 + \gamma \|\mathbf{z}^{(2)}\|_1 + i_{\mathcal{C}}(\mathbf{z}^{(3)})}_{g(\mathbf{B}\mathbf{z})} \quad \text{s.t.} \quad \underbrace{\begin{bmatrix} \mathbf{z}^{(1)} \\ \mathbf{z}^{(2)} \\ \mathbf{z}^{(3)} \end{bmatrix}}_z = \underbrace{\begin{bmatrix} \mathbf{D}_{xy} \\ \Psi_\lambda \\ \mathbf{I} \end{bmatrix}}_G \boldsymbol{\nu}$$

where  $\boldsymbol{\nu} = \text{vec}(\mathbf{X})$  and  $\mathbf{B} = \mathbf{I}$ , and follow the procedure given in Algorithm 2.1. The resulting algorithm is outlined in Algorithm 3.1. The implementation details are as follows: As discussed previously (see Section 2.5.2.1), since  $\mathbf{D}_{xy}$  is diagonalizable by the discrete

**Algorithm 3.1** Algorithm for solving (3.6)

---

```

1: Set  $k = 0$ ,  $\mathbf{u}_k^{(1)}, \mathbf{u}_k^{(2)}, \mathbf{u}_k^{(3)}, \mathbf{z}_k^{(1)}, \mathbf{z}_k^{(2)}, \mathbf{z}_k^{(3)}$ , and choose  $\gamma, \rho > 0$ 
2: repeat
3:    $\boldsymbol{\nu}_{k+1} = (\mathbf{D}_{xy}^\top \mathbf{D}_{xy} + 2\mathbf{I})^{-1} \left( \mathbf{D}_{xy}^\top (\mathbf{z}_k^{(1)} - \mathbf{u}_k^{(1)}) + \boldsymbol{\Psi}_\lambda^\top (\mathbf{z}_k^{(2)} - \mathbf{u}_k^{(2)}) + (\mathbf{z}_k^{(3)} - \mathbf{u}_k^{(3)}) \right)$ 
4:    $\mathbf{z}_{k+1}^{(1)} = \text{prox}_{\|\cdot\|_1, \rho} \left( \mathbf{u}_k^{(1)} + \mathbf{D}_{xy} \boldsymbol{\nu}_{k+1} \right)$ 
5:    $\mathbf{z}_{k+1}^{(2)} = \text{prox}_{\|\cdot\|_1, \gamma/\rho} \left( \mathbf{u}_k^{(2)} + \boldsymbol{\Psi}_\lambda \boldsymbol{\nu}_{k+1} \right)$ 
6:    $\mathbf{z}_{k+1}^{(3)} = \text{prox}_{i_{\mathcal{C}}, \rho} \left( \mathbf{u}_k^{(3)} + \boldsymbol{\nu}_{k+1} \right)$ 
7:    $\mathbf{u}_{k+1}^{(1)} = \left( \mathbf{u}_k^{(1)} + \mathbf{D}_{xy} \boldsymbol{\nu}_{k+1} \right) - \mathbf{z}_{k+1}^{(1)}$ 
8:    $\mathbf{u}_{k+1}^{(2)} = \left( \mathbf{u}_k^{(2)} + \boldsymbol{\Psi}_\lambda \boldsymbol{\nu}_{k+1} \right) - \mathbf{z}_{k+1}^{(2)}$ 
9:    $\mathbf{u}_{k+1}^{(3)} = \left( \mathbf{u}_k^{(3)} + \boldsymbol{\nu}_{k+1} \right) - \mathbf{z}_{k+1}^{(3)}$ 
10:   $k = k + 1$ 
11: until  $\|\mathbf{G}\boldsymbol{\nu}_{k+1} - \mathbf{z}_{k+1}\|_2 \leq \mu_{\text{thresh}}$ 

```

---

Fourier transform, Step 3 of Algorithm 3.1 can be computed via the 3D-FFT algorithm as

$$\boldsymbol{\nu}_{k+1} = \text{IFFT3} \left( \frac{\text{FFT3} \left( \mathbf{D}_{xy}^\top (\mathbf{z}_k^{(1)} - \mathbf{u}_k^{(1)}) + \boldsymbol{\Psi}_\lambda^\top (\mathbf{z}_k^{(2)} - \mathbf{u}_k^{(2)}) + (\mathbf{z}_k^{(3)} - \mathbf{u}_k^{(3)}) \right)}{\text{diag} \left( |\tilde{\mathbf{D}}_x|^2 + |\tilde{\mathbf{D}}_y|^2 + 2\mathbf{I} \right)} \right).$$

where  $\tilde{\mathbf{D}}_x = \tilde{\mathbf{F}} \mathbf{D}_x \tilde{\mathbf{F}}^*$ ,  $\tilde{\mathbf{D}}_y = \tilde{\mathbf{F}} \mathbf{D}_y \tilde{\mathbf{F}}^*$  are the diagonalized matrices and the division is carried out element-wise. Step 4 and 5 is computed via the soft-thresholding operator (2.28), while Step 6 is the orthogonal projection onto the convex set  $\mathcal{C}$  given by:

$$\text{prox}_{i_{\mathcal{C}}, \rho}(\boldsymbol{\delta}) = \mathbf{z} + \begin{cases} 0 & \text{if } \|\mathbf{y} - \mathbf{A}\boldsymbol{\delta}\|_2 \leq \epsilon \\ \alpha \mathbf{A}^* (\mathbf{A}\mathbf{A}^*)^{-1} (\mathbf{y} - \mathbf{A}\boldsymbol{\delta}) & \text{if } \|\mathbf{y} - \mathbf{A}\boldsymbol{\delta}\|_2 > \epsilon \end{cases} \quad (3.8)$$

where  $\alpha = 1 - \epsilon \|\mathbf{y} - \mathbf{A}\boldsymbol{\delta}\|_2^{-1}$  is a scalar constant depending on  $\boldsymbol{\delta} \in \mathbb{R}^{n_x n_y n_\lambda}$ .

Alternatively, instead of inducing sparsity (penalizing the  $l_1$ -norm) in a certain transform to promote smoothness along the spectral dimension as in (3.6), such characteristic can also be enforced by promoting a gradual change in the difference between successive samples along the spectral direction. One could then consider instead penalizing the  $l_2$ -norm of the discrete

**Algorithm 3.2** Algorithm for solving (3.9)

---

1: Set  $k = 0$ ,  $\mathbf{u}_k^{(1)}, \mathbf{u}_k^{(2)}, \mathbf{u}_k^{(3)}, \mathbf{z}_k^{(1)}, \mathbf{z}_k^{(2)}, \mathbf{z}_k^{(3)}$ , and choose  $\gamma, \rho > 0$   
2: **repeat**  
3:    $\boldsymbol{\nu}_{k+1} = (\mathbf{D}_{xy}^\top \mathbf{D}_{xy} + \mathbf{D}_\lambda^\top \mathbf{D}_\lambda + \mathbf{I})^{-1} \left( \mathbf{D}_{xy}^\top (\mathbf{z}_k^{(1)} - \mathbf{u}_k^{(1)}) + \mathbf{D}_\lambda^\top (\mathbf{z}_k^{(2)} - \mathbf{u}_k^{(2)}) + (\mathbf{z}_k^{(3)} - \mathbf{u}_k^{(3)}) \right)$   
4:    $\mathbf{z}_{k+1}^{(1)} = \text{prox}_{\|\cdot\|_1, \rho} \left( \mathbf{u}_k^{(1)} + \mathbf{D}_{xy} \boldsymbol{\nu}_{k+1} \right)$   
5:    $\mathbf{z}_{k+1}^{(2)} = \text{prox}_{\|\cdot\|_2^2, \gamma/\rho} \left( \mathbf{u}_k^{(2)} + \mathbf{D}_\lambda \boldsymbol{\nu}_{k+1} \right)$   
6:    $\mathbf{z}_{k+1}^{(3)} = \text{prox}_{i_C, \rho} \left( \mathbf{u}_k^{(3)} + \boldsymbol{\nu}_{k+1} \right)$   
7:    $\mathbf{u}_{k+1}^{(1)} = \left( \mathbf{u}_k^{(1)} + \mathbf{D}_{xy} \boldsymbol{\nu}_{k+1} \right) - \mathbf{z}_{k+1}^{(1)}$   
8:    $\mathbf{u}_{k+1}^{(2)} = \left( \mathbf{u}_k^{(2)} + \mathbf{D}_\lambda \boldsymbol{\nu}_{k+1} \right) - \mathbf{z}_{k+1}^{(2)}$   
9:    $\mathbf{u}_{k+1}^{(3)} = \left( \mathbf{u}_k^{(3)} + \boldsymbol{\nu}_{k+1} \right) - \mathbf{z}_{k+1}^{(3)}$   
10:    $k = k + 1$   
11: **until**  $\|\mathbf{G}\boldsymbol{\nu}_{k+1} - \mathbf{z}_{k+1}\|_2 \leq \mu_{\text{thresh}}$

---

gradient:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \sum_{i=1}^{n_\lambda} \|\mathbf{X} \mathbf{e}_i\|_{TV} + \gamma \sum_{i=1}^{n_x n_y} \|\mathbf{D}(\mathbf{X}^\top \boldsymbol{\epsilon}_i)\|_2^2 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{A} \text{vec}(\mathbf{X})\|_2 \leq \epsilon. \quad (3.9)$$

where  $\mathbf{D} \in \mathbb{R}^{n_\lambda \times n_\lambda}$  is a first-order difference matrix. As before, we can re-write the objective in (3.9) as the matrix-vector product

$$\underbrace{\sum_{i=1}^{n_\lambda} \|\mathbf{X} \mathbf{e}_i\|_{TV}}_{\|\mathbf{D}_{xy} \text{vec}(\mathbf{X})\|_1} + \gamma \underbrace{\sum_{i=1}^{n_x n_y} \|\mathbf{D}(\mathbf{X}^\top \boldsymbol{\epsilon}_i)\|_2^2}_{\|\mathbf{D}_\lambda \text{vec}(\mathbf{X})\|_2^2} = \|\mathbf{D}_{xy} \text{vec}(\mathbf{X})\|_1 + \gamma \|\mathbf{D}_\lambda \text{vec}(\mathbf{X})\|_2^2.$$

where  $\mathbf{D}_\lambda \in \mathbb{R}^{n_x n_y n_\lambda \times n_x n_y n_\lambda}$  computes the discrete gradients of each voxel (along the spectral dimension). To solve (3.9), we follow the same procedure as described in solving (3.6) (via ADMM). The resulting algorithm is given in Algorithm 3.2. One can easily see that  $\mathbf{D}_\lambda$  also exhibits the similar triple block-circulant structure as  $\mathbf{I}_{n_\lambda} \otimes \mathbf{D}_x$  and  $\mathbf{I}_{n_\lambda} \otimes \mathbf{D}_y$  in  $\mathbf{D}_{xy}$ , it is therefore also diagonalizable by the 3D discrete Fourier Transform. Step 3 of Algorithm 3.2 can consequently be computed as

$$\boldsymbol{\nu}_{k+1} = \text{IFFT3} \left( \frac{\text{FFT3} \left( \mathbf{D}_{xy}^\top (\mathbf{z}_k^{(1)} - \mathbf{u}_k^{(1)}) + \mathbf{D}_\lambda^\top (\mathbf{z}_k^{(2)} - \mathbf{u}_k^{(2)}) + (\mathbf{z}_k^{(3)} - \mathbf{u}_k^{(3)}) \right)}{\text{diag} \left( |\tilde{\mathbf{D}}_x|^2 + |\tilde{\mathbf{D}}_y|^2 + |\tilde{\mathbf{D}}_\lambda|^2 + \mathbf{I} \right)} \right).$$

where  $\tilde{\mathbf{D}}_\lambda = \tilde{\mathbf{F}}\mathbf{D}_\lambda\tilde{\mathbf{F}}^*$ . Step 4 is computed as before via (2.28), while a direct calculation shows that the proximity operator in Step 5 becomes:

$$\mathbf{z}_{k+1}^{(2)} = \text{prox}_{\|\cdot\|_2^2, \gamma/\rho} \left( \mathbf{u}_k^{(2)} + \mathbf{D}_\lambda \boldsymbol{\nu}_{k+1} \right) = \frac{\rho}{\rho + 2\gamma} \left( \mathbf{u}_k^{(2)} + \mathbf{D}_\lambda \boldsymbol{\nu}_{k+1} \right).$$

### 3.3 Simulations and Performance Analysis

To demonstrate the potential of the proposed methods, we present numerical results based on applying the two methods on simulated compressive measurements to both synthetic and real data. In all experiments, the compressive measurements are obtained using a randomly sub-sampled 2D discrete Fourier orthogonal basis as the measurement matrix, applied to each spatial band of the hyperspectral image. Based on the findings in Chapter 2, a different sensing matrix was used for each band in order to exploit structures beyond the measured dimensions.

Algorithm 3.1 was initialized with all variables set to 0 and the algorithm is terminated when  $\|\mathbf{G}\boldsymbol{\nu}_{k+1} - \mathbf{z}_{k+1}\|_2$  falls below  $\mu_{thresh} = 10^{-4}$ . As discussed in Section 2.6.1.1 and Section 2.6.1.2, the penalty,  $\rho$ , and noise,  $\epsilon$ , were selected appropriately by the speed of convergence and the expected noise distribution respectively. As a quantitative metric, the signal-to-reconstruction error ratio was used to measure the quality of the HSI reconstruction:

$$\text{SRE}(dB) = 10 \log_{10} \left( \frac{\|\text{vec}(\mathbf{X})\|_2^2}{\|\text{vec}(\mathbf{X} - \hat{\mathbf{X}})\|_2^2} \right)$$

All experiments were performed under MATLAB v7.13 (R2011b) running on an Desktop PC with an Intel Core2 Quad CPU at 2.6 GHz and 4 GB of RAM.

#### 3.3.1 Experimental Results with Synthetic Hyperspectral Data

The synthetic hyperspectral data used in this test is a  $128 \times 128 \times 128$ -pixel hyperspectral cube composed of four randomly selected spectral signatures from the USGS spectral library [96] in the range of 0.2 to 3  $\mu\text{m}$  (see Figure 3.2). The composition maps of the four materials

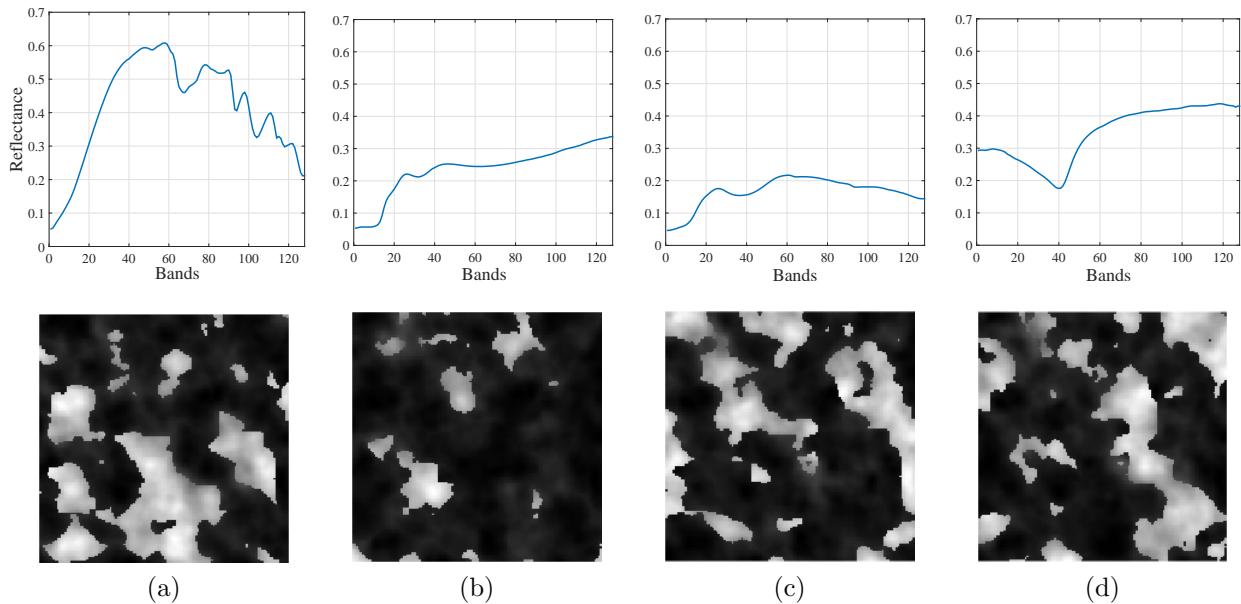


Figure 3.2: Material constituents making up the synthetic hyperspectral data. The composition maps indicate the percentage of contribution to each and every pixels of the hyperspectral data by the corresponding material.

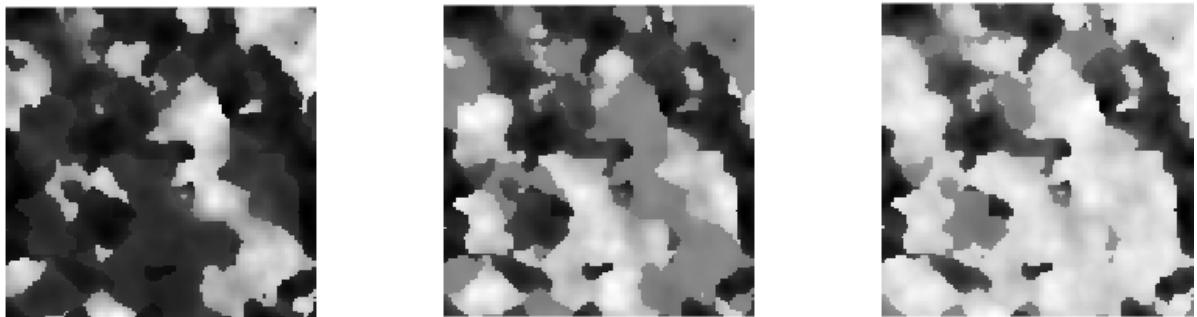


Figure 3.3: Synthetic HSI displayed (from left to right) using spectral band 5, 64, and 100.

at each of the spatial locations are also shown in Figure 3.2, these were generated using the publicly available Hyperspectral Imagery Synthesis toolbox<sup>4</sup>. Three different bands of the resulting hyperspectral image are shown in Figure 3.3.

The average performance (over 5 Monte-Carlo runs) evaluation results are plotted in Figure 3.4 at different under-sampling ratios, with and without additive white Gaussian noise. Note that in accordance with CS theory, there were minimal variations between each of the 5 runs as the measurement matrices were well-conditioned for the cases considered, thus resulting in uniform recovery. Figure 3.5 shows the recovered voxels at the same locations as

<sup>4</sup>Available online: [http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral\\_Imagery\\_Synthesis\\_tools\\_for\\_MATLAB](http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Imagery_Synthesis_tools_for_MATLAB).

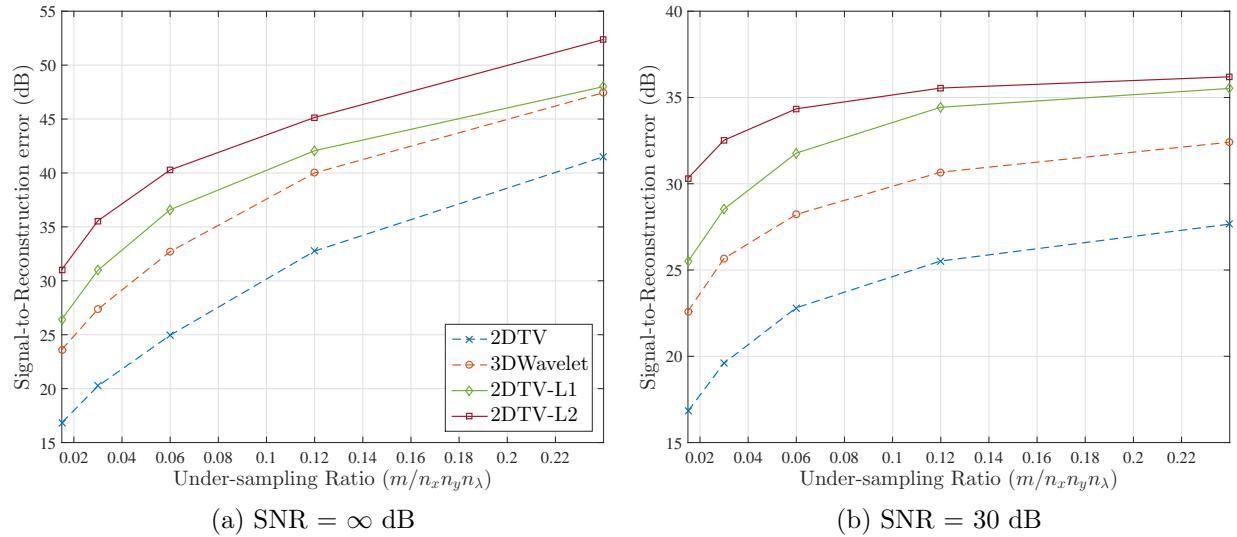


Figure 3.4: Synthetic HSI reconstruction performance at various sampling rates and SNR.

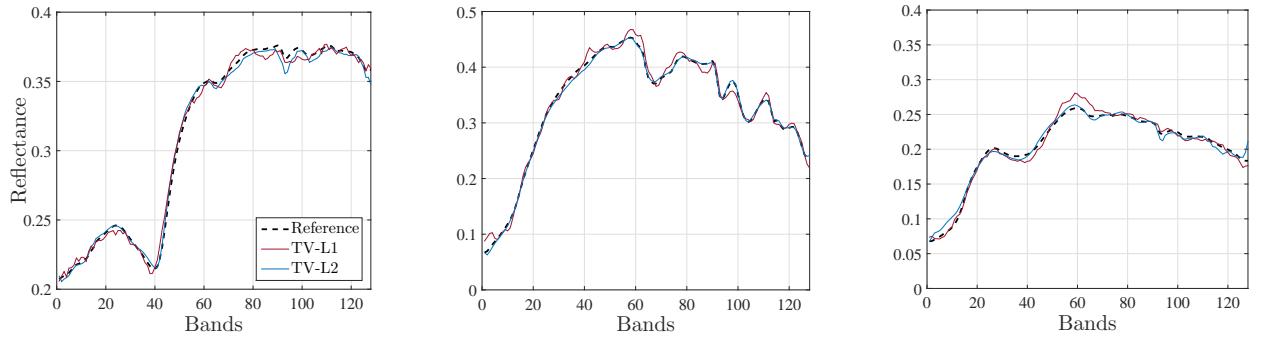


Figure 3.5: Reconstructed voxels from 1.5% CS measurements using Algorithm 3.1. Voxel locations are (10,120), (64,10), and (120,119) respectively of the Synthetic HSI. A direct comparison with Figure 3.1 shows that the proposed priors successfully mitigated the discrepancy between adjacent bands.

shown in Figure 3.1 using the two proposed methods, and the reconstructed HSI using 1.5% CS measurements (at three different bands) is shown in Figure 3.6 for the various recovery strategies. For our methods, the results were obtained with  $\gamma = 4.5$  for (3.6) and  $\gamma = 3 \times 10^3$  for (3.9) based on preliminary simulations, carried out in order to find the appropriate values.

From Figure 3.4 we can observe that solutions that exploit both spatial and spectral redundancies improve the reconstruction SRE significantly when compared to the channel-by-channel scheme (2DTV). Further improvements over the 3D-Wavelet minimization from the proposed method indicate that the inter-channel structure enforced by our methods (i.e., the smooth variations along the spectral dimension) together with the intra-channel structure

(i.e., the spatial piece-wise smoothness) accurately model the synthetic HSI. In Figure 3.5 the measurements and initial conditions were identical to that of Figure 3.1, thus the only variability between the two experiments is the addition of the proposed priors, and has convincingly mitigated the discrepancy exhibited in the channel-by-channel reconstruction.

Qualitatively, Figure 3.6 shows that, with just 1.5% CS measurements, the proposed method (Figure 3.6(c) and Figure 3.6(d)) is able accurately recover the underlying HSI for which the standard 2D-TV and 3D-Wavelet minimization (Figure 3.6(a) and Figure 3.6(b)) fail to do so. In particular, solving (3.9) achieves almost perfect visual reconstruction with a relative error of less than 2% (31.05 dB). The success on synthetic data suggests that modelling the spectral correlations through separate smooth-inducing priors is potentially more efficient than a 3D Wavelet representation for HSI, and encourages further tests on more realistic data.

### 3.3.2 Experimental Results with Real Hyperspectral Data

In this section, we test the proposed methods using two real hyperspectral datasets. The first is the publicly available SpecTIR Gulf of Mexico HSI<sup>5</sup>, cropped to a spatial resolution of  $128 \times 128$ -pixels with 128 spectral bands (ignoring water absorption bands). Figure 3.7 shows three different bands of the resulting HSI. The second HSI is the HYDICE Washington DC Mall HSI used in Chapter 2 (see Section 2.6.2) of size  $128 \times 128 \times 128$ , with three bands shown in Figure 2.14.

The performance evaluation results (averaged over 5 Monte-Carlo runs) are plotted in Figure 3.8 and Figure 3.10. Figure 3.9 and Figure 3.11 display the recovered HSI by the various methods for different under-sampling ratios. The regularization parameter was set to  $\gamma = 10$  for (3.6) and  $\gamma = 9 \times 10^3$  for (3.9) during the runs for both datasets based on preliminary tests.

As expected, we observe that in all cases, exploiting redundancies along both spatial and spectral directions improved the reconstruction quality drastically. In particular, solving (3.9) resulted in the best performance, with an enhancement in the SRE by 5 to 7 dB

---

<sup>5</sup>Available online at: <http://www.spectir.com/free-data-samples/>

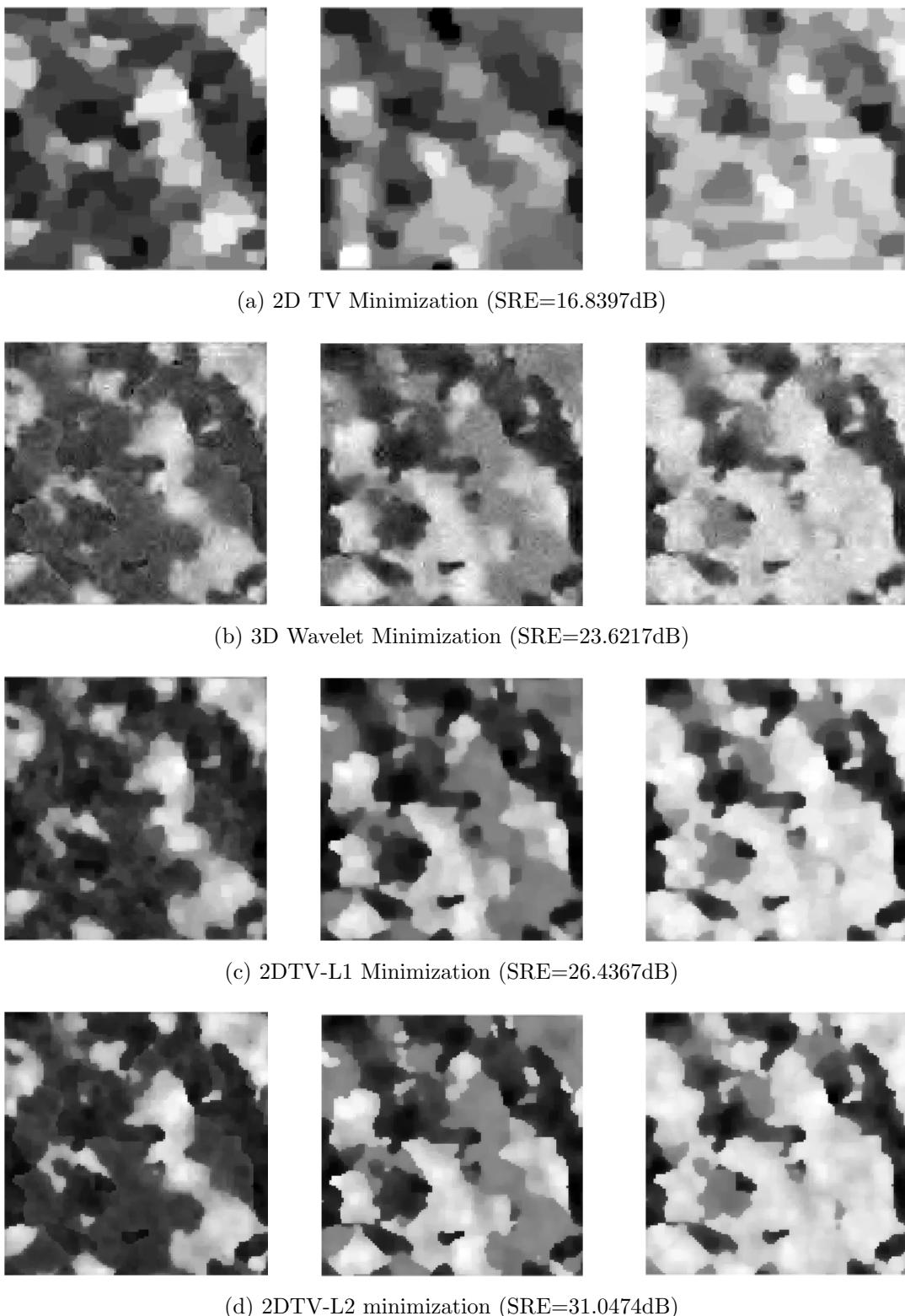


Figure 3.6: Synthetic HSI reconstructed from 1.5% CS measurements using (a) Channel-by-channel TV, (b) 3D Wavelet, (c) Spatial-Spectral TV-L1, and (d) Spatial-Spectral TV-L2 minimization. The displayed bands are 5, 64, and 100.



Figure 3.7: Gulf of Mexico HSI displayed (from left to right) using spectral band 5, 50, and 100.

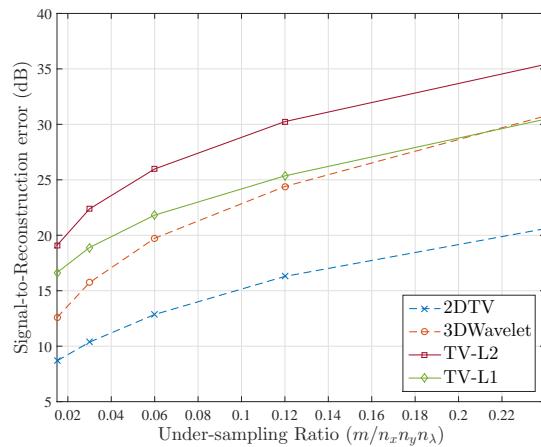


Figure 3.8: Gulf of Mexico HSI reconstruction performance at various sampling rates.

when compared to the standard 3D Wavelet minimization. For both datasets, Figure 3.9 and Figure 3.11 reveal that the solution to (3.9) (TV-L2) accurately recovers the pertinent details of the HSI even in severely under-sampled situations (e.g., when the number of CS measurements is only 3% of the original data size) for which the other methods cannot. Interestingly, for the HYDICE dataset, Figure 3.11 shows that the solution to (3.6) (TV-L1) performed much worse than the 3D Wavelet minimization. We speculate that this phenomenon occurs because the dataset has a much sparser representation in the 3D Wavelets as shown in Figure 3.12 in which we plot the sorted magnitude of the transformed coefficients for the two datasets. Noticeably, the 3D Wavelet coefficients have a higher rate of decay when compared to the spectral DCT for the HYDICE dataset, which is less apparent in the SpecTIR data.

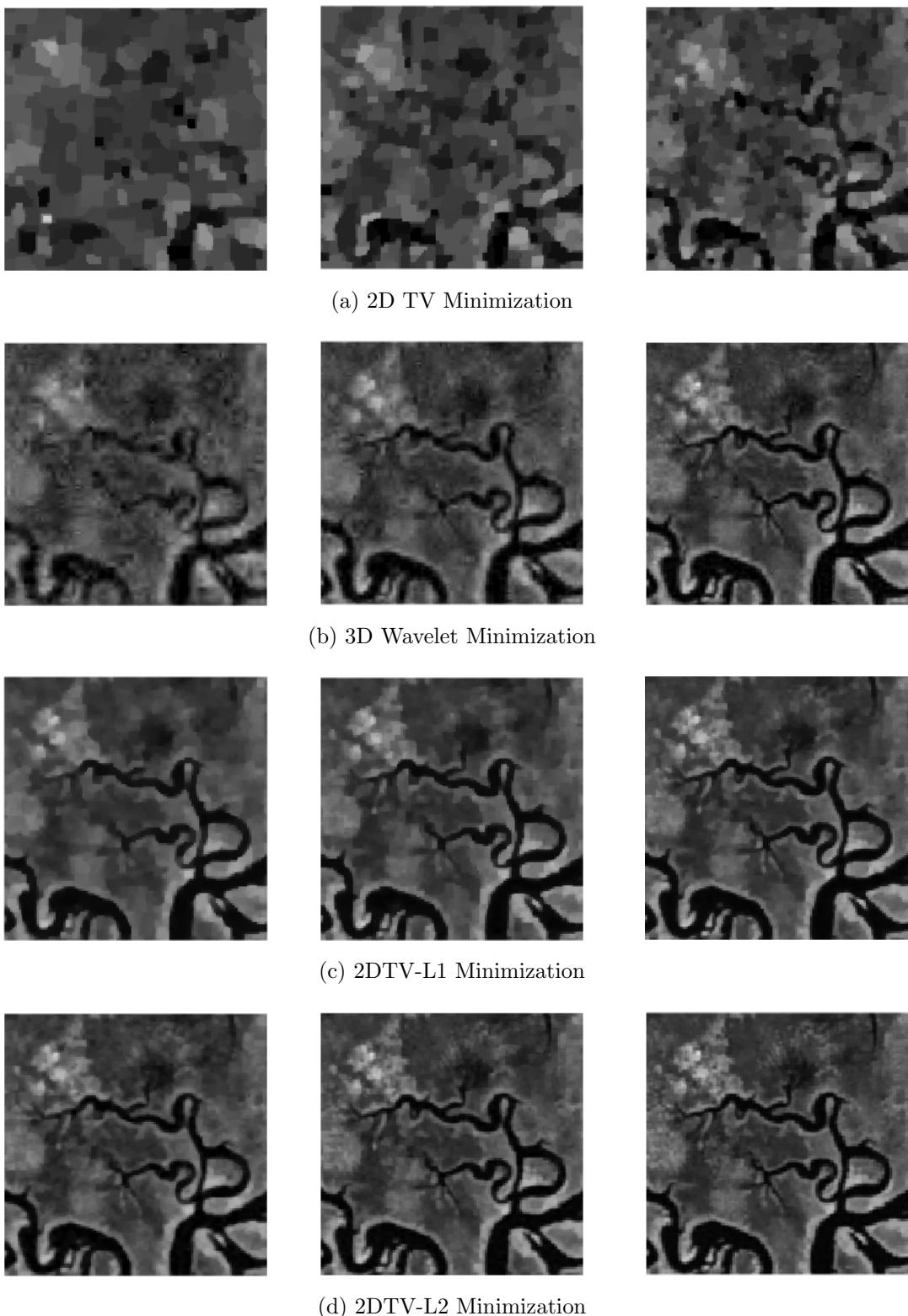


Figure 3.9: Reconstructed Gulf of Mexico HSI (showing spectral band 50) using (going from left to right) 1.5%, 3%, and 6% CS measurements.

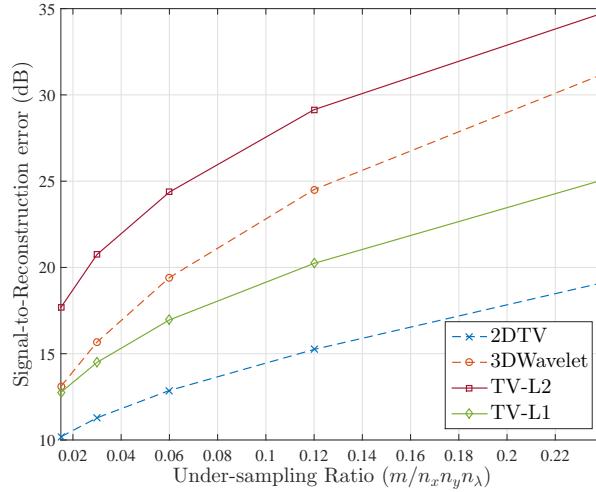


Figure 3.10: HYDICE Washington DC Mall HSI reconstruction performance at various sampling rates.

### 3.4 Conclusions

Reconstructing hyperspectral imagery from compressive sensing imagers requires finding a solution to an underdetermined system of linear equations. As suggested by the fundamental theorem of linear algebra, in general such systems admit infinitely many solutions. Finding appropriate priors that narrow down the infinite solutions to the original data as well as being computationally feasible therefore lies at the core of any method aiming at successful compressive hyperspectral image recovery. Mathematically, these priors usually appear in the form of objectives to be minimized, constrained by the CS measurements that characterise the linear system. With enough measurements, the solution under these objectives should accurately represent the original image.

Hyperspectral images typically possess a high degree of correlation along both their spatial and spectral dimensions. Ideal priors should therefore encourage all of the possible data structure and inherent correlations that the underlying data exhibit. Unfortunately, there are no priors that can capture both the spatial redundancies within each spectral band as well as the spectral similarities in each voxel across the bands, let alone efficient algorithms for minimizing such objectives. It is therefore desirable to linearly combine multiple pertinent priors together so that the minimizer to these objectives will simultaneously exhibit properties induced by those priors.

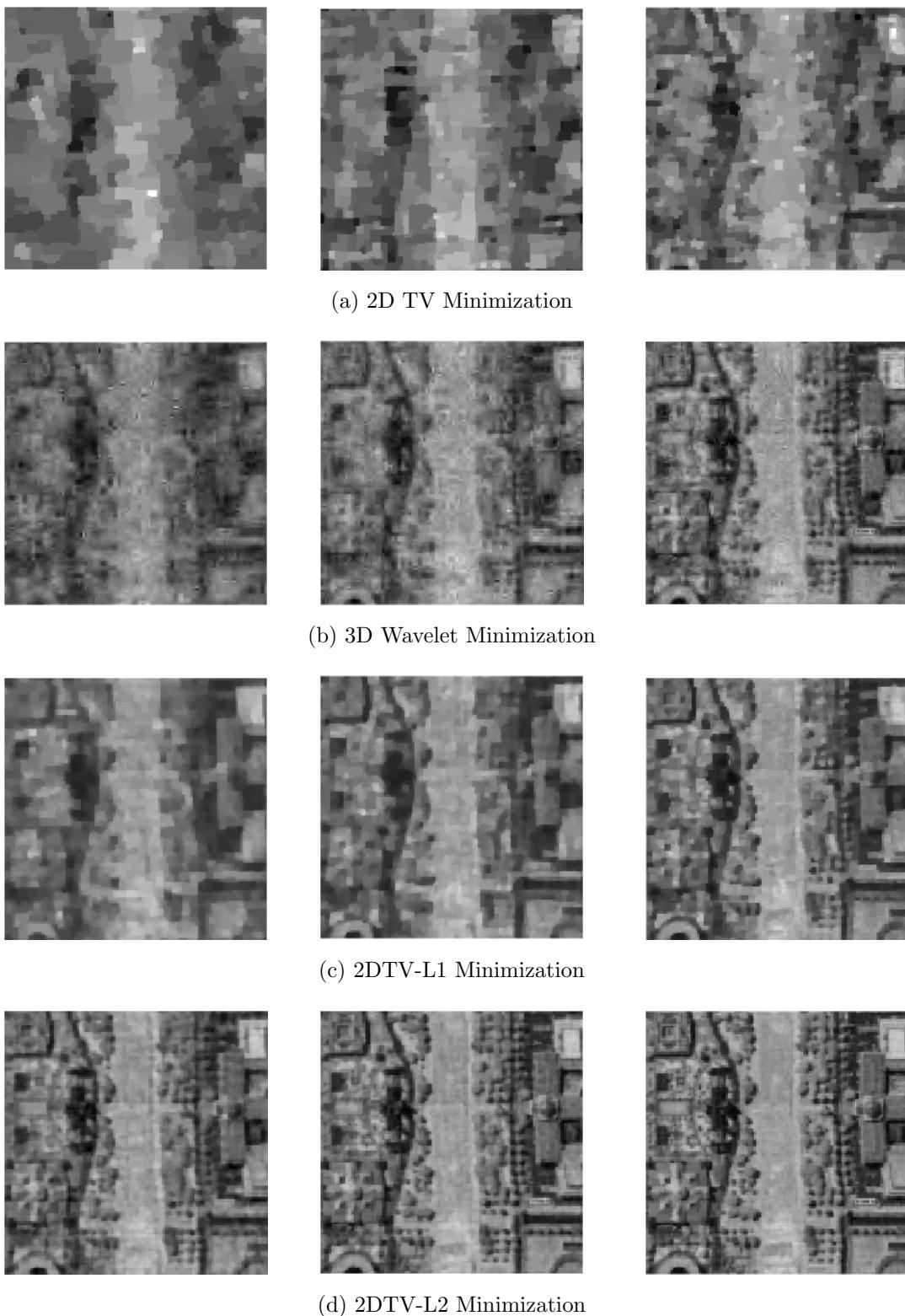


Figure 3.11: Reconstructed HYDICE Washington DC Mall HSI (showing spectral band 64) using (going from left to right) 1.5%, 3%, and 6% CS measurements.

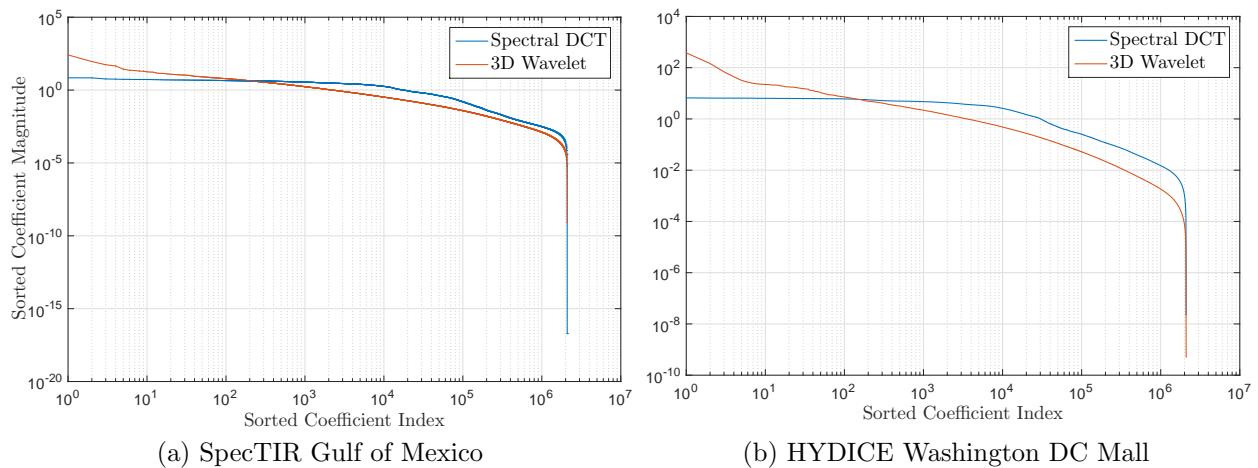


Figure 3.12: Sparsity of the considered datasets in the 3D Wavelet (red) and spectral DCT coefficients (blue) demonstrated by the decay of their sorted magnitudes. The Wavelet coefficients are much sparser in the HYDICE dataset compared to the spectral DCT as evidenced by steeper rate of decay.

In this chapter, we have investigated two recovery models which seek a solution that is spatially divided into homogeneous regions and spectrally smooth along each voxel. Both properties are based on the observation that, globally, a hyperspectral image is a concatenation of multiple 2D images at various spectral bands that are piece-wise continuous while, locally, each pixel along the spectral direction is a linear combination of a few material constituents whose spectral signatures are usually smooth. The former property is induced by a sum of the total variation in each band of the HSI and the latter is induced by (1) a sum of the  $l_1$ -norm of each voxel in the discrete cosine basis, and (2) a sum of the  $l_2$ -norm of the discrete gradient associated with each voxel. Both methods are tested and implemented via the ADMM algorithm in which the intermediate steps can be computed efficiently using the fast Fourier transform. Experiments on both synthetic and real hyperspectral data revealed, quantitatively and qualitatively, that modelling the data structure through the proposed priors accurately recovers the underlying HSI. In particular, minimizing the  $l_2$ -norm of the discrete gradient of the voxels significantly improves the reconstruction quality, even over existing joint spatio-spectral models such as the 3D Wavelet minimization and successfully reduces the number of CS measurements necessary for accurate rendition of the underlying data.



# Chapter 4

## Compressive Hyperspectral Unmixing

In many practical applications of hyperspectral imaging (e.g., terrain classification, mineral detection, environmental monitoring, etc.) what is often of particular interest is being able to identify the spectra (or *endmembers*) corresponding to the different materials encompassing the measured spectrum of a given pixel, together with their relative contributions (or *abundances*) to each and every pixel within a given hyperspectral image. This process, as discussed in Section 1.2.2, is called *spectral unmixing*. Under the most frequently used model for spectral unmixing, the LMM reduces the problem roughly to finding a relatively small number of endmembers of which their linear combinations describe the hyperspectral image accurately (usually in a least-square sense). If the endmembers of a given scene are known beforehand (e.g., from spectral libraries such as ASTER [54] or USGS [96]), we can mathematically decompose each pixel’s spectrum to find the abundance of each endmember. On the other hand, when the endmembers are unknown but the hyperspectral data is fully accessible, many algorithms exist for determining the endmembers in the scene including N-FINDR [97], VCA [98], MVSA [99], MVC-NMF [100], etc. (see [59] for a recent overview of various endmember determination algorithms).

In the previous chapter we have examined several approaches in reconstructing the hyperspectral volume from its compressive measurements, but it is apparent that many (if not most) hyperspectral applications (in industry, agriculture, and military, for example) do not actually require reconstructing the entire hyperspectral volume. Instead, it is particularly

favourable to be able to perform spectral unmixing directly from the compressive measurements. While one could always attempt to recover the full hyperspectral image from the compressive measurements via the various CS recovery methods, and then perform the unmixing using any of the well-established unmixing algorithms, such a costly and inefficient procedure naturally leads to a strong motivation for methods that combine both data reconstruction and unmixing into a single step without ever reconstructing or storing the entire hyperspectral cube itself. Studies have therefore stemmed from such need, with initial works on determining abundances via compressive measurements assuming prior knowledge of the endmembers within the scene [71, 101, 102]. Although useful in situations where the materials within the scene have previously been identified or known to be contained within a selection, unfortunately variability in illumination, environmental, atmospheric, and temporal conditions within and between hyperspectral data collections means that the measured spectra will inevitably vary in comparison to those found in the libraries which have been collected under controlled laboratory conditions [103]. It is therefore particularly interesting to consider the problem of inferring both the materials present together with their compositions. Although recent works in this area have demonstrated promising results [104, 105], they come at the cost of strict assumptions on the underlying data: *pure endmember spectra need to be present dominantly over pixels containing a mixture of them.* However, as discussed in Chapter 1, this is rarely the case in practice due to limited spatial resolution and mixing phenomena at different scales.

In line with these works, this chapter explores and develops a blind compressive spectral unmixing technique that aims to recover simultaneously the endmembers and their abundances directly from compressive measurements of an hyperspectral image, bypassing the need to recover and store the image itself. However, in contrast to previous works, our rationale behind the work is to mitigate the assumption of the dominant presence of pure endmembers within the underlying data. Through simulation, we demonstrate that our approach accurately recovers the underlying sources from only a few CS measurements even when the hyperspectral data is highly mixed.

## 4.1 Prior Art

For a given hyperspectral image  $\mathbf{X} \in \mathbb{R}^{n_x n_y \times n_\lambda}$  with  $n_e$  endmembers, the LMM (see Section 1.2.2) asserts that  $\mathbf{X}$  can be decomposed as

$$\mathbf{X} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_{n_e}] \begin{bmatrix} \mathbf{e}_1^\top \\ \mathbf{e}_2^\top \\ \vdots \\ \mathbf{e}_{n_e}^\top \end{bmatrix} = \mathbf{A}\mathbf{E}^\top \quad \text{subject to} \quad \sum_{i=1}^{n_e} \mathbf{a}_i = \mathbf{1}_{n_x n_y} \quad (4.1)$$

$$\mathbf{a}_i, \mathbf{e}_i \succeq 0 \quad \forall i = \{1, \dots, n_e\}$$

where the matrix  $\mathbf{E} = [\mathbf{e}_1, \dots, \mathbf{e}_{n_e}] \in \mathbb{R}^{n_\lambda \times n_e}$  contains a column-wise arrangement of the endmembers,  $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_{n_e}] \in \mathbb{R}^{n_x n_y \times n_e}$  is the matrix whose columns represents the abundance maps of each corresponding endmember, and  $\mathbf{1}$  is a vector in the subscripted dimension whose elements are all ones. The abundance maps indicate the percentage (as a fraction) of each corresponding endmember within a certain pixel of  $\mathbf{X}$ , they must therefore sum to one as enforced by the first constraint in (4.1) (i.e., the STU constraint). The non-negativity constraints on the other hand signify that real-world materials do not have negative reflectance spectra. Each pixel is therefore a linear combination of the endmembers that reside within the hyperspectral image. The idea of spectral unmixing from the linear measurements of HSI,  $y = \mathcal{A}(\mathbf{X}) = \mathcal{A}(\mathbf{A}\mathbf{E}^\top)$ , has gained popularity in light of recent advances in CS and as a result, several CS based unmixing methods have been developed.

### 4.1.1 Compressive Unmixing with Endmember Priors

The common assumption in these works is the availability of spectral libraries containing the materials within a given scene in the HSI. As such,  $\mathbf{E}$  is either known in advance or represents an over-complete dictionary of known spectra, thus reducing the problem to finding  $\mathbf{A}$  (and the correct subset of  $\mathbf{E}$ ). The two key priors exploited for recovery here are:

- (i) Each voxel is a non-negative linear mixture of the endmembers present (i.e., the linear mixing model), and

- (ii) the piecewise smooth variations exhibited by each abundance map (columns of  $\mathbf{A}$ ) leading to a sparse representation in a properly chosen 2D wavelet basis, or equivalently sparse 2D gradients along the spatial domain.

Based on these two priors, the following optimization problem is commonly considered:

$$\begin{aligned} \hat{\mathbf{A}} = \arg \min_{\mathbf{A}} \sum_{i=1}^{n_e} \text{TV}(\mathbf{a}_i) \quad & \text{subject to} \quad \sum_{i=1}^{n_e} \mathbf{a}_i = \mathbf{1}_{n_x n_y} \\ & \mathbf{A} \succeq 0 \end{aligned} \quad \| \mathbf{y} - \mathcal{A}(\mathbf{A} \mathbf{E}^\top) \|_2 \leq \epsilon \quad (4.2)$$

In [71], the authors proposed to recover  $\mathbf{A}$  using (4.2) for the particular case when  $\epsilon = 0$ , with the measurement operator  $\mathcal{A}$  as the partial Walsh-Hadamard orthogonal basis (each band is sampled using the same matrix). The non-negativity constraint was ignored as their experimental results seemed satisfactory without it. While their specific model did not account for model imprecision or noise, to compensate, the authors proposed a pre-processing procedure using the Singular Value Decomposition (SVD) to act as a de-noising step (also as a dimensionality reduction). This involved computing the rank- $n_e$  economy-sized SVD of the measurements  $\mathbf{Y} = \Phi \mathbf{A} \mathbf{E}^\top = \mathbf{U} \Sigma \mathbf{V}^\top$  where  $\mathbf{U} \in \mathbb{R}^{m \times n_e}$ ,  $\Sigma \in \mathbb{R}^{n_e \times n_e}$ ,  $\mathbf{V} \in \mathbb{R}^{n_\lambda \times n_e}$  and multiplying both side of the fidelity constraint by  $\mathbf{V}$  to obtain

$$\mathbf{Y} = \Phi \mathbf{A} \mathbf{E}^\top \Leftrightarrow \mathbf{U} \Sigma = \Phi \mathbf{A} (\mathbf{E}^\top \mathbf{V})$$

It was proved in [71] that the new linear system has an identical solution set to the original provided  $\text{rank}(\mathbf{E}^\top \mathbf{V}) = n_e$ . An ADMM algorithm (see Section 2.5.1) was proposed therein to solve the problem.

In the work of [102], the authors similarly proposed to solve (4.2) but without the STU constraint and with  $\mathcal{A}$  acting on individual voxels of the hyperspectral data using i.i.d. Gaussian ensembles. An equivalent unconstrained TV regularized least-squares problem to (4.2) is also considered in their work:

$$\hat{\mathbf{A}} = \arg \min_{\mathbf{A}} \frac{1}{2} \| \mathbf{y} - \mathcal{A}(\mathbf{A} \mathbf{E}^\top) \|_2^2 + \alpha \sum_{i=1}^{n_e} \text{TV}(\mathbf{a}_i) \quad \text{subject to} \quad \mathbf{A} \succeq 0 \quad (4.3)$$

however, (4.2) circumvents the need to tune the regularization parameter  $\alpha$ . (4.2) and (4.3) are equivalent in the sense that for a properly chosen  $\epsilon \geq 0$ , there exists an  $\alpha \geq 0$  for which the solution to both problems coincide [106]. Both problems were solved via ADMM.

The work of [101] on the other hand was the first to tackle the complete unmixing problem, incorporating all constraints as in (4.2). The authors also considered the 2D wavelet as a sparsity prior in addition to the TV. The sensing matrix used depicts that of a partial random convolution [43] (a special case of the partial random Toeplitz matrices discussed in Chapter 1) applied to each individual band of  $\mathbf{X}$ . A de-correlation scheme was also introduced that reduces the dimensionality of the data and discards the endmember matrix in the formulation by multiplying the rows of the measurement matrix  $\mathbf{Y}$  by the psuedo-inverse of  $\mathbf{E}$ ,  $(\mathbf{E}^\dagger)^\top = ((\mathbf{E}^\top \mathbf{E})^{-1} \mathbf{E}^\top)^\top$ :

$$\mathbf{Y} (\mathbf{E}^\dagger)^\top = \Phi \mathbf{A} + \mathbf{N} (\mathbf{E}^\dagger)^\top$$

where  $\mathbf{N}$  is the noise matrix. It was proved that this de-correlation scheme reduces the number of measurements needed for recovery of  $\mathbf{A}$  from the standard result  $m = \mathcal{O}(kn_\lambda \log(\frac{n_x n_y}{k}))$  to  $m = \mathcal{O}(k \log(\frac{n_e n_x n_y}{k}))$ , i.e., independent of the number  $n_\lambda$  of channels, provided that the sensing matrix  $\Phi$  satisfies an extended notion of the RIP [107]. To solve (4.2), a Parallel Proximal Algorithm [79], closely related to ADMM, was used (which consists of computing the proximity operator of the functions involved and averaging their results).

More recently, borrowing from the work in CS, in [108] the authors considered a regularized least-squares using two sparsity regularizers for recovering  $\mathbf{A}$  from blurred hyperspectral data with  $\Phi$  as a blurring operator:

$$\hat{\mathbf{A}} = \arg \min_{\mathbf{A}} \frac{1}{2} \|\mathbf{Y} - \Phi \mathbf{A} \mathbf{E}^\top\|_2^2 + \alpha_1 \|\mathbf{A}\|_{1,1} + \alpha_2 \sum_{i=1}^{n_e} \text{TV}(\mathbf{a}_i) \quad \text{subject to} \quad \mathbf{A} \succeq 0$$

where  $\|\mathbf{A}\|_{1,1}$  sums the absolute value of the elements of  $\mathbf{A}$  and enforces the abundance maps to be spatially sparse based on the assumption that only a few endmembers are associated with each pixel.

### 4.1.2 Blind Compressive Unmixing

When the endmembers of a scene are unknown, both  $\mathbf{A}$  and  $\mathbf{E}$  need to be estimated simultaneously from the given CS measurements  $\mathbf{y}$  of the scene. A direct approach is then to minimize the Euclidean distance between  $\mathbf{y}$  and  $\mathcal{A}(\mathbf{AE}^\top)$  with respect to both  $\mathbf{A}$  and  $\mathbf{E}$ , i.e.,

$$(\hat{\mathbf{A}}, \hat{\mathbf{E}}) = \arg \min_{\mathbf{A}, \mathbf{E}} \frac{1}{2} \|\mathbf{y} - \mathcal{A}(\mathbf{AE}^\top)\|_2^2 \quad \text{subject to} \quad \begin{aligned} \sum_{i=1}^{n_e} \mathbf{a}_i &= \mathbf{1} \\ \mathbf{A} &\succeq 0 \end{aligned} \quad (4.4)$$

However since the objective is non-convex, it may contain several local minima and thus in order to obtain solutions of physical significance (relating to real-world materials' spectra), additional constraints are usually enforced to limit the degrees of freedom.

In [104], the authors tackled the problem by posing two additional constraints: (i) the abundance maps (columns of  $\mathbf{A}$ ) are sparse in the 2D spatial wavelet basis (i.e.,  $\mathbf{A} = \Psi_{2D}\Lambda$  where columns of  $\Lambda = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_{n_e}]$  are sparse), and (ii) the abundance maps are disjoint, i.e., each spatial pixel is made up of only one endmember:

$$\begin{aligned} & \sum_{i=1}^{n_e} \mathbf{a}_i = \mathbf{1} \\ (\hat{\Lambda}, \hat{\mathbf{E}}) = \arg \min_{\Lambda, \mathbf{E}} \frac{1}{2} \|\mathbf{Y} - \Phi(\underbrace{\Psi_{2D}\Lambda}_{\mathbf{A}} \mathbf{E}^\top)\|_F^2 & \quad \text{subject to} \quad \begin{aligned} \mathbf{A}, \mathbf{E} &\succeq 0 \\ \|\boldsymbol{\alpha}_i\|_0 &\leq K, \forall i = 1, \dots, n_e \end{aligned} \\ & \text{Offdiag}(\Lambda \Lambda^\top) = \mathbf{0} \end{aligned} \quad (4.5)$$

where elements of  $\Phi$  are i.i.d. Gaussian. The  $l_0$  “norm” constraint imposes the wavelet coefficients to be  $K$ -sparse and the operator Offdiag returns the off-diagonal elements. The last constraint therefore forces the wavelet coefficients between the abundance maps to be orthogonal which implies abundance maps disjointness. To solve (4.5), an alternating minimization algorithm was proposed that involves approximating  $\Lambda$  with a fixed  $\hat{\mathbf{E}}$  (the current estimate of  $\mathbf{E}$ ) and updating  $\mathbf{E}$  for a given  $\hat{\Lambda}$  (the new estimate of  $\Lambda$ ). To approximate  $\Lambda$ , an iterative gradient projection algorithm was used where the constraints of (4.5) are imposed sequentially after a gradient descent step at each iteration. While updating  $\mathbf{E}$  is achieved by directly minimizing the objective with respect to  $\mathbf{E}$  giving  $\hat{\mathbf{E}} = \mathbf{Y}^\top (\Phi \mathbf{A}) ((\Phi \mathbf{A})^\top (\Phi \mathbf{A}))^{-1}$

and setting the non-positive values to zero to impose the non-negativity constraint.

The work of [105] proposed to recover both  $\mathbf{A}$  and  $\mathbf{E}$  from CASSI measurements (Section 2.2.3) of the HSI by solving the regularized problem

$$(\hat{\mathbf{A}}, \hat{\mathbf{E}}) = \arg \min_{\mathbf{A}, \mathbf{E}} \frac{1}{2} \|\mathbf{y} - \mathcal{A}(\mathbf{A}\mathbf{E}^\top)\|_2^2 + \alpha \sum_{i=1}^{n_e} \text{TV}(\mathbf{a}_i) + \beta \|\mathbf{E}\|_F^2 \quad \text{subject to } \mathbf{E} \succeq 0 \quad (4.6)$$

where  $\alpha, \beta$  are regularization parameters and  $\|\mathbf{E}\|_F^2 = \sum_{i=1}^{n_e} \|\mathbf{e}_i\|_2^2$  is the Tikhonov regularizer for reducing the effect of noise. However, as noted by the authors, the proposed solution is only ideal for situations where the endmembers lie in spatially homogenous areas, i.e., each pixel contains only one endmember. To solve (4.6), an alternating least-squares approach was presented that alternates between solving for  $\mathbf{A}$  (for a fixed current estimate  $\hat{\mathbf{E}}$ ) using an iterative TV minimization algorithm similar to the well-known Chambolle's projection algorithm [87], and minimizing  $\mathbf{E}$  (given the current estimate  $\hat{\mathbf{A}}$ ) by projecting the Tikhonov regularized psuedo-inverse solution onto the non-negative orthant.

## 4.2 A Blind Compressive HSI Unmixing Scheme

In this section, we explore and conduct a proof-of-concept study on a blind compressive hyperspectral unmixing method that does not rely on the dominant presence of pure end-members. Instead we are interested in developing a method that accurately unmixes scenes with highly mixed pixels as these scenarios are more likely to be observed in practice [7]. To validate the proposed method, experimental and numerical evidence will be provided using simulated CS measurements.

### 4.2.1 Additivity Regularized Least-Squares

We begin by considering a relaxed version of problem (4.4) that converts the STU constraint to a regularization term:

$$\arg \min_{\mathbf{A}, \mathbf{E}} \underbrace{\frac{1}{2} \|\mathbf{Y} - \Phi(\mathbf{A}\mathbf{E}^\top)\|_F^2 + \frac{\alpha_{STU}}{2} \|\mathbf{A}\mathbf{1}_{n_e} - \mathbf{1}_{n_x n_y}\|_2^2}_{f(\mathbf{A}, \mathbf{E})} \quad \text{subject to } \mathbf{A}, \mathbf{E} \succeq 0 \quad (4.7)$$

where a unique CS sensing matrix  $\Phi$  is used for each individual spatial band of the HSI, and  $\alpha_{STU}$  is the associated regularization parameter. The reason behind such a relaxation is that in practice the STU condition on  $\mathbf{A}$  does not usually hold strictly due to imprecisions and noise from various stages of the acquisition. By imposing the STU as a regularization term, we account for these non-idealities in the model by allowing slight violation in the STU requirement in  $\mathbf{A}$ .

It is well known that the decomposition of  $\mathbf{X} = \mathbf{AE}^\top$  under the linear mixing model is non-unique. This is because

- (i) algebraically speaking, for a carefully chosen square invertible matrix  $\mathbf{D} \in \mathbb{R}^{n_e \times n_e}$  it can be easily seen that  $(\mathbf{AD}^{-1})(\mathbf{DE}^\top)$  can also be a decomposition of  $\mathbf{X}$  provided  $\mathbf{AD}^{-1}, \mathbf{DE}^\top \succeq 0$  and  $\mathbf{AD}^{-1}\mathbf{1}_{n_e} \approx \mathbf{1}_{n_x n_y}$ , and
- (ii) geometrically speaking, the combined effect of the STU and non-negativity constraints means each pixel of the HSI lies in an  $(n_e - 1)$ -dimensional simplex with  $n_e$  vertices corresponding to the endmembers as illustrated in Figure 1.5. It is obvious from Figure 1.5 that any simplex with  $n_e$  vertices that encloses the data will also be a solution.

With further complication from the ill-posed nature of the under-determined system in the CS scheme, one can only anticipate that minimizing the representation error in the linear mixing model (4.7) is unlikely to be sufficient in obtaining meaningful results, and this is indeed the case as shown by the results in this section. Nonetheless, we begin by developing an algorithm for solving (4.7) which will form the basis for the remaining sections.

In order to solve (4.7), while it may look logical to adopt the standard alternating minimization strategy and alternately minimize the objective of (4.7) with respect to  $\mathbf{A}$  then  $\mathbf{E}$  while keeping the other variable fixed, one can easily check that minimizing with respect to  $\mathbf{A}$  results in the Sylvester equation:

$$\Phi^\top \Phi \mathbf{AE}^\top \mathbf{E} + \alpha_{STU} \mathbf{A} \mathbf{1}_{n_e} \mathbf{1}_{n_e}^\top = \Phi \mathbf{Y} \mathbf{E} + \mathbf{1}_{n_x n_y} \mathbf{1}_{n_e}^\top$$

whose solution can be shown to be<sup>1</sup>

$$\text{vec}(\mathbf{A}) = ((\mathbf{E}^\top \mathbf{E} \otimes \Phi^\top \Phi) + (\mathbf{1}_{n_e} \mathbf{1}_{n_e}^\top \otimes \alpha_{STU} \mathbf{I}_{n_x n_y}))^{-1} \text{vec}(\Phi \mathbf{Y} \mathbf{E} + \mathbf{1}_{n_x n_y} \mathbf{1}_{n_e}^\top)$$

Unfortunately, without efficient means of computing the matrix inversion, this is computationally unbearable due to the typical dimensions of hyperspectral data: a modest size of  $\mathbf{X} \in \mathbb{R}^{128 \times 128 \times 128}$  would require manipulating a dense  $128^2 n_e \times 128^2 n_e$  matrix. Furthermore, it is usually infeasible to explicitly form the CS measurement matrix  $\Phi$  for a problem of this size. To avoid the associated computations, we take advantage of the low-rank structure inherent in  $\mathbf{X}$  (under the linear mixing model) and decompose the hyperspectral data as a sum of products of the endmembers and their abundances:

$$\mathbf{X} = \mathbf{A} \mathbf{E}^\top = \sum_{i=1}^{n_e} \mathbf{a}_i \mathbf{e}_i^\top \quad (4.8)$$

and seek to update successively the columns of  $\mathbf{A}$  and  $\mathbf{E}$  alternately. This amounts to approximately solving the minimization problem with respect to  $\mathbf{A}$  and  $\mathbf{E}$  with a single complete round of a block-coordinate descent method with  $n_e$  blocks corresponding to the columns of  $\mathbf{A}$  and  $\mathbf{E}$ .

Formally, we approximate the minimization with respect to  $\mathbf{A}$  by sequentially updating each  $\{\mathbf{a}_i\}_{i=1}^{n_e}$ . Substituting (4.8) into (4.7) and fixing all variables but a single column  $\mathbf{a}_i$  of  $\mathbf{A}$ , the objective  $f(\mathbf{A}, \mathbf{E})$  can be re-written as

$$f(\mathbf{A}, \mathbf{E}) = \frac{1}{2} \left\| \mathbf{Y} - \underbrace{\Phi \sum_{j \neq i}^{n_e} \mathbf{a}_j \mathbf{e}_j^\top - \Phi \mathbf{a}_i \mathbf{e}_i^\top}_{\tilde{\mathbf{Y}}} \right\|_F^2 + \frac{\alpha_{STU}}{2} \left\| \mathbf{a}_i + \left( \sum_{j \neq i}^{n_e} \mathbf{a}_j - \mathbf{1}_{n_x n_y} \right) \right\|_2^2 \quad (4.9)$$

We then minimize (4.9) with respect to  $\mathbf{a}_i$ , and direct calculation gives the update equation:

$$\mathbf{a}_i = \left( \|\mathbf{e}_i\|_2^2 \Phi^\top \Phi + \alpha_{STU} \mathbf{I}_{n_x n_y} \right)^{-1} \left( \Phi^\top \tilde{\mathbf{Y}} \mathbf{e}_i - \alpha_{STU} \left( \sum_{j \neq i}^{n_e} \mathbf{a}_j - \mathbf{1}_{n_x n_y} \right) \right) \quad (4.10)$$

---

<sup>1</sup>Take the vec operator on both side of the equation and use the identity  $\text{vec}(\mathbf{AXB}) = (\mathbf{B}^\top \otimes \mathbf{A}) \text{vec}(\mathbf{X})$ . The result then follows from direct inversion.

As suggested in the preceding paragraph, we are interested in situations where it is infeasible to explicitly form  $\Phi$ , one might conclude that it is no more feasible to compute the inverse in (4.10) than before. Fortunately, it turns out we can sidestep this inversion by invoking the well-known matrix inversion lemma<sup>2</sup> to obtain:

$$(\|e_i\|_2^2 \Phi^\top \Phi + \alpha_{STU} I_{n_x n_y})^{-1} = \frac{1}{\alpha_{STU}} I_{n_x n_y} - \left( \frac{\|e_i\|_2^2}{\alpha_{STU}} \right)^2 \Phi^\top \left( I_m + \frac{\|e_i\|_2^2}{\alpha_{STU}} \Phi \Phi^\top \right)^{-1} \Phi$$

for  $\alpha_{STU} > 0$ . Noting that for a range of practical CS measurement matrices (e.g., partial orthogonal basis, partial random convolutions, etc.), the particular structure of  $\Phi$  satisfies  $\Phi \Phi^\top = I_m$ , this in turn simplifies the inversion to

$$(\|e_i\|_2^2 \Phi^\top \Phi + \alpha_{STU} I_{n_x n_y})^{-1} = \frac{1}{\alpha_{STU}} \left( I_{n_x n_y} - \frac{\|e_i\|_2^2}{\alpha_{STU} + \|e_i\|_2^2} \Phi^\top \Phi \right)$$

which gives an exact solution to (4.10) as the matrix-vector product

$$\mathbf{a}_i = \left( I_{n_x n_y} - \frac{\|e_i\|_2^2}{\alpha_{STU} + \|e_i\|_2^2} \Phi^\top \Phi \right) \left( \frac{1}{\alpha_{STU}} \Phi^\top \tilde{\mathbf{Y}} e_i - \left( \sum_{j \neq i}^{n_e} \mathbf{a}_j - \mathbf{1}_{n_x n_y} \right) \right) \quad (4.11)$$

By substituting  $\tilde{\mathbf{Y}} = \mathbf{Y} - \Phi \sum_{j \neq i}^{n_e} \mathbf{a}_j e_j^\top$  and expanding, it is possible to re-write (4.11) as

$$\mathbf{a}_i = \left( c \Phi^\top \mathbf{Y} e_i - \left( \sum_{j \neq i}^{n_e} \mathbf{a}_j - \mathbf{1}_{n_x n_y} \right) \right) - c \Phi^\top \Phi \left( \sum_{j \neq i}^{n_e} \mathbf{a}_j (e_j^\top e_i) - (e_i^\top e_i) \left( \sum_{j \neq i}^{n_e} \mathbf{a}_j - \mathbf{1}_{n_x n_y} \right) \right)$$

where  $c = (\alpha_{STU} + \|e_i\|_2^2)^{-1}$ . Since  $\mathbf{E}$  is fixed (during the  $\mathbf{A}$  update), we can therefore pre-compute  $\mathbf{R} = \Phi^\top \mathbf{Y} \mathbf{E}^{(k)\top}$  and  $\mathbf{S} = \mathbf{E}^{(k)\top} \mathbf{E}^{(k)}$  and update  $\{\mathbf{a}_i\}_{i=1}^{n_e}$  as follows:

$$\mathbf{a}_i^{(k+1)} = (c \mathbf{R}_i - \mathbf{q}) - c \Phi^\top \Phi \left( \left( \sum_{j=1}^{i-1} \mathbf{S}_{j,i} \mathbf{a}_j^{(k+1)} + \sum_{j=i+1}^{n_e} \mathbf{S}_{j,i} \mathbf{a}_j^{(k)} \right) - \mathbf{S}_{i,i} \mathbf{q} \right) \quad (4.12)$$

---

<sup>2</sup>For matrices  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$ ,  $\mathbf{D}$  of correct sizes, we have  $(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{D}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{D}\mathbf{A}^{-1}$ .

where

$$c = (\alpha_{STU} + \mathbf{S}_{i,i})^{-1}, \quad \mathbf{q} = \left( \sum_{j=1}^{i-1} \mathbf{a}_j^{(k+1)} + \sum_{j=i+1}^{n_e} \mathbf{a}_j^{(k)} \right) - \mathbf{1}_{n_x n_y}$$

The non-negativity constraint is then imposed by projecting the minimizer (4.12) onto the non-negative orthant. In essence, we are updating  $\mathbf{A}$  via an inexact block-coordinate descent of  $n_e$  blocks with each block computed using a projected Newton's method (see Remark 3; below).

The minimization of  $f(\mathbf{A}, \mathbf{E})$  with respect to  $\mathbf{E}$  on the other hand is computed by minimizing the following equivalent objective (after omitting the STU term which is independent of  $\mathbf{E}$ ):

$$\frac{1}{2} \|\mathbf{Y} - \Phi(\mathbf{A}\mathbf{E}^\top)\|_F^2 = \frac{1}{2} \|\mathbf{Y}^\top - \mathbf{E}\mathbf{A}^\top\Phi^\top\|_F^2 = \frac{1}{2} \|\mathbf{Y}^\top - \sum_{j \neq i}^{n_e} \mathbf{e}_j(\Phi\mathbf{a}_j)^\top - \mathbf{e}_i(\Phi\mathbf{a}_i)^\top\|_F^2 \quad (4.13)$$

sequentially with respect to each  $\{\mathbf{e}_i\}_{i=1}^{n_e}$ . Standard calculation reveals the update equation:

$$\mathbf{e}_i = \frac{1}{\|\Phi\mathbf{a}_i\|_2^2} \left( \mathbf{Y}^\top(\Phi\mathbf{a}_i) - \sum_{j \neq i}^{n_e} \mathbf{e}_j(\Phi\mathbf{a}_j)^\top(\Phi\mathbf{a}_j) \right) \quad (4.14)$$

Since  $\mathbf{A}$  is fixed, we can therefore pre-compute  $\mathbf{R} = \mathbf{Y}^\top\Phi\mathbf{A}^{(k+1)}$  and  $\mathbf{S} = (\Phi\mathbf{A}^{(k+1)})^\top(\Phi\mathbf{A}^{(k+1)})$  and update  $\{\mathbf{e}_i\}_{i=1}^{n_e}$  as follows:

$$\mathbf{e}_i^{(k+1)} = \mathbf{S}_{i,i}^{-1} \left( \mathbf{R}_i - \left( \sum_{j=1}^{i-1} \mathbf{S}_{j,i} \mathbf{e}_j^{(k+1)} + \sum_{j=i+1}^{n_e} \mathbf{S}_{j,i} \mathbf{e}_j^{(k)} \right) \right) \quad (4.15)$$

As before, the minimizer (4.15) is then projected onto the non-negative orthant. However different to the  $\mathbf{A}$  update, this projection results in an *exact* block-coordinate descent for the update of  $\mathbf{E}$  (see Remark 2). The entire procedure is presented in Algorithm 4.1 for clarity.

**Remark 1.** *Although  $\Phi$  usually cannot be explicitly formed, Algorithm 4.1 only requires multiplication of  $\Phi$  and its transpose which, for most practical CS measurement matrices, can be computed efficiently via fast transforms.*

**Remark 2.** *The projection onto the non-negative orthant for  $\mathbf{e}_i$  (Step 12) results in the*

optimal minimizer to (4.7) with respect to the individual  $\mathbf{e}_i$ 's. This can be seen by noting that (4.13) can be further decomposed as

$$\frac{1}{2} \|\mathbf{Y}^\top - \sum_{j \neq i}^{n_e} \mathbf{e}_j (\Phi \mathbf{a}_j)^\top - \mathbf{e}_i (\Phi \mathbf{a}_i)^\top\|_F^2 = \sum_{l=1}^{n_\lambda} \|\mathbf{y}_l^\top - \sum_{j \neq i}^{n_e} (\mathbf{e}_j)_l (\Phi \mathbf{a}_j)^\top - (\mathbf{e}_i)_l (\Phi \mathbf{a}_i)^\top\|_F^2$$

where  $(\mathbf{e}_j)_l$  denotes the  $l$ th element of the  $j$ th endmember. That is, the objective can be decoupled into  $n_\lambda$  independent quadratic programs in one variable  $(\mathbf{e}_i)_l$ . Direct calculation shows the optimal solution  $(\mathbf{e}_i)_l^*$  of these sub-problems is given by

$$(\mathbf{e}_i)_l^* = \max \left( 0, \frac{\left( \mathbf{y}_l^\top - \sum_{j \neq i}^{n_e} (\mathbf{e}_j)_l (\Phi \mathbf{a}_j)^\top \right) (\Phi \mathbf{a}_i)}{\|\Phi \mathbf{a}_i\|_2^2} \right), \quad l = 1, \dots, n_\lambda$$

which is equivalent to projecting (4.14) onto the non-negative orthant.

**Remark 3.** The update for  $\mathbf{A}$  (Step 7 and 8) can be considered as a projected Newton's method [109] since (4.10) can be re-written as:

$$\begin{aligned} \mathbf{a}_i &= (\|\mathbf{e}_i\|_2^2 \Phi^\top \Phi + \alpha \mathbf{I}_{n_x n_y})^{-1} \left( \Phi^\top \left( \mathbf{Y} - \Phi \sum_{j \neq i}^{n_e} \mathbf{a}_j \mathbf{e}_j^\top \right) \mathbf{e}_i - \alpha \left( \sum_{j \neq i}^{n_e} \mathbf{a}_j - \mathbf{1}_{n_x n_y} \right) \right) \\ &= (\|\mathbf{e}_i\|_2^2 \Phi^\top \Phi + \alpha \mathbf{I})^{-1} \left( \Phi^\top \left( \mathbf{Y} - \Phi \left( \sum_{j=1}^{n_e} \mathbf{a}_j \mathbf{e}_j^\top - \mathbf{a}_i \mathbf{e}_i^\top \right) \right) \mathbf{e}_i - \alpha \left( \sum_{j=1}^{n_e} \mathbf{a}_j - \mathbf{a}_i - \mathbf{1} \right) \right) \\ &= (\|\mathbf{e}_i\|_2^2 \Phi^\top \Phi + \alpha \mathbf{I})^{-1} \left( (\|\mathbf{e}_i\|_2^2 \Phi^\top \Phi + \alpha \mathbf{I}) \mathbf{a}_i + \Phi^\top \left( \mathbf{Y} - \Phi \sum_{j=1}^{n_e} \mathbf{a}_j \mathbf{e}_j^\top \right) \mathbf{e}_i - \alpha \left( \sum_{j=1}^{n_e} \mathbf{a}_j - \mathbf{1} \right) \right) \\ &= \mathbf{a}_i - \underbrace{(\|\mathbf{e}_i\|_2^2 \Phi^\top \Phi + \alpha \mathbf{I})^{-1}}_{(\nabla_{\mathbf{a}_i}^2 f(\mathbf{A}, \mathbf{E}))^{-1}} \underbrace{\left( -\Phi^\top \left( \mathbf{Y} - \Phi \sum_{j=1}^{n_e} \mathbf{a}_j \mathbf{e}_j^\top \right) \mathbf{e}_i + \alpha \left( \sum_{j=1}^{n_e} \mathbf{a}_j - \mathbf{1} \right) \right)}_{\nabla f(\mathbf{A}, \mathbf{E})} \end{aligned}$$

where  $\nabla_{\mathbf{a}_i}^2 f$  denotes the Hessian of  $f$  with respect to  $\mathbf{a}_i$ . Thus in essence, Step 7 and 8 of Algorithm 4.1 together performs an exact Newton step followed by a projection.

**Algorithm 4.1** Algorithm for solving (4.7)

---

```

1: Set  $k = 0$ ,  $n_e$ ,  $\mathbf{A}_0$ ,  $\mathbf{E}_0$ , and choose  $\alpha_{STU} > 0$ 
2: repeat
3:   Compute  $\mathbf{R} = \Phi^\top \mathbf{Y} \mathbf{E}^{(k)}$  and  $\mathbf{S} = \mathbf{E}^{(k)\top} \mathbf{E}^{(k)}$ 
4:   for  $i = 1, 2, \dots, n_e$  do
5:      $c = (\alpha_{STU} + \mathbf{S}_{i,i})^{-1}$ 
6:      $\mathbf{q} = \left( \sum_{j=1}^{i-1} \mathbf{a}_j^{(k+1)} + \sum_{j=i+1}^{n_e} \mathbf{a}_j^{(k)} \right) - \mathbf{1}_{n_x n_y}$ 
7:      $\mathbf{a}_i^{(k+1)} = (c \mathbf{R}_i - \mathbf{q}) - c \Phi^\top \Phi \left( \left( \sum_{j=1}^{i-1} \mathbf{S}_{j,i} \mathbf{a}_j^{(k+1)} + \sum_{j=i+1}^{n_e} \mathbf{S}_{j,i} \mathbf{a}_j^{(k)} \right) - \mathbf{S}_{i,i} \mathbf{q} \right)$ 
8:      $\mathbf{a}_i^{(k+1)} = \max \left( 0, \mathbf{a}_i^{(k+1)} \right)$ 
9:   end for
10:  Compute  $\mathbf{Q} = \Phi \mathbf{A}^{(k+1)}$ ,  $\mathbf{R} = \mathbf{Y}^\top \mathbf{Q}$ , and  $\mathbf{S} = \mathbf{Q}^\top \mathbf{Q}$ 
11:  for  $i = 1, 2, \dots, n_e$  do
12:     $\mathbf{e}_i^{(k+1)} = \mathbf{S}_{i,i}^{-1} \left( \mathbf{R}_i - \left( \sum_{j=1}^{i-1} \mathbf{S}_{j,i} \mathbf{e}_j^{(k+1)} + \sum_{j=i+1}^{n_e} \mathbf{S}_{j,i} \mathbf{e}_j^{(k)} \right) \right)$ 
13:     $\mathbf{e}_i^{(k+1)} = \max \left( 0, \mathbf{e}_i^{(k+1)} \right)$ 
14:  end for
15:   $k = k + 1$ 
16: until stopping criterion is satisfied

```

---

Similarly, (4.14) can be re-written as

$$\begin{aligned}
\mathbf{e}_i &= \frac{1}{\|\Phi \mathbf{a}_i\|_2^2} \left( \mathbf{Y}^\top - \sum_{j \neq i}^{n_e} \mathbf{e}_j (\Phi \mathbf{a}_j)^\top \right) (\Phi \mathbf{a}_i) \\
&= \frac{1}{\|\Phi \mathbf{a}_i\|_2^2} \left( \mathbf{Y}^\top - \sum_{j=1}^{n_e} \mathbf{e}_j (\Phi \mathbf{a}_i)^\top + \mathbf{e}_i (\Phi \mathbf{a}_i)^\top \right) (\Phi \mathbf{a}_i) \\
&= \mathbf{e}_i - \underbrace{\frac{1}{\|\Phi \mathbf{a}_i\|_2^2}}_{(\nabla_{\mathbf{e}_i}^2 f(\mathbf{A}, \mathbf{E}))^{-1}} \underbrace{\left( - \left( \mathbf{Y}^\top - \sum_{j=1}^{n_e} \mathbf{e}_j (\Phi \mathbf{a}_j)^\top \right) (\Phi \mathbf{a}_i) \right)}_{\nabla f(\mathbf{A}, \mathbf{E})}
\end{aligned}$$

revealing the update for  $\mathbf{E}$  (Step 12 and 13) is also a projected Newton's method.

**Remark 4.** The algorithm proposed here is, in essence, a variant of the Hierarchical Alternating Least Squares (HALS) [110–112]. However, HALS solves the problem (4.7) for the special case  $\Phi = \mathbf{I}_{n_x n_y}$  i.e., when the entire hyperspectral data  $\mathbf{X}$  is fully accessible. Under this setting, each  $n_e$  block update for  $\mathbf{A}$  in HALS is also exact whereas in Algorithm 4.1 they are not (due to the coupling of  $\Phi$ ). While this raises concerns regarding the convergence, our results indicate the proposed algorithm works well in practice, at least for the scenarios we have tested.

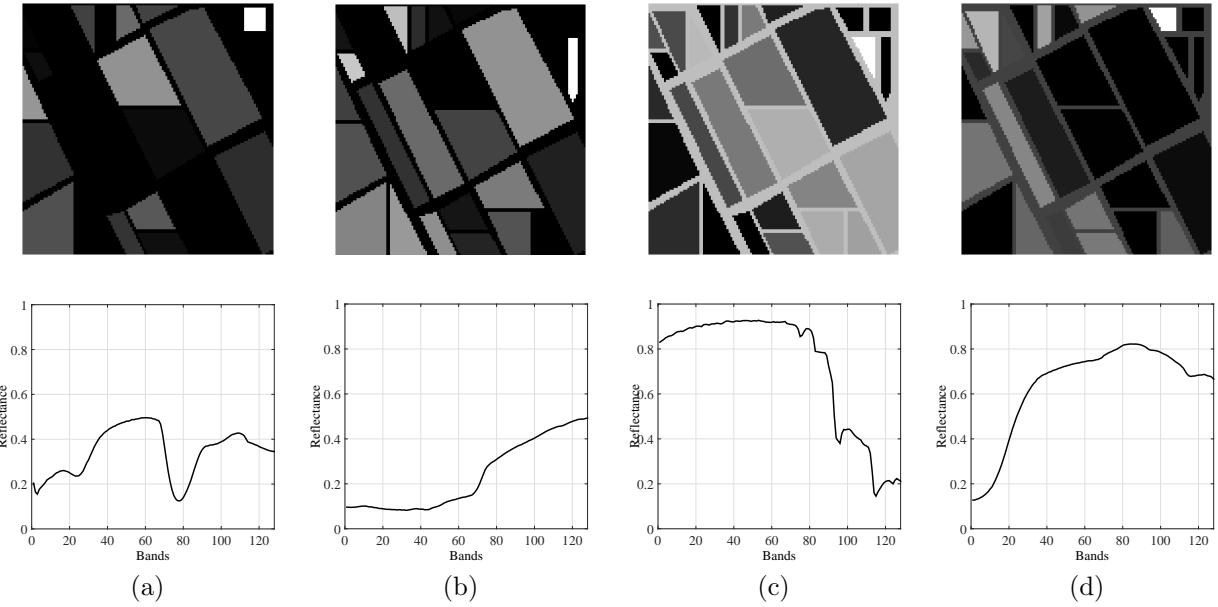


Figure 4.1: Simulated endmembers and their corresponding abundances.

#### 4.2.1.1 Results

To evaluate the performance of Algorithm 4.1, we selected four spectral signatures from the ASTER Spectral Library [54] constituting the columns of  $\mathbf{E}$ . The columns of  $\mathbf{A}$  are formed by vectorizing a cropped and modified  $128 \times 128$ -pixel ground truth map of the AVIRIS Salinas Valley data containing 30 distinct regions. The endmembers and their corresponding abundances are shown in Figure 4.1. For each distinct region shown in Figure 4.1, we randomly selected between two, three, or four endmembers for mixing with a maximal abundance value of the endmembers at each region being no greater than  $\rho \in (0, 1]$ , which gives us control over the mixing level of the data. We default the image to contain  $< 1\%$  pure pixels with the rest set using  $\rho = 0.8$ . Constructing the hyperspectral cube in such a synthetic but controllable way helps us with the assessment of the computed results.

For compressive measurements, we have used a randomly sub-sampled 2D discrete Fourier orthogonal basis as  $\Phi$  for computational simplicity. All tests, unless specified otherwise, were done using 25% measurements. White Gaussian noise with variance  $\sigma_N^2$  is then added to the measurements and we default the contamination to a SNR ( $= 10 \log(\|\mathbf{Y}\|_F^2 / mn_\lambda \sigma_N^2)$ ) of 40 dB.

To evaluate the performance, the standard Spectral Angle Distance (SAD) [113] between the

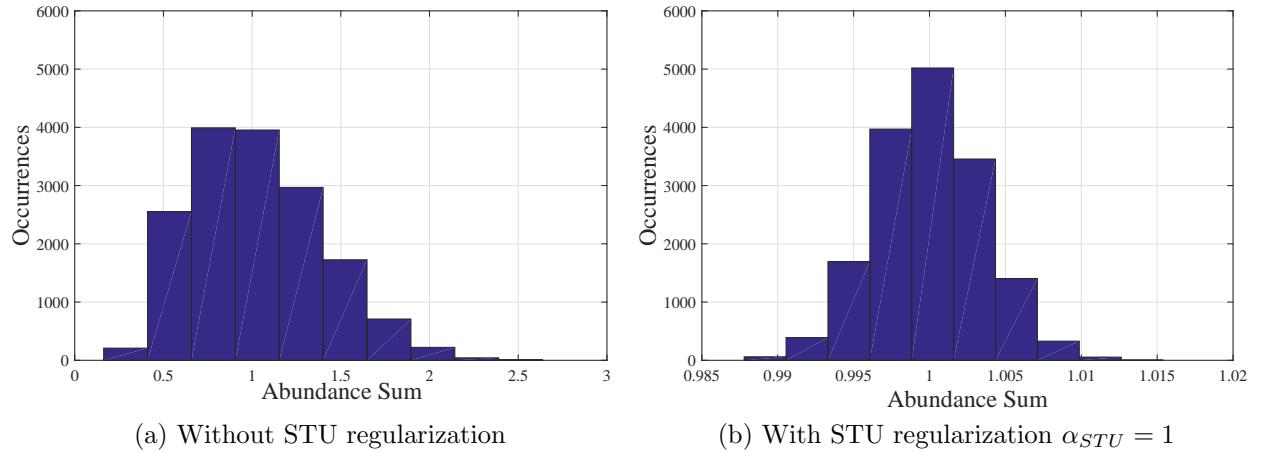


Figure 4.2: Abundance sum distribution with and without the STU constraint.

true endmember  $\mathbf{e}_j$  and its estimate  $\hat{\mathbf{e}}_j$  was used to compare their similarity:

$$\text{SAD}(\mathbf{e}_j, \hat{\mathbf{e}}_j) = \cos^{-1} \left( \frac{\mathbf{e}_j^\top \hat{\mathbf{e}}_j}{\|\mathbf{e}_j\|_2 \|\hat{\mathbf{e}}_j\|_2} \right).$$

In addition, the relative Mean-Squared Error (MSE) for the abundance maps, the endmembers, as well as the reconstructed hyperspectral cube (generated from the recovered  $A$  and  $E$ ) are also computed to quantify the unmixing result as a whole. These are defined, respectively, as

$$\text{AMSE} = \frac{\|\mathbf{A} - \hat{\mathbf{A}}\|_F^2}{\|\mathbf{A}\|_F^2}, \quad \text{EMSE} = \frac{\|\mathbf{E} - \hat{\mathbf{E}}\|_F^2}{\|\mathbf{E}\|_F^2}, \quad \text{and} \quad \text{XMSE} = \frac{\|\mathbf{X} - \hat{\mathbf{A}}\hat{\mathbf{E}}^\top\|_F^2}{\|\mathbf{X}\|_F^2}.$$

The first simulation, shown in Figure 4.2, shows the behaviour of the abundance sum distribution with and without the additivity regularization. Clearly, the additional regularization successfully constrains the abundance sum distribution to approximately unity, with a tolerance accounting for the existence of noise in the measurements. Note that in this simulation, the choice of the regularization parameter  $\alpha_{STU}$  was rather arbitrary, and the results shown are merely to demonstrate the effectiveness of the regularization. Suitable choice of  $\alpha_{STU}$  is discussed in the following sections.

Figure 4.3 shows the estimated endmembers and abundance maps using Algorithm 4.1 over 5 Monte-Carlo runs of the algorithm, each with elements of  $A_0$  and  $E_0$  initialized uniformly

Table 4.1: Unmixing metrics from 25% noisy CS measurements using Algorithm 4.1

EMSE	AMSE	XMSE	SAD1	SAD2	SAD3	SAD4
30.24%	42.19%	11.85%	19.43°	14.66°	3.90°	5.18°

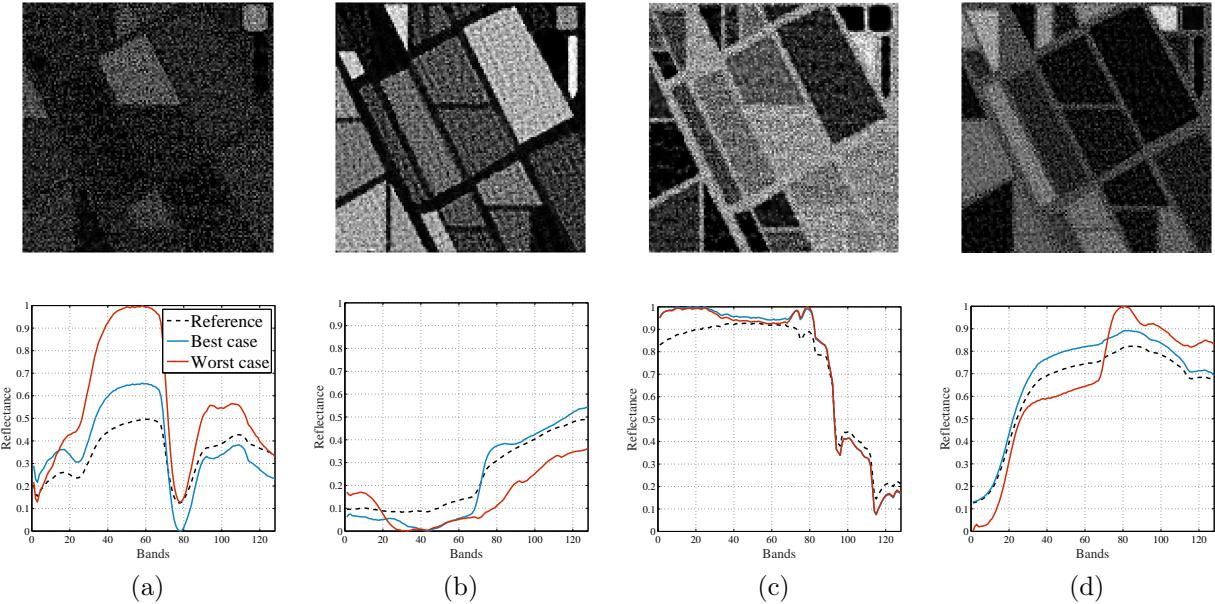


Figure 4.3: Reconstructed endmembers and abundances from 25% noisy CS measurements with SNR of 40 dB using Algorithm 4.1.

at random between 0 and 1. The best- and worst-cases are displayed for the endmembers while the abundance maps shown are the best-case scenario. Table 4.1 details the average unmixing results over the 5 runs. The noticeable variations in the reconstructed endmembers in Figure 4.3 emphasises the non-convexity nature of the problem, leading Algorithm 4.1 to converge to one of multiple stationary points. While visually, the abundance distribution seems to follow the ground truth despite the impulsive noise artefact, the significant errors in all quantitative metrics in Table 4.1 agree with our prior suspicion of the insufficiency in minimizing the representation error in the linear mixing model Equation 4.7.

### 4.2.2 Additivity-TV Regularized Least-Squares

Results from the previous section show that the recovered abundances and endmembers using Algorithm 4.1 do not accurately resemble the ground truths. This necessitates the use of prior knowledge about the underlying sources in order to recover them. In CS, TV regularization

puts emphasis on sparsity in the discrete gradients of images and is particularly suitable when the underlying image is piece-wise smooth. Since piece-wise smoothness in images implies similar material compositions over clustered regions, the fractional abundances of the materials (endmembers) involved are expected to preserve this structure and have strong resemblance to the underlying image. As seen from Section 4.1, this assumption has been widely accepted for the abundance fractions. Therefore, here we incorporate this spatial information into account and consider regularizing the objective (4.7) with the TV of the abundance maps as an additional prior and solve the following unmixing model:

$$\arg \min_{\mathbf{A}, \mathbf{E}} \frac{1}{2} \|\mathbf{Y} - \Phi(\mathbf{A}\mathbf{E}^\top)\|_F^2 + \frac{\alpha_{STU}}{2} \|\mathbf{A}\mathbf{1}_{n_e} - \mathbf{1}_{n_x n_y}\|_2^2 + \alpha_{TV} \sum_{j=1}^{n_e} \text{TV}(\mathbf{a}_j) \quad (4.16)$$

subject to       $\mathbf{A}, \mathbf{E} \succeq 0$

In order to solve (4.16) we follow the same algorithmic framework for solving (4.7). The minimization of (4.16) w.r.t.  $\{\mathbf{e}_i\}_{i=1}^{n_e}$  remains unchanged and is identical to minimizing (4.13) (corresponds to Step 10 to 14 of Algorithm 4.1). Since the TV is non-differentiable, the minimization w.r.t. each  $\{\mathbf{a}_i\}_{i=1}^{n_e}$  on the other hand no longer admits a closed-form solution and iterative methods must be used. Here, we use the proximal gradient method [114, 115] which involves taking a step along the negative gradient of the differentiable terms in (4.16):

$$f(\mathbf{a}_i) = \frac{1}{2} \left\| \left( \mathbf{Y} - \Phi \sum_{j \neq i}^{n_e} \mathbf{a}_j \mathbf{e}_j^\top \right) - \Phi \mathbf{a}_i \mathbf{e}_i^\top \right\|_2^2 + \frac{\alpha_{STU}}{2} \|\mathbf{a}_i + \sum_{j \neq i}^{n_e} \mathbf{a}_j - \mathbf{1}\|_2^2,$$

whose gradient is given by

$$\nabla f(\mathbf{a}_i) = -\Phi^\top \left( \mathbf{Y} - \Phi \sum_{j=1}^{n_e} \mathbf{a}_j \mathbf{e}_j^\top \right) \mathbf{e}_i + \alpha_{STU} \left( \sum_{j=1}^{n_e} \mathbf{a}_j - \mathbf{1} \right),$$

followed by the proximity operator (see Section 2.5.2) of the non-differentiable term (i.e., the TV) involved. The update can be derived as follows: we first upper-bound  $f(\mathbf{a}_i)$  by the

well-known surrogate for smooth functions with Lipschitz continuous gradient [116]:

$$\begin{aligned} f(\mathbf{a}_i) &\leq \underbrace{f(\boldsymbol{\theta}) + \nabla f(\boldsymbol{\theta})^\top (\mathbf{a}_i - \boldsymbol{\theta}) + \frac{L}{2} \|\mathbf{a}_i - \boldsymbol{\theta}\|_2^2}_{f(\mathbf{a}_i, \boldsymbol{\theta})} \\ &= \frac{L}{2} \|\mathbf{a}_i - \left( \boldsymbol{\theta} - \frac{1}{L} \nabla f(\boldsymbol{\theta}) \right)\|_2^2 + f(\boldsymbol{\theta}) - \frac{L}{2} \|\frac{1}{L} \nabla f(\boldsymbol{\theta})\|_2^2 \quad \forall \mathbf{a}_i, \boldsymbol{\theta} \end{aligned}$$

where  $L$  is the Lipschitz constant associated with  $\nabla f(\mathbf{a}_i)$  such that

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2 \leq L \|\mathbf{x} - \mathbf{y}\|_2 \quad \forall \mathbf{x}, \mathbf{y}.$$

By setting  $\boldsymbol{\theta} = \mathbf{a}_i^{(k)}$  and defining

$$\begin{aligned} \mathbf{a}_i^{(k+1)} &= \arg \min_{\mathbf{a}_i} f\left(\mathbf{a}_i, \mathbf{a}_i^{(k)}\right) + \alpha_{TV} \text{TV}(\mathbf{a}_i) \\ &= \arg \min_{\mathbf{a}_i} \frac{L/\alpha_{TV}}{2} \|\mathbf{a}_i - \left( \mathbf{a}_i^{(k)} - \frac{1}{L} \nabla f\left(\mathbf{a}_i^{(k)}\right) \right)\|_2^2 + \text{TV}(\mathbf{a}_i) \\ &= \text{prox}_{\text{TV}, \frac{L}{\alpha_{TV}}} \left( \mathbf{a}_i^{(k)} - \frac{1}{L} \nabla f\left(\mathbf{a}_i^{(k)}\right) \right) \end{aligned} \quad (4.17)$$

which corresponds to applying the TV proximity operator over the resulting descent step along the negative gradient of the least-squares and STU terms at the current iterate. One can easily verify that minimizing (4.16) with the surrogate  $f\left(\mathbf{a}_i, \mathbf{a}_i^{(k)}\right)$  in place of  $f(\mathbf{a}_i)$  has the property:

$$\begin{aligned} f\left(\mathbf{a}_i^{(k+1)}\right) + \alpha_{TV} \text{TV}\left(\mathbf{a}_i^{(k+1)}\right) &\leq f\left(\mathbf{a}_i^{(k+1)}, \mathbf{a}_i^{(k)}\right) + \alpha_{TV} \text{TV}\left(\mathbf{a}_i^{(k+1)}\right) \\ &\leq f\left(\mathbf{a}_i^{(k)}, \mathbf{a}_i^{(k)}\right) + \alpha_{TV} \text{TV}\left(\mathbf{a}_i^{(k)}\right) \\ &= f\left(\mathbf{a}_i^{(k)}\right) + \alpha_{TV} \text{TV}\left(\mathbf{a}_i^{(k)}\right) \end{aligned}$$

where the first inequality follows from the definition of  $f(\mathbf{a}_i, \boldsymbol{\theta})$  and the second inequality follows from the definition of  $\mathbf{a}_i^{(k+1)}$ . Updating  $\mathbf{a}_i$  using (4.17) will therefore result in a monotonically non-increasing sequence of  $\left\{ f\left(\mathbf{a}_i^{(k)}\right) + \alpha_{TV} \text{TV}\left(\mathbf{a}_i^{(k)}\right) \right\}_{k=1}$  and the sequence of  $\left\{ \mathbf{a}_i^{(k)} \right\}_{k=1}$  generated will therefore converge to a stationary point of (4.16). Thus, the

**Algorithm 4.2** Algorithm for solving (4.16)

---

```

1: Set  $k = 0$ ,  $n_e$ ,  $\mathbf{A}_0$ ,  $\mathbf{E}_0$ , and choose  $\alpha_{STU}$ ,  $\alpha_{TV} > 0$ 
2: repeat
3:   Compute  $\mathbf{R} = \Phi^\top \mathbf{Y} \mathbf{E}^{(k)}$  and  $\mathbf{S} = \mathbf{E}^{(k)\top} \mathbf{E}^{(k)}$ 
4:   for  $i = 1, 2, \dots, n_e$  do
5:      $L = \alpha_{STU} + \mathbf{S}_{i,i}$ 
6:      $\mathbf{q} = \left( \sum_{j=1}^{i-1} \mathbf{a}_j^{(k+1)} + \sum_{j=i+1}^{n_e} \mathbf{a}_j^{(k)} \right) - \mathbf{1}_{n_x n_y}$ 
7:      $\mathbf{a}_i^{(k+1)} = \mathbf{a}_i^{(k)} - L^{-1} \left( \Phi^\top \Phi \left( \sum_{j=1}^{i-1} \mathbf{S}_{j,i} \mathbf{a}_j^{(k+1)} + \sum_{j=i+1}^{n_e} \mathbf{S}_{j,i} \mathbf{a}_j^{(k)} \right) - \mathbf{R}_i + \alpha_{STU} \mathbf{q} \right)$ 
8:      $\mathbf{a}_i^{(k+1)} = \text{prox}_{\text{TV}, \frac{L}{\alpha_{TV}}} \left( \mathbf{a}_i^{(k+1)} \right)$ 
9:      $\mathbf{a}_i^{(k+1)} = \max \left( 0, \mathbf{a}_i^{(k+1)} \right)$ 
10:    end for
11:   Compute  $\mathbf{Q} = \Phi \mathbf{A}^{(k+1)}$ ,  $\mathbf{R} = \mathbf{Y}^\top \mathbf{Q}$ , and  $\mathbf{S} = \mathbf{Q}^\top \mathbf{Q}$ 
12:   for  $i = 1, 2, \dots, n_e$  do
13:      $\mathbf{e}_i^{(k+1)} = \mathbf{S}_{i,i}^{-1} \left( \mathbf{R}_i - \left( \sum_{j=1}^{i-1} \mathbf{S}_{j,i} \mathbf{e}_j^{(k+1)} + \sum_{j=i+1}^{n_e} \mathbf{S}_{j,i} \mathbf{e}_j^{(k)} \right) \right)$ 
14:      $\mathbf{e}_i^{(k+1)} = \max \left( 0, \mathbf{e}_i^{(k+1)} \right)$ 
15:   end for
16:    $k = k + 1$ 
17: until stopping criterion is satisfied

```

---

minimization w.r.t. each  $\{\mathbf{a}_i\}_{i=1}^{n_e}$  can be carried out by computing (4.17) repeatedly up to a convergence point. In our implementation however, we only compute (4.17) once for each  $\mathbf{a}_i$  for two reasons: (1) an optimal solution is actually not required for the minimization w.r.t. to each  $\mathbf{a}_i$  before the update for the others as seen in previous algorithms (see e.g., [117, 118]), and (2) we have observed, empirically, a faster convergence with just one step than with more steps. For the same reasons, the TV proximity operator is carried out using a single iteration of Chambolle's algorithm [87]. The resulting procedure is presented in Algorithm 4.2.

**Remark 5.** *The Lipschitz constant  $L$  of  $\nabla f$  for the differentiable terms (Step 5 of Algorithm 4.2) is calculated as follows:*

$$\begin{aligned}
\|\nabla f(\mathbf{x}) - \nabla(\mathbf{y})\|_2 &= \|\mathbf{e}_i^\top \mathbf{e}_i \Phi^\top \Phi (\mathbf{x} - \mathbf{y}) + \alpha_{STU} (\mathbf{x} - \mathbf{y})\|_2 \\
&= \|(\mathbf{e}_i^\top \mathbf{e}_i \Phi^\top \Phi + \alpha_{STU} \mathbf{I})(\mathbf{x} - \mathbf{y})\|_2 \\
&\leq \|(\mathbf{e}_i^\top \mathbf{e}_i \Phi^\top \Phi + \alpha_{STU} \mathbf{I})\|_2 \|\mathbf{x} - \mathbf{y}\|_2 \\
&= \underbrace{\lambda_{\max}(\mathbf{e}_i^\top \mathbf{e}_i \Phi^\top \Phi + \alpha_{STU} \mathbf{I})}_L \|\mathbf{x} - \mathbf{y}\|_2
\end{aligned}$$

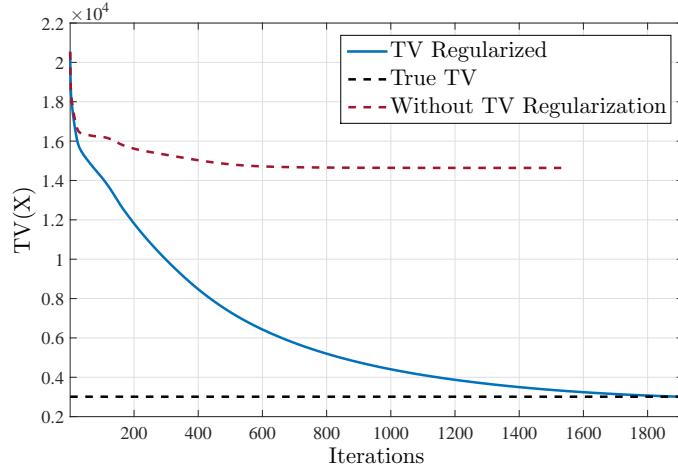


Figure 4.4: Functional value of  $\text{TV}(X)$  over iterations using Algorithm 4.1 (red) and Algorithm 4.2 (blue).

where  $\lambda_{\max}(\cdot)$  denotes the largest eigenvalue of the input. For a range of CS measurement matrices (e.g., partial orthogonal transforms, partial random convolutions, etc.), their rows are orthogonal, i.e.,  $\Phi\Phi^\top = \mathbf{I}_m$  and therefore  $\lambda_{\max}(\Phi^\top\Phi) = 1$ . For these CS matrices, we have

$$\begin{aligned}\lambda_{\max}(\mathbf{e}_i^\top \mathbf{e}_i \Phi^\top \Phi + \alpha_{STU} \mathbf{I}) &= \lambda_{\max}(\mathbf{e}_i^\top \mathbf{e}_i \Phi^\top \Phi) + \lambda_{\max}(\alpha_{STU} \mathbf{I}) \\ &= |\mathbf{e}_i^\top \mathbf{e}_i| \lambda_{\max}(\Phi^\top \Phi) + \alpha_{STU} \\ &= \mathbf{e}_i^\top \mathbf{e}_i + \alpha_{STU}\end{aligned}$$

On the other hand, when the dominant eigenvalue cannot be inferred analytically, a few iterations of the power method can be used for estimation. While in general, one could implement line search strategies for determining the step-size, they are usually costly, requiring multiple evaluations of  $\nabla f$  and  $\text{prox}_{\text{TV}, \frac{L}{\alpha_{TV}}}$  per iteration. Various acceleration strategies can also be implemented efficiently when  $L$  is known [114, 119].

#### 4.2.2.1 Results

In order to determine the regularization parameters in Equation 4.16, we performed preliminary Monte-Carlo simulations to find the values of  $\alpha_{STU}$  and  $\alpha_{TV}$  that gives, on average, the smallest error between the TV norm of the reconstituted hyperspectral cube,  $\text{TV}(\hat{\mathbf{A}}\hat{\mathbf{E}}^\top)$ , and

Table 4.2: Unmixing metrics from 25% noisy CS measurements using Algorithm 4.2

EMSE	AMSE	XMSE	SAD1	SAD2	SAD3	SAD4
25.38%	26.27%	1.03%	17.77°	12.28°	3.59°	3.55°

the ground truth,  $\text{TV}(\mathbf{X})$ . The chosen values were  $\alpha_{STU} = 1$  and  $\alpha_{TV} = 3 \times 10^{-4}$ . Figure 4.4 shows the TV values of the recovered hyperspectral data over iterations with and without the TV regularization. In both instances, the measurements and initial conditions were the same, thus the only variability between the two experiments is the regularization parameter  $\alpha_{TV}$ . The outcome suggests that the TV regularization successfully guides the algorithm toward a solution with the true TV value that would otherwise converge to a stationary point with a much higher TV value than the ground truth.

Figure 4.5 shows the recovered endmembers and abundances using Algorithm 4.2 and Table 4.2 details the unmixing quantitatively averaged over 5 runs of the algorithm. Clearly, taking into account of the abundance smoothness had significantly mitigated the noise artefact in the recovered abundances in Figure 4.5 compared to Figure 4.3. However, comparing with the ground truth, the difference between the abundance values and endmembers is still noticeable with average relative errors exceeding 25%. However, the average relative error of the reconstituted hyperspectral cube dropped to only 1.03%. Thus despite the recovered data cube matching the ground truth in terms of their TV norms, the ambiguity in estimated abundances and endmembers suggests that the reason for the failure is not the lack of CS measurements, but the lack of constraints defining the desired solution.

### 4.2.3 Joint Endmember-Abundance Regularized Least-Squares

As seen from the results in the previous section, the highly mixed nature of the data and the non-uniqueness of the factorization entail the solution obtained is not necessarily the desired one. This is revealed by the incorrectly identified endmembers and abundances (due to the non-convexity of the problem) whose product gives a well-estimated hyperspectral cube. Further constraints should therefore be imposed in order to narrow down the solution space. To achieve this, here we propose the use of additional priors on the endmembers

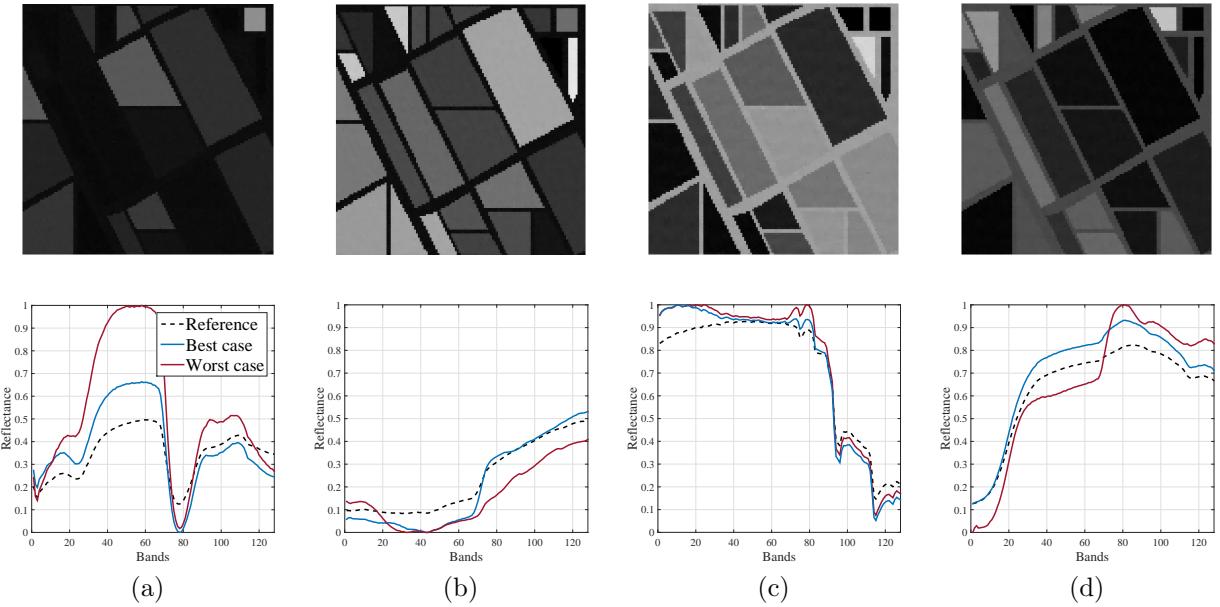


Figure 4.5: Reconstructed endmembers and abundances from 25% noisy CS measurements with SNR of 40 dB using Algorithm 4.2.

to integrate the spectral information inherent to the materials within the scene with the spatial arrangement of the HSI amenable to its image representation. By incorporating both abundance and endmember priors, the expression for the objective to be minimized now takes the form of

$$f(\mathbf{A}, \mathbf{E}) = \frac{1}{2} \|\mathbf{Y} - \Phi(\mathbf{A}\mathbf{E}^\top)\|_F^2 + \sum_{i=1}^{n_e} \alpha_i \phi_i(\mathbf{A}) + \sum_{i=1}^{n_e} \beta_i \phi_i(\mathbf{E}) \quad (4.18)$$

with  $\phi_1(\mathbf{A}) = \phi_{STU}(\mathbf{A}) = \|\sum_{j=1}^{n_e} \mathbf{a}_j - \mathbf{1}\|_2^2$  and  $\phi_2(\mathbf{A}) = \phi_{TV}(\mathbf{A}) = \sum_{j=1}^{n_e} \text{TV}(\mathbf{a}_j)$  as before.

We employ a minimum variance (MV) regularizer which encourages the *a priori* knowledge of the smooth nature in most real material spectra by minimizing the variance in each endmember:

$$\phi_1(\mathbf{E}) = \phi_{MV}(\mathbf{E}) = \frac{1}{2} \sum_{j=1}^{n_e} \|\mathbf{e}_j - \mathbf{1}_{n_\lambda} \frac{1}{n_\lambda} \sum_{l=1}^{n_\lambda} (\mathbf{e}_j)_l\|_2^2.$$

where  $\mathbf{1}_{n_\lambda}$  denotes a vector of ones in  $\mathbb{R}^{n_\lambda}$ . This type of regularizer has been previously used in [120]. In addition, we enforce a compactness measure, minimizing the distance to the centroid of the endmembers (MD), based on the observation that the endmembers tend to

lie on the vertices of the smallest simplex volume enclosing the hyperspectral dataset:

$$\phi_2(\mathbf{E}) = \phi_{MD}(\mathbf{E}) = \frac{1}{2} \sum_{j=1}^{n_e} \|\mathbf{e}_j - \frac{1}{n_e} \sum_{l=1}^{n_e} \mathbf{e}_l\|_2^2.$$

We thus bring together the intrinsic nature of each endmember and the geometric interdependency between them: two important priors which cannot be inferred by either regularizers alone.

Solving (4.18) with the specified regularizers amounts to incorporating the two endmember priors into Step 11 to 15 of Algorithm 4.2 while the other steps remain unchanged. Direct calculation shows that the minimization of (4.18) w.r.t. an individual  $\mathbf{e}_i$  results in the equation:

$$\begin{aligned} & \left[ \left( \|\Phi \mathbf{a}_i\|_2^2 + \beta_{MV} + \left( 1 - \frac{1}{n_e} \right) \beta_{MD} \right) \mathbf{I} - \frac{\beta_{MV}}{n_\lambda} \mathbf{1}_{n_\lambda} \mathbf{1}_{n_\lambda}^T \right] \mathbf{e}_i = \\ & \quad \left[ \mathbf{Y}^\top (\Phi \mathbf{a}_i) - \sum_{j \neq i}^{n_e} \mathbf{e}_j (\Phi \mathbf{a}_j)^\top (\Phi \mathbf{a}_i) + \frac{\beta_{MD}}{n_e} \sum_{j \neq i}^{n_e} \mathbf{e}_j \right] \end{aligned}$$

The matrix on the left hand side is a rank-one perturbation of the identity in the form  $a\mathbf{I} - b\mathbf{u}\mathbf{v}^\top$  where  $a = \|\Phi \mathbf{a}_i\|_2^2 + \beta_{MV} + (1 - n_e^{-1}) \beta_{MD}$ ,  $b = \beta_{MV}/n_\lambda$ , and  $\mathbf{u} = \mathbf{v} = \mathbf{1}_{n_\lambda}$ . An inverse therefore exists and is equal to  $c\mathbf{I} + d\mathbf{u}\mathbf{v}^\top$  where

$$c = \left( \|\Phi \mathbf{a}_i\|_2^2 + \beta_{MV} + \left( 1 - \frac{1}{n_e} \right) \beta_{MD} \right)^{-1}, \quad d = c\beta_{MV} \left( n_\lambda \left( \|\Phi \mathbf{a}_i\|_2^2 + \left( 1 - \frac{1}{n_e} \right) \beta_{MD} \right) \right)^{-1}.$$

The  $\mathbf{e}_i$ 's can thus be updated as:

$$\mathbf{e}_i = (c\mathbf{I} + d\mathbf{1}_{n_\lambda} \mathbf{1}_{n_\lambda}^T) \left[ \mathbf{Y}^\top (\Phi \mathbf{a}_i) - \sum_{j \neq i}^{n_e} \mathbf{e}_j (\Phi \mathbf{a}_j)^\top (\Phi \mathbf{a}_i) + \frac{\beta_{MD}}{n_e} \sum_{j \neq i}^{n_e} \mathbf{e}_j \right].$$

By pre-computing  $R = \mathbf{Y}^\top \Phi \mathbf{A}^{(k+1)}$  and  $\mathbf{S} = (\Phi \mathbf{A}^{(k+1)})^\top (\Phi \mathbf{A}^{(k+1)})$ , this gives the update equation

$$\mathbf{e}_i^{(k+1)} = c\boldsymbol{\theta} + d \left( \sum_{l=1}^{n_\lambda} \boldsymbol{\theta}_l \right) \mathbf{1}_{n_\lambda}$$

**Algorithm 4.3** Algorithm for solving (4.18)

---

```

1: Set  $k = 0$ ,  $n_e$ ,  $\mathbf{A}_0$ ,  $\mathbf{E}_0$ , and choose  $\alpha_{STU}$ ,  $\alpha_{TV}$ ,  $\beta_{MV}$ ,  $\beta_{MD} > 0$ 
2: repeat
3:   Compute  $\mathbf{R} = \Phi^\top \mathbf{Y} \mathbf{E}^{(k)}$  and  $\mathbf{S} = \mathbf{E}^{(k)\top} \mathbf{E}^{(k)}$ 
4:   for  $i = 1, 2, \dots, n_e$  do
5:      $L = \alpha_{STU} + \mathbf{S}_{i,i}$ 
6:      $\mathbf{q} = \left( \sum_{j=1}^{i-1} \mathbf{a}_j^{(k+1)} + \sum_{j=i+1}^{n_e} \mathbf{a}_j^{(k)} \right) - \mathbf{1}_{n_x n_y}$ 
7:      $\mathbf{a}_i^{(k+1)} = \mathbf{a}_i^{(k)} - L^{-1} \left( \Phi^\top \Phi \left( \sum_{j=1}^{i-1} \mathbf{S}_{j,i} \mathbf{a}_j^{(k+1)} + \sum_{j=i+1}^{n_e} \mathbf{S}_{j,i} \mathbf{a}_j^{(k)} \right) - \mathbf{R}_i + \alpha_{STU} \mathbf{q} \right)$ 
8:      $\mathbf{a}_i^{(k+1)} = \text{prox}_{TV, \frac{L}{\alpha_{TV}}} \left( \mathbf{a}_i^{(k+1)} \right)$ 
9:      $\mathbf{a}_i^{(k+1)} = \max \left( 0, \mathbf{a}_i^{(k+1)} \right)$ 
10:    end for
11:   Compute  $\mathbf{Q} = \Phi \mathbf{A}^{(k+1)}$ ,  $\mathbf{R} = \mathbf{Y}^\top \mathbf{Q}$ , and  $\mathbf{S} = \mathbf{Q}^\top \mathbf{Q}$ 
12:   for  $i = 1, 2, \dots, n_e$  do
13:      $c = (\mathbf{S}_{i,i} + \beta_{MV} + (1 - n_e^{-1}) \beta_{MD})^{-1}$ 
14:      $d = c \beta_{MV} (n_\lambda (\mathbf{S}_{i,i} + (1 - n_e^{-1}) \beta_{MD}))^{-1}$ 
15:      $\boldsymbol{\theta} = \mathbf{R}_i - \left( \sum_{j=1}^{i-1} \mathbf{S}_{j,i} \mathbf{e}_j^{(k+1)} + \sum_{j=i+1}^{n_e} \mathbf{S}_{j,i} \mathbf{e}_j^{(k)} \right) + \frac{\beta_{MD}}{n_e} \left( \sum_{j=1}^{i-1} \mathbf{e}_j^{(k+1)} + \sum_{j=i+1}^{n_e} \mathbf{e}_j^{(k)} \right)$ 
16:      $\mathbf{e}_i^{(k+1)} = c \boldsymbol{\theta} + d \left( \sum_{l=1}^{n_\lambda} \boldsymbol{\theta}_l \right) \mathbf{1}_{n_\lambda}$ 
17:      $\mathbf{e}_i^{(k+1)} = \max \left( 0, \mathbf{e}_i^{(k+1)} \right)$ 
18:   end for
19:    $k = k + 1$ 
20: until stopping criterion is satisfied

```

---

where  $c$ ,  $d$  are defined as above, and

$$\boldsymbol{\theta} = \mathbf{R}_i - \left( \sum_{j=1}^{i-1} \mathbf{S}_{j,i} \mathbf{e}_j^{(k+1)} + \sum_{j=i+1}^{n_e} \mathbf{S}_{j,i} \mathbf{e}_j^{(k)} \right) + \frac{\beta_{MD}}{n_e} \left( \sum_{j=1}^{i-1} \mathbf{e}_j^{(k+1)} + \sum_{j=i+1}^{n_e} \mathbf{e}_j^{(k)} \right)$$

The entire procedure is presented in Algorithm 4.3.

#### 4.2.3.1 Results

As before, the regularization parameters in Equation 4.18 are selected based on preliminary Monte-Carlo simulations to find the appropriate  $\alpha_{STU}, \alpha_{TV}, \beta_{MV}, \beta_{MD}$  that gives, on average, the smallest error in the functional value of the regularizers simultaneously between the estimate and the ground truth. Unfortunately, due to the large number of degrees of freedom and the lack of theoretical guidance, the procedure was carried out systematically, with a few trial-and-error attempts and the retained values in this scenario were  $\alpha_{STU} = 1$ ,  $\alpha_{TV} = 0.03$ ,

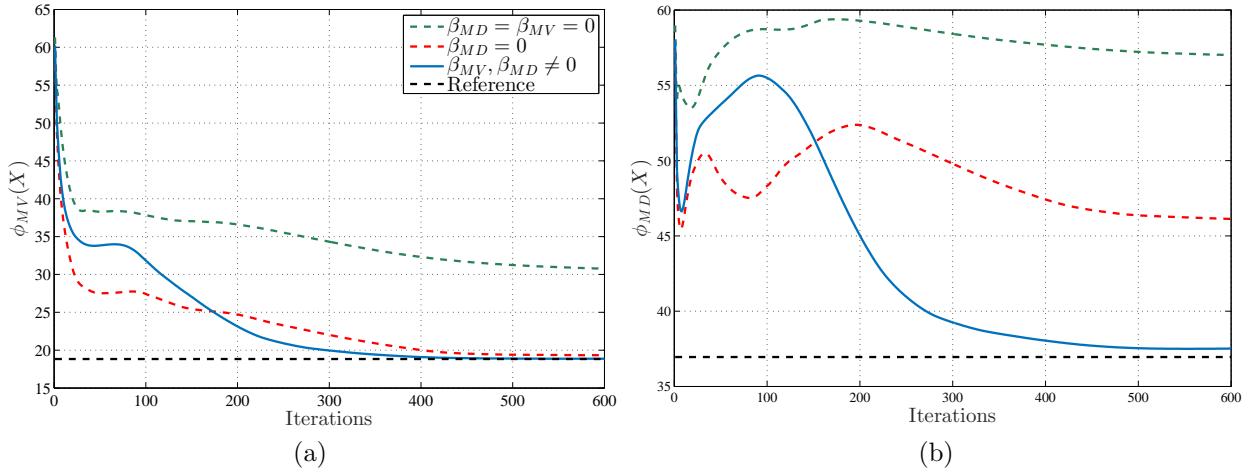


Figure 4.6: Functional value of (a) endmember variance and (b) endmember distance vs. iterations using Algorithm 4.2 (green), and Algorithm 4.3 (red and blue).

$\beta_{MV} = 2.5$ ,  $\beta_{MD} = 2$ . While the chosen parameters may not be optimal, they demonstrate, convincingly, the effectiveness of the proposed joint regularization. Figure 4.6 shows functional values of the two endmember regularizers over iterations using Algorithm 4.2 (green) and Algorithm 4.3 (blue). We also considered the case when only the minimum variance regularizer,  $\phi_{MV}$ , is in effect (red) with  $\alpha_{STU} = 1$ ,  $\alpha_{TV} = 0.2$ ,  $\beta_{MV} = 5$ ,  $\beta_{MD} = 0$  chosen experimentally. As expected, the additional regularization(s) successfully guides the algorithm to solutions with the desired properties exhibited by the ground truth.

Figure 4.7 demonstrates the quantitative improvement in combining the two endmember priors over each individual regularizer by plotting the unmixed EMSEs against the relative deviation from the optimal regularization parameters found experimentally to give the minimal MSE for (a)  $\beta_{MV} \neq 0, \beta_{MD} = 0$ , (b)  $\beta_{MV} = 0, \beta_{MD} \neq 0$ , and (c)  $\beta_{MV}, \beta_{MD} \neq 0$ . It is apparent that the joint regularization achieved better results than using either regularizers alone.

Figure 4.8 shows the unmixed results visually when only considering the minimum variance regularization, while Figure 4.9 considered both minimum variance and distance. In both cases, the recovered sources match faithfully with the ground truths with the joint regularization performing perceptibly and quantitatively (see Table 4.3) better. This improvement should not come as a surprise since in the absence of the minimum distance prior, the solution may converge to a larger simplex that encapsulates the hyperspectral data as suggested by

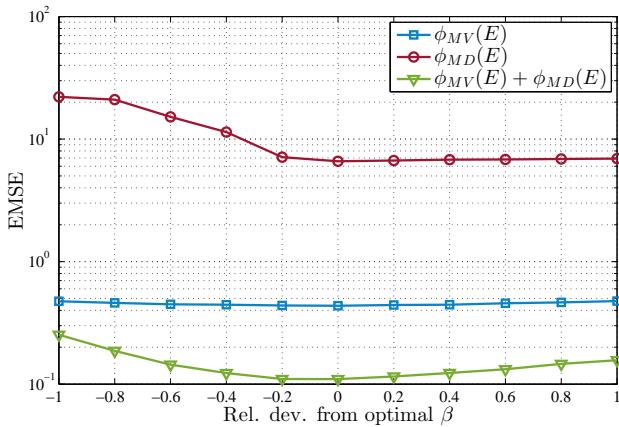


Figure 4.7: Unmixed EMSE from 25% noisy CS measurements as a function of the regularization parameter using Algorithm 4.3 with (a)  $\beta_{MD} = 0$  (blue); (b)  $\beta_{MV} = 0$  (red); and (c)  $\beta_{MV} \neq 0, \beta_{MD} \neq 0$  (green).

Table 4.3: Unmixing metrics from 25% noisy CS measurements using Algorithm 4.3

	EMSE	AMSE	XMSE	SAD1	SAD2	SAD3	SAD4
$\beta_{MV} = 5, \beta_{MD} = 0$	6.536%	9.690%	0.4780%	4.355°	3.018°	0.5682°	0.9565°
$\beta_{MV} = 2.5, \beta_{MD} = 2$	1.771%	5.764%	0.2623%	2.044°	0.5330°	0.2682°	0.1383°

the higher  $\phi_{MD}$  value seen in Figure 4.6(b) (red). Figure 4.10 reinforces this observation by plotting the evolution of the algorithm path on the first three principal components of the dataset  $\mathbf{X}$ . Figure 4.10(a) reveals that even though the minimizer without the minimum distance regularization correctly encloses the dataset, the resultant simplex is much larger than the ground truth. By incorporating the additional prior on the other hand, Figure 4.10(b) shows that the minimizer correctly converges towards the underlying endmembers.

In Figure 4.11 we computed the relative errors in the reconstruction as a function of the undersampling ratio. The result indicates that we can recover the endmembers and abundances with relatively few measurements: a measurement rate of 30% gives a relative error of no greater than 2% and 5% respectively, and achieves comparable quality to that if one were to sample the HSI fully. We note that with 100% measurement, the algorithm did not achieve exact identification of the underlying endmembers and abundances. This is due to (1) the regularization parameters not being selected optimally, (2) the early termination criterion limiting the numerical precision of the algorithm, (3) the presence of noise, and (4) the regularization terms provide *a priori* information but are by no means optimal. This reveals

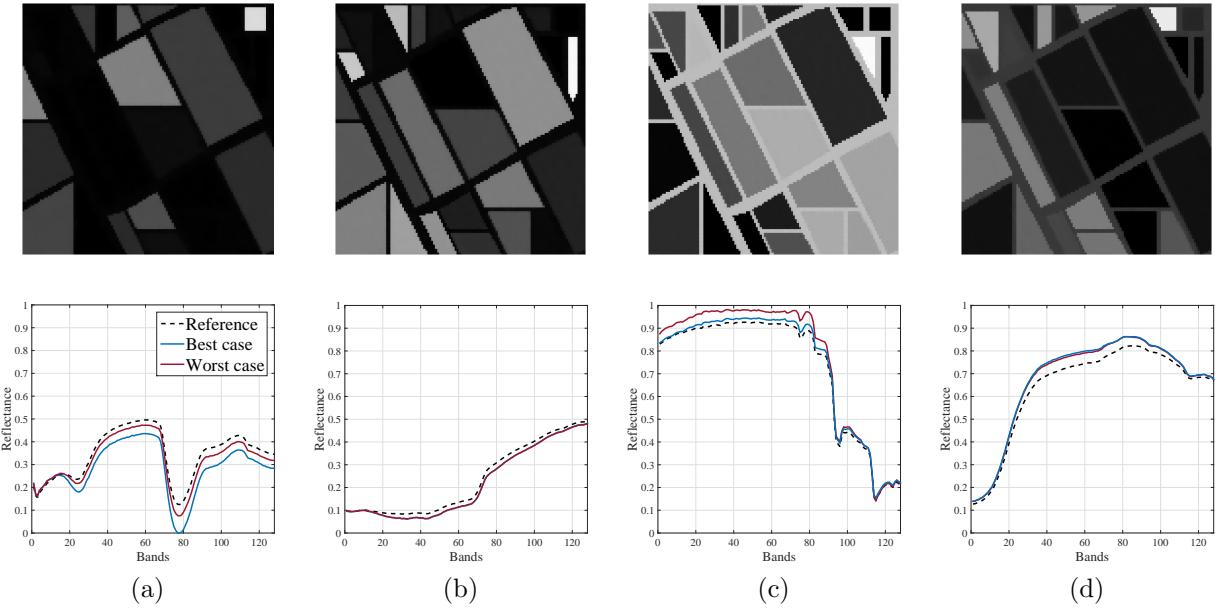


Figure 4.8: Reconstructed endmembers and abundances from 25% noisy CS measurements with SNR of 40 dB using Algorithm 4.3 with  $\beta_{MV} = 5, \beta_{MD} = 0$ .

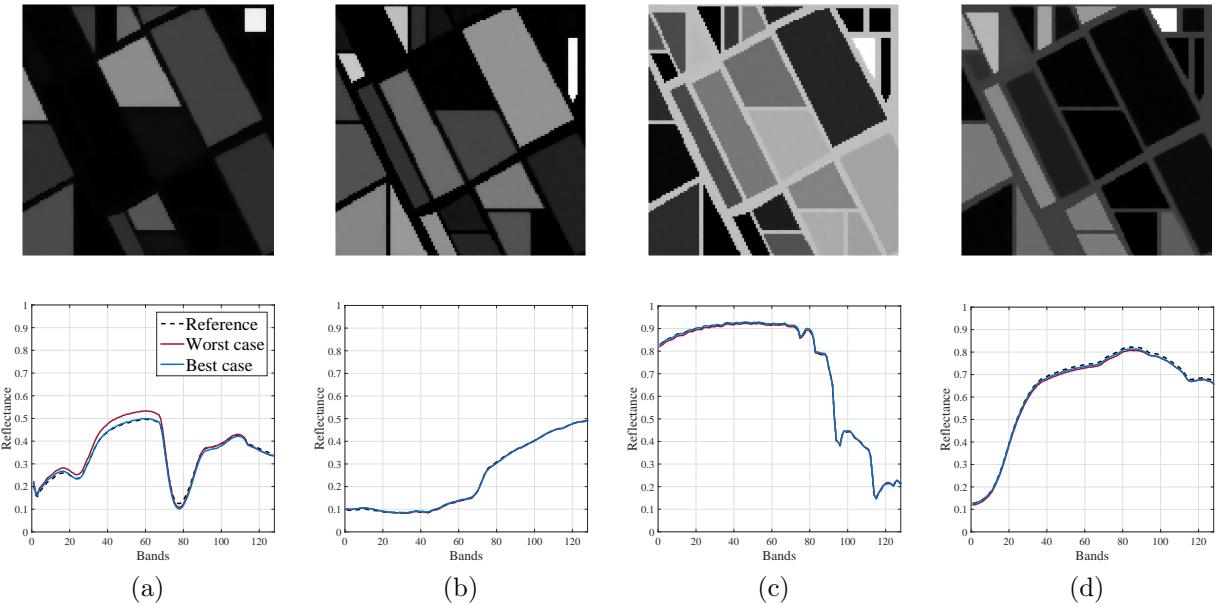


Figure 4.9: Reconstructed endmembers and abundances from 25% noisy CS measurements with SNR of 40 dB using Algorithm 4.3 with  $\beta_{MV} = 2.5, \beta_{MD} = 2$ .

that the number of CS measurements used for recovery does not necessarily correlate to the success of blind unmixing. However, by choosing the appropriate relevant priors, the number of measurements necessary to achieve similar performance may be significantly reduced.

Table 4.4 gives the unmixing results at different values of the mixing parameter  $\rho$ . It appears that the algorithm is particularly sensitive to this parameter, in particular, the relative errors

Table 4.4: Unmixing metrics from noisy CS measurements for different values of the mixing parameter  $\rho$ .

	<b>EMSE</b>	<b>AMSE</b>	<b>XMSE</b>	<b>SAD1</b>	<b>SAD2</b>	<b>SAD3</b>	<b>SAD4</b>
$\rho = 0.6$	3.214%	12.75%	0.8842%	4.367°	0.7417°	0.4125°	0.3994°
$\rho = 0.5$	6.347%	18.15%	1.227%	7.778°	0.4747°	0.2683°	0.7851°

in abundance increases significantly as the purity level decreases. Interestingly, the errors in the recovered endmembers are still relatively low as indicated by the EMSE and SADs. This seems to suggest that additional constraints on the abundances may be necessary to overcome the nonconvexity of the objective.

Lastly, we note that Algorithm 4.3 requires, as one of its inputs, an initial estimation of the number of endmembers,  $n_e$ , present within the hyperspectral data. Accurate estimation of this parameter is a delicate issue and merits a separate problem of its own as a pre-processing step (see [59]). Nonetheless, it is interesting to see the effect of an inaccurate estimate of  $n_e$  to the proposed method. In Figure 4.12 we plot the relative MSE of the reconstituted hyperspectral data for various values of  $n_e$ . The result shows that the minimum relative MSE is indeed attained at  $n_e = 4$ . When  $n_e < 4$ , endmembers are not enough to correctly describe the scene leading to an increase in error. Surprisingly the algorithm is not sensitive to overestimations and positively reveals the effectiveness of the joint regularization. Figure 4.13 shows recovered abundances and endmembers when  $n_e$  is overestimated by 2. The endmembers in excess (Figure 4.13(e) and Figure 4.13(f)) are encouraged to be flat (due to  $\phi_{MV}$ ) with their reflectance values near the average of all endmembers (due to  $\phi_{MD}$ ). This in turn limits the estimation errors on the endmembers that are actually present, and forces the contributions of excess endmembers to the scene to be negligible as shown by their corresponding abundances in Figure 4.13.

### 4.3 Conclusions

Practical applications of hyperspectral imaging usually require the identification of the end-members that constitute the scene. To achieve this from CS measurements of the scene, it

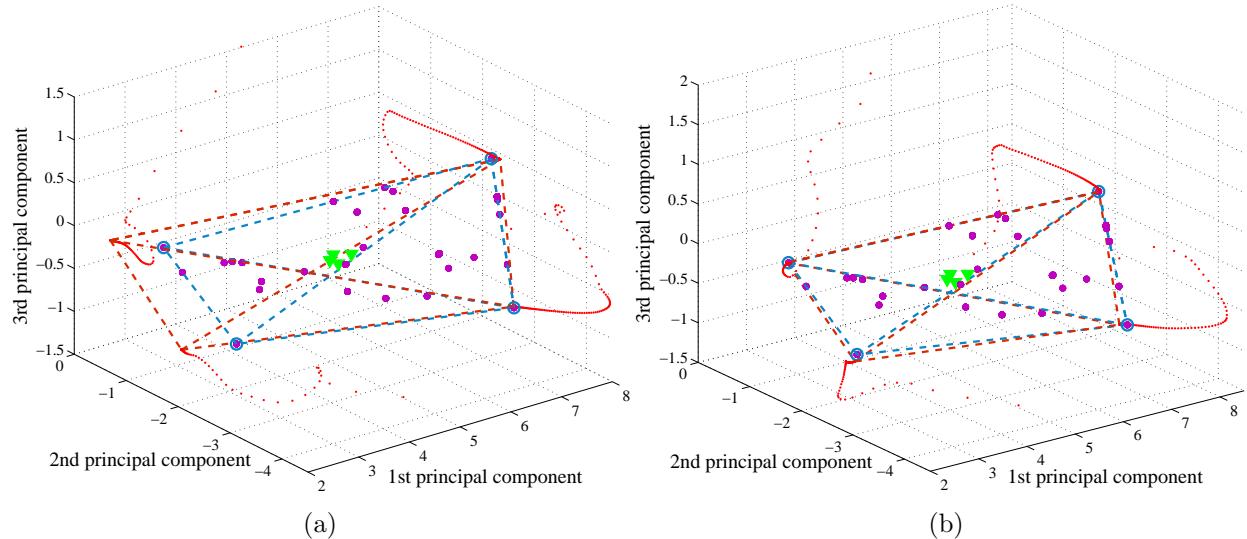


Figure 4.10: Evolution of the algorithm path (red) of Algorithm 4.3 for simulated data (purple) from random initialization (green). (a)  $\beta_{MV} = 5, \beta_{MD} = 0$ , (b)  $\beta_{MV} = 2.5, \beta_{MD} = 2$ .

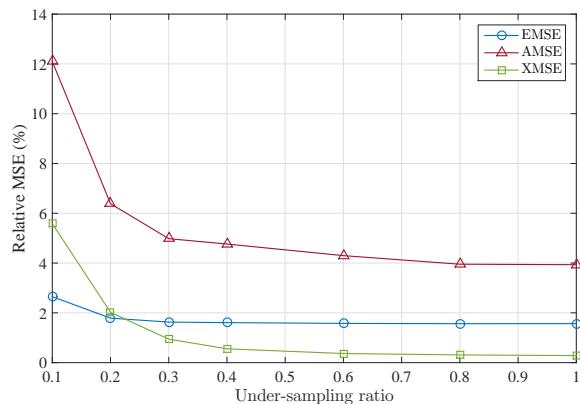


Figure 4.11: Relative MSE of recovered hyperspectral data at various sampling rates.

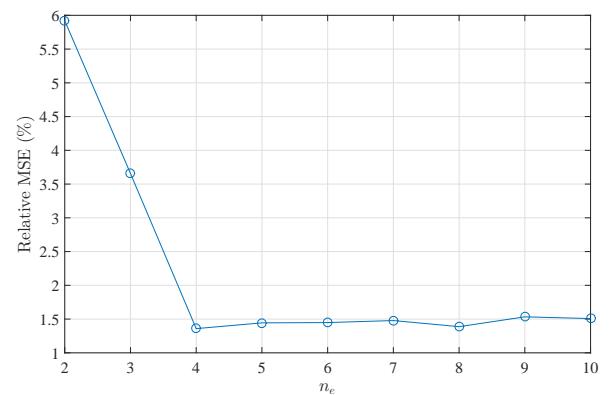


Figure 4.12: Relative MSE of recovered hyperspectral data for various  $n_e$ .

is logical to first recover the full hyperspectral image via existing CS recovery methods, and then perform spectral unmixing using well-established unmixing algorithms. However, inefficiencies in this paradigm naturally lead to a strong motivation for development of methods that determine the endmembers and their compositions directly via the CS measurements.

Existing methods usually make several assumptions about the underlying data in order to introduce convexity to an otherwise ill-posed nonconvex problem. These include (1) the dominant presence of pure endmembers within the hyperspectral data, and/or (2) the availability of standard libraries containing spectral signatures of the materials likely to be present within the scene. In practice however, it is unlikely for pure endmembers to be present predom-

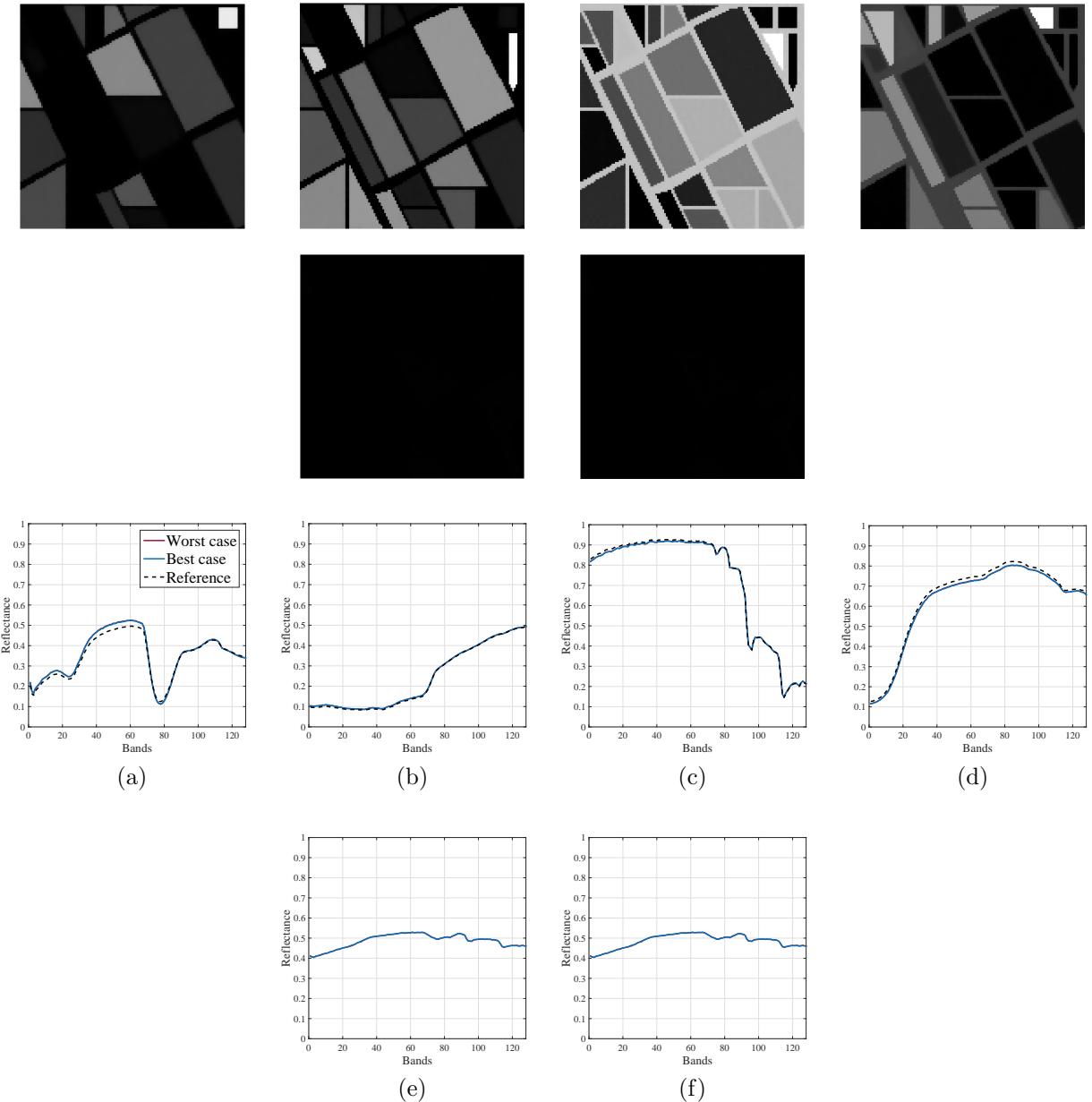


Figure 4.13: Reconstructed endmembers and abundances from CS measurements of a hyperspectral scene consisting of 4 endmembers with the number of endmembers input to Algorithm 4.3 overestimated by 2.

inantly due to limited spatial resolution and mixing phenomena. Furthermore, deviations from standard libraries will undoubtedly exist due to variability in environmental and temporal conditions between the captures.

Throughout this chapter, we have developed and conducted a proof-of-concept study on a blind compressive spectral unmixing technique that aims to directly determine the material spectra present and their relative abundances in a hyperspectral scene from its CS mea-

surements without the assumptions mentioned previously. The developed method brings together multiple important priors in the hyperspectral data by penalizing the TV norm of the abundances, the variance of the endmembers, as well as the geometric distances between them. The solution therefore simultaneously accounts for the internal characteristics within each material constituting the underlying data, and the external geometry between the materials under the linear mixing model. The optimization scheme is based on an alternating minimization strategy that successively updates the individual endmembers and abundances via block-coordinate descent using a projected Newton's method. Experimental results positively shows the effectiveness of the proposed method in unmixing highly mixed scenarios, subject to various parameters important to spectral unmixing.



# Chapter 5

## Future Work

A number of opportunities to expand on the current work were identified throughout the research process. This chapter outlines the various possible directions that can be pursued in continuation of the research contained within the previous chapters. These opportunities naturally fall under the three sub-modules investigated in this thesis that make up a compressive hyperspectral system: acquisition, recovery, and unmixing.

### 5.1 Acquisition

The multimodal acquisition framework detailed within this thesis was explored and implemented with the restriction that the total number of measurements made is divided equally among the utilized modes. Despite the superior reconstruction observed under such a configuration, we have not made any attempt to optimize the measurement allocations. While a uniform distribution may be logical in lieu of prior knowledge of the imaged scene, the noticeable difference in recovery performance when utilizing three modes instead of two observed in the results with real data (Figure 2.15) hints that further investigation into the effect of varying the distribution of measurements may be fruitful. The difference in performance observed between the modes in Figure 2.15 compared to the more uniform increment with the synthetic data in Figure 2.8 and Figure 2.10 suggests that the distribution among the modalities can potentially have a sizeable impact on the reconstruction quality depending

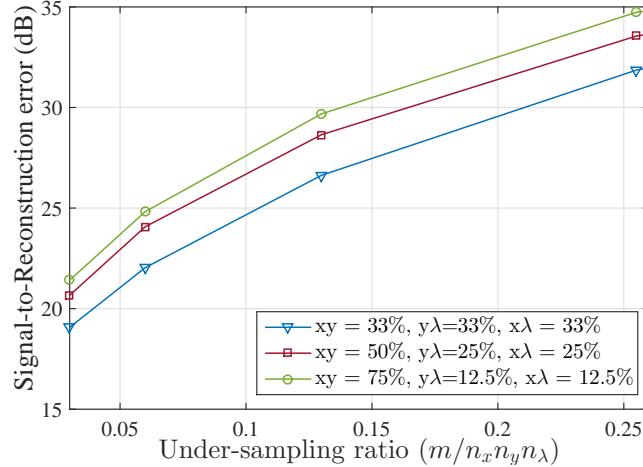


Figure 5.1: Reconstruction SRE vs. under-sampling ratio using three modes of measurements with various measurement allocations between the modes.

on the underlying data. Figure 5.1 shows some preliminary reconstruction results using the real data detailed in Section 2.6.2 with different measurement allocations. We have fixed the number of measurements along the  $x$ - $\lambda$  and  $y$ - $\lambda$  plane to be the same (for they can be argued to contain equally important information). The improvements observed by adaptively distributing the measurements (red and green) over the uniform distribution (blue) reinforce our contention. We speculate that such improvement occurs because the hyperspectral data is more sparse along the  $x$ - $\lambda$  and  $y$ - $\lambda$  imaging planes when compared to the spatial  $x$ - $y$  plane. Consequently, fewer measurements are necessary along the former planes for successful recovery and more measurements should be allocated to the latter in order to augment the performance.

The measurement matrices used in all simulations within this thesis were constructed using randomly sub-sampled discrete Fourier basis for computational convenience. While this could be achieved using specialized lenses and spatial light modulators (SLM) that perform the optical Fourier transforms and the appropriate filtering [43], in practice the DMD is a more tractable and accessible alternative. In conjunction with the proposed acquisition system shown in Figure 2.6 that uses DMD exclusively, a natural direction is therefore to investigate the performance of our proposed framework using a more practical measurement matrix that is implementable via the DMD such as the scrambled block Hadamard matrix [121]. Furthermore, the proposed system could be used to capture two of the modalities ( $x$ - $y$  and one of

$x\text{-}\lambda$  or  $y\text{-}\lambda$ ) simultaneously at the expense of forgoing a certain degree of randomness. Since the DMD D1 is used as both a measurement matrix (to effect the compressive measurement along the  $x\text{-}y$  plane) and a pushbroom scanner (to effect the measurement along the  $x\text{-}\lambda$  or  $y\text{-}\lambda$  plane), a concurrent acquisition necessitates rows (or columns) of D1 to be sequentially omitted from the Hadamard matrix construction. While we suspect such structured construction may naturally lead to an inferior performance [39], we believe the possible gain in temporal resolution and overall improvement in reconstruction quality over traditional methods is worthy of further investigation.

Lastly, even though the proposed multimodal framework has been empirically (and rather convincingly) validated using both synthetic and real hyperspectral data, a formal theoretical study is nevertheless lacking. The RIP is a convenient tool in CS theory: provided a matrix satisfies the RIP, a range of reconstruction error bounds can be established (subject to domain-specific optimizations). The RIP, however, is known to be too strict a property and is NP-hard to verify [122], making it far from ideal for construction or analysis of practical sensing matrices. RIP-based theoretical bounds also require prior knowledge of the sparsity of the input which, in practice, is often unknown. Some authors (e.g., [123, 124]) also point out that RIP is of little relevance in a wide range of imaging problems, as for realistic parameter values (i.e., problem resolution, sub-sampling percentage, and sparsity) the RIP usually does not hold, yet the measurement matrix has been experimentally shown to work well. Consequently, recent attention has shifted towards the study of “RIPless” theory based on the notion of isotropy and incoherence between the sensing and sparsity basis [25, 125, 126]. We therefore hope that the results shown in this thesis provides future researchers with convincing evidence for a thorough theoretical analysis into the proposed framework as well as stimulating further research into the practicality of such an acquisition regime.

## 5.2 Recovery

The work presented in both Chapter 2 and Chapter 3 relies heavily on the ADMM for solving the associated optimization problems. As seen in Section 2.5.1, the application of ADMM to imaging problems generally necessitates the use of variable splitting whereby terms in the objective are substituted with auxiliary variables in order to carry out the algorithm. Needless to say, any methods involving a change in variables will inevitably result in different algorithmic implementations for a fixed problem, contingent on the specific choices of what is substituted. Consequently, finding the optimal choices for an algorithmic framework that is computationally feasible and result in efficient iterates is an art in itself and deserves some special attention. While the focus of the work presented was not in this direction, during the research process we have noticed that certain substitutions can lead to more efficient iterations. We give a contrived example here representative of what we hope to be further investigated and applied to our work. Consider the unconstrained TV minimization problem in CS:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \alpha \|\mathbf{D}_{xy}\mathbf{x}\|_1. \quad (5.1)$$

where  $\mathbf{D}_{xy}$  is the anisotropic TV operator. Suppose we make two similar but subtly different change of variables for solving (5.1) using the ADMM algorithm (Algorithm 2.1):

$$\underbrace{\begin{bmatrix} \mathbf{z}^{(1)} \\ \mathbf{z}^{(2)} \end{bmatrix}}_{\mathbf{z}} = \underbrace{\begin{bmatrix} \mathbf{A} \\ \mathbf{D}_{xy} \end{bmatrix}}_{\mathbf{G}} \boldsymbol{\nu} \quad (5.2)$$

and

$$\begin{bmatrix} \mathbf{z}^{(1)} \\ \mathbf{z}^{(2)} \end{bmatrix} = \begin{bmatrix} \mathbf{A} \\ \mathbf{I} \end{bmatrix} \boldsymbol{\nu} \quad (5.3)$$

where  $\boldsymbol{\nu} = \mathbf{x}$ . The resulting algorithm for each mapping is shown, respectively, in Algorithm 5.1 and Algorithm 5.2. The key observations are as follows:

- (i) The  $\boldsymbol{\nu}$ -update in Algorithm 5.1 involves an inversion of the summation of two symmetric matrices  $\mathbf{A}^\top \mathbf{A}$  and  $\mathbf{D}_{xy}^\top \mathbf{D}_{xy}$  which does not, by and large, admit to a closed form

---

**Algorithm 5.1** The ADMM algorithm for solving (5.1) using variable splitting (5.2)

---

```

1: Set  $k = 0$ ,  $\mathbf{z}_0^{(1)}$ ,  $\mathbf{z}_0^{(2)}$ ,  $\mathbf{u}_0^{(1)}$ ,  $\mathbf{u}_0^{(2)}$ , and choose  $\rho > 0$ 
2: repeat
3:    $\boldsymbol{\nu}_{k+1} = (\mathbf{A}^\top \mathbf{A} + \mathbf{D}_{xy}^\top \mathbf{D}_{xy})^{-1} \left( \mathbf{A}^\top (\mathbf{z}_k^{(1)} - \mathbf{u}_k^{(1)}) + \mathbf{D}_{xy}^\top (\mathbf{z}_k^{(2)} - \mathbf{u}_k^{(2)}) \right)$ 
4:    $\mathbf{z}_{k+1}^{(1)} = \frac{1}{1+\rho} \left( \mathbf{y} + \rho \left( \mathbf{A}\boldsymbol{\nu}_{k+1} + \mathbf{u}_k^{(1)} \right) \right)$ 
5:    $\mathbf{z}_{k+1}^{(2)} = \text{prox}_{\|\cdot\|_1, \alpha/\rho} \left( \mathbf{D}_{xy}\boldsymbol{\nu}_{k+1} + \mathbf{u}_k^{(2)} \right)$ 
6:    $\mathbf{u}_{k+1}^{(1)} = \mathbf{u}_k^{(1)} + \left( \mathbf{A}\boldsymbol{\nu}_{k+1} - \mathbf{z}_{k+1}^{(1)} \right)$ 
7:    $\mathbf{u}_{k+1}^{(2)} = \mathbf{u}_k^{(2)} + \left( \mathbf{D}_{xy}\boldsymbol{\nu}_{k+1} - \mathbf{z}_{k+1}^{(2)} \right)$ 
8:    $k = k + 1$ 
9: until stopping criterion is satisfied

```

---

solution and necessitates iterative methods (e.g., conjugate gradient method) resulting in additional sub-iterations as well as an inexact  $\boldsymbol{\nu}$ -update. On the other hand, the inversion in Algorithm 5.2 can be solved exactly via the Fast Fourier Transform (Section 2.5.2.1). Furthermore, for the same  $\boldsymbol{\nu}$ -update, Algorithm 5.1 requires an additional matrix multiplication by  $\mathbf{A}^\top$ .

- (ii) While the  $\mathbf{z}^{(1)}$ -update in Algorithm 5.2 contains a matrix inversion, direct application of the matrix inversion lemma<sup>1</sup> and noting that  $\mathbf{A}^\top \mathbf{y}$  can be pre-computed reveals that the computational complexity is, more or less, on par with that in Algorithm 5.1 (assuming  $\mathbf{A}$  can be computed efficiently via fast Transforms).
- (iii) The  $\mathbf{u}^{(1)}$ -update in Algorithm 5.1 requires an additional matrix multiplication by  $\mathbf{A}$ .

While the analysis above ignores the rate of convergence between the two instances, judging by the complexity of each iteration, we believe that Algorithm 5.2 clearly has the upper hand. Further analysis into this matter may lead to efficient implementations as well as novel systematic approaches in automating variable splitting selections.

All instances of ADMM used in Chapter 2 and Chapter 3 require the user to supply or tune the penalty parameter  $\rho$  as well as various regularization parameter. While the algorithms are not overly sensitive to  $\rho$ , it nonetheless needs to vary in order to maintain near-optimal performance as detailed in Section 2.6.1.1. Despite the lack of consensus on the appropriate choices, adaptive schemes exist that iteratively update  $\rho$  based on a set of pre-determined

---

<sup>1</sup>Recall that for matrices  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$ ,  $\mathbf{D}$  of correct sizes, we have  $(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{D}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{D}\mathbf{A}^{-1}$ .

---

**Algorithm 5.2** The ADMM algorithm for solving (5.1) using variable splitting (5.3)

---

- 1: Set  $k = 0$ ,  $\mathbf{z}_0^{(1)}$ ,  $\mathbf{z}_0^{(2)}$ ,  $\mathbf{u}_0^{(1)}$ ,  $\mathbf{u}_0^{(2)}$ , and choose  $\rho > 0$
  - 2: **repeat**
  - 3:      $\boldsymbol{\nu}_{k+1} = (\mathbf{D}_{xy}^\top \mathbf{D}_{xy} + \mathbf{I})^{-1} \left( (\mathbf{z}_k^{(1)} - \mathbf{u}_k^{(1)}) + \mathbf{D}_{xy}^\top (\mathbf{z}_k^{(2)} - \mathbf{u}_k^{(2)}) \right)$
  - 4:      $\mathbf{z}_{k+1}^{(1)} = (\mathbf{A}^\top \mathbf{A} + \mathbf{I})^{-1} \left( \mathbf{A}^\top \mathbf{y} + (\boldsymbol{\nu}_{k+1} + \mathbf{u}_k^{(1)}) \right)$
  - 5:      $\mathbf{z}_{k+1}^{(2)} = \text{prox}_{\|\cdot\|_1, \alpha/\rho} \left( \mathbf{D}_{xy} \boldsymbol{\nu}_{k+1} + \mathbf{u}_k^{(2)} \right)$
  - 6:      $\mathbf{u}_{k+1}^{(1)} = \mathbf{u}_k^{(1)} + (\boldsymbol{\nu}_{k+1} - \mathbf{z}_{k+1}^{(1)})$
  - 7:      $\mathbf{u}_{k+1}^{(2)} = \mathbf{u}_k^{(2)} + (\mathbf{D}_{xy} \boldsymbol{\nu}_{k+1} - \mathbf{z}_{k+1}^{(2)})$
  - 8:      $k = k + 1$
  - 9: **until** stopping criterion is satisfied
- 

conditions on the primal and dual ADMM variables [76, 77, 127–129] and is an active research topic. Incorporating an appropriate scheme to our methods will make the performance less dependent on the initial choices of the penalty parameter. On the other hand, the values for the regularization parameter  $\gamma$  in Section 3.2 was selected manually in our experiments. The selections were merely to demonstrate the benefits of our methods and are by no means optimal. Proper automatic selection of the parameter using existing techniques (e.g., L-curve analysis [130], generalized cross-validation [131], etc.) could prove to be valuable.

In order to keep each contribution in our work independent so that individual sections are made comparable with the existing literature, it was necessary to isolated our work into the three separate modules. However, one of the original intentions of this research was to develop an integrated compressive hyperspectral system that addresses both the acquisition and recovery problems within the field. As such, a natural extension to the work presented in Chapter 2 and Chapter 3 is to consider integrating the proposed acquisition strategy with the developed recovery methods.

### 5.3 Unmixing

Despite the encouraging results presented in Chapter 4, these are nonetheless produced based on simulated datasets. As a proof-of-concept study, this is reasonable as it facilitates a highly controllable environment, enabling tuning of specific parameters. It is likely, however, that

the proposed method may not perform well in practice as many more influential factors now come in play. Shadowing and illumination variations, either as a result of non-uniform lighting or the topography of the scene, are present in virtually all image data. Non-linearity in the mixing will also hinder the applicability of our methods' underlying assumptions. Environmental variability and digitization error may lead to significant inaccuracies. Therefore, further experimentation with real hyperspectral data is highly desirable and the results from these experiments would better inform on the applicability of the methods. Proper characterization of these issues would ultimately stimulate appropriate compensation techniques necessary for real-world adaptations.

The unmixing procedures in Chapter 4 relied on an initial estimation of the number of endmembers present for the underlying hyperspectral data. Even with a prior knowledge of the scene, determining the optimal number of endmembers that minimizes the error metrics and/or compensate for the likely non-linearity is a difficult process in practice. The results presented in Section 4.2.3.1 (Figure 4.12) seem to suggest that the method presented is not overly sensitive to an overestimation in the number of endmembers and could potentially be modified for initial development of an automated procedure using the process outlined for generating Figure 4.12, i.e., using the residual error as a metric for appropriate selection.

The focus of Chapter 4 was on the simultaneous recoverability of the endmembers and abundances directly from CS measurements, thus by-passing the entire HSI reconstruction process. Conversely, it may be interesting to, instead, reconstruct the underlying HSI using the recovered endmembers and abundances from the methods described herein. Comparisons with the performance in Chapter 3 and with recent work along similar lines (e.g., [72, 102]) may be revealing and insightful.

As before, integrating the described methods with the detailed acquisition strategy in Chapter 2 is also a research direction deserving future consideration.



# Chapter 6

## Conclusions

Hyperspectral imaging systems offers significantly more information than standard colour cameras and has received increasing attention in a vast range of terrestrial-based applications where non-invasive measurements are preferred or essential. However, the associated costs and challenges faced in deploying such an inherently data-intensive technology have unfortunately resisted many prospective developments. Recent advances in signal processing on the other hand, despite their immaturity, have suggested plausible alternatives that alleviate these issues. The primary goal of this research was to investigate barriers in integrating one such alternative, known as compressive sensing, into hyperspectral systems. The contributions in this thesis fall into the three main components making up a compressive hyperspectral system: acquisition of hyperspectral images via compressive sensing, reconstruction of hyperspectral data from their compressive measurements, and unmixing of hyperspectral images in the compressive domain.

Practical compressive acquisition of hyperspectral images typically captures a sets of measurements, each containing independent information describing disjoint partitions of the underlying hyperspectral volume. Such independent multiplexing was found to have significant impact on the performance of many existing reconstruction methods that exploit inter-sparsity structures across the union of these hyperspectral partitions. In lieu of available technology for higher dimensional sensing methods, Chapter 2 investigated a pragmatic strategy that exploits the various acquisition modalities available for hyperspectral capture

and produced a hybrid reconstruction method, MuRecAL, that uses a combination of measurements obtained via the different modes for recovery. The new acquisition technique successfully facilitates existing reconstruction methods and was shown to outperform its single-mode counterparts by as much as 14 dB in the reconstruction error metric.

Ideal compressive hyperspectral reconstruction methods requires accurate modelling of both spatial and spectral data sparsity exhibited by the imaged scene. These two properties unfortunately cannot be captured by a single tractable prior for use in the related inverse problem. Models that linearly combine separate priors associated with each of the dimensions to account for the properties simultaneously were therefore considered in Chapter 3. Two convex optimizations of this kind were investigated in which the desired properties are induced by minimizing the spatial total variation norm and (1) the  $l_1$ -norm of each voxel in the DCT basis , and (2) the  $l_2$ -norm of the discrete gradient of each voxel. The efficacy of both methods were demonstrated via algorithmic implementation using ADMM, an improvement of at least 6 dB in the reconstruction error metric over existing techniques was achieved and shown through experimentation.

The identification of the endmembers and their abundances from CS measurements of hyperspectral images can be naively achieved through the sequential procedure of full CS HSI recovery followed by application of well-established unmixing algorithms. Inefficiencies in such a paradigm and the limited practical use of raw hyperspectral imagery have, instead, steered focus towards spectral unmixing directly via CS measurements. Recent works make strict assumptions about the underlying data in order to provide tractability to the problem, including pure pixel dominance and *a priori* knowledge of endmember spectra. Following the philosophy behind the success in the developed CS HSI recovery methods, a blind unmixing technique was developed in Chapter 4, based on the LMM, that uses multiple pertinent, but generic, priors to accounts for both the internal characteristics and the external geometry of the material constituents within the scene. The method was shown to be effective in unmixing highly mixed simulated scenarios and encourages further experimentation with real datasets.

Many directions worthy of future investigation still exists and are detailed in Chapter 5.

These include an examination into the allocation of measurement modes, a formal theoretical analysis of the detailed acquisition strategy, development of automated means for optimal parameter and variable selections for the proposed algorithms, an analysis of the applicability of the developed unmixing technique under practical scenarios, and the integration of the proposed methods into a unified system. The work within this thesis is not an end in itself, but a means of opening up new opportunities for discovery in the ever-growing spectrum of spectroscopic techniques



# Bibliography

- [1] E. Weinstein, K. Steele, A. Agarwal, and J. Glass, “LOUD: A 1020-Node Microphone Array and Acoustic Beamformer,” in *International Congress on Sound and Vibration (ICSV)*, Cairns, Australia, 2007.
- [2] C. P. Jayapandian, C.-H. Chen, A. Bozorgi, S. D. Lhatoo, G.-Q. Zhang, and S. S. Sahoo, “Cloudwave: Distributed Processing of ”Big Data” From Electrophysiological Recordings for Epilepsy Clinical Research Using Hadoop,” in *AMIA Annual Symposium*, 2013, pp. 691–700. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3787533/>
- [3] H. Akil, M. E. Martone, and D. C. Van Essen, “Challenges and Opportunities in Mining Neuroscience Data,” *Science*, vol. 331, no. 6018, pp. 708–712, 2011.
- [4] R. G. Baraniuk, “More Is Less: Signal Processing and the Data Deluge,” *Science*, vol. 331, no. 6018, pp. 717–719, 2011.
- [5] J. Harsanyi and C.-I. Chang, “Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 4, pp. 779–785, Jul. 1994. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=298007>
- [6] D. Landgrebe, “Hyperspectral image data analysis,” *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 17–28, 2002. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=974718>
- [7] N. Keshava and J. Mustard, “Spectral unmixing,” *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 44–57, 2002. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=974727>

- [8] D. S. Taubman and M. W. Marcellin, "Introduction to JPEG2000," in *JPEG2000 Image Compression Fundamentals, Standards and Practice*. Boston, MA: Springer US, 2002, pp. 399–415. [Online]. Available: <http://link.springer.com/10.1007/978-1-4615-0799-4>
- [9] I. Jolliffe, *Principal Component Analysis*. Springer US, 2002. [Online]. Available: <http://www.springer.com/gp/book/9780387954424>
- [10] Q. Du and J. E. Fowler, "Hyperspectral Image Compression Using JPEG2000 and Principal Component Analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 2, pp. 201–205, Apr. 2007. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4156154>
- [11] B. Penna, T. Tillo, E. Magli, and G. Olmo, "Transform Coding Techniques for Lossy Hyperspectral Data Compression," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 5, pp. 1408–1421, May 2007. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4156336>
- [12] J. Xing, C. Bravo, P. T. Jancsók, H. Ramon, and J. De Baerdemaeker, "Detecting Bruises on 'Golden Delicious' Apples using Hyperspectral Imaging with Multiple Wavebands," *Biosystems Engineering*, vol. 90, no. 1, pp. 27–36, Jan. 2005. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1537511004001515>
- [13] J. Gomez-Sanchis, L. Gomez-Chova, N. Aleixos, G. Camps-Valls, C. Montesinos-Herrero, E. Molto, and J. Blasco, "Hyperspectral systems for early detection of rotteness caused by Penicillium digitatum in mandarins," *Journal of Food Engineering*, vol. 89, no. 1, pp. 80–86, 2008.
- [14] L. N. Brewer, J. A. Ohlhausen, P. G. Kotula, and J. R. Michael, "Forensic analysis of bioagents by X-ray and TOF-SIMS hyperspectral imaging." *Forensic science international*, vol. 179, no. 2-3, pp. 98–106, Aug. 2008. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/18571885>
- [15] S. Dowler, "Applications of hyperspectral imaging techniques to forensic image analysis," Ph.D. dissertation, The University of Auckland, 2010.

- [16] T. Nagaoka, A. Nakamura, H. Okutani, Y. Kiyohara, and T. Sota, “A possible melanoma discrimination index based on hyperspectral data: a pilot study.” *Skin research and technology*, vol. 18, no. 3, pp. 301–10, Aug. 2012. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/22092570>
- [17] S. V. Panasyuk, S. Yang, D. V. Faller, D. Ngo, R. A. Lew, J. E. Freeman, and A. E. Rogers, “Medical hyperspectral imaging to facilitate residual tumor identification during surgery.” *Cancer biology & therapy*, vol. 6, no. 3, pp. 439–46, Mar. 2007. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/17374984>
- [18] E. J. Candès and T. Tao, “Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?” *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4016283>
- [19] E. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1580791>
- [20] E. J. Candès, J. K. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006. [Online]. Available: <http://doi.wiley.com/10.1002/cpa.20124>
- [21] D. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1614066>
- [22] R. Coifman, F. Geshwind, and Y. Meyer, “Noiselets,” *Applied and Computational Harmonic Analysis*, vol. 10, no. 1, pp. 27–44, Jan. 2001. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1063520300903130>
- [23] R. Vershynin, “Introduction to the non-asymptotic analysis of random matrices,” in *Compressed Sensing: Theory and Applications*, Y. C. Eldar and G. Kutyniok,

- Eds. Cambridge University Press, Nov. 2010, p. 62. [Online]. Available: <http://arxiv.org/abs/1011.3027>
- [24] D. Donoho and X. Huo, “Uncertainty principles and ideal atomic decomposition,” *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2845–2862, 2001. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=959265>
- [25] E. J. Candès and Y. Plan, “A Probabilistic and RIPless Theory of Compressed Sensing,” *IEEE Transactions on Information Theory*, vol. 57, no. 11, pp. 7235–7254, Nov. 2011. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5967912>
- [26] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic Decomposition by Basis Pursuit,” *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, Jan. 1998. [Online]. Available: <http://pubs.siam.org/doi/abs/10.1137/S1064827596304010>
- [27] E. Candès and J. Romberg, “Sparsity and incoherence in compressive sampling,” *Inverse Problems*, vol. 23, no. 3, pp. 969–985, Jun. 2007. [Online]. Available: <http://stacks.iop.org/0266-5611/23/i=3/a=008>
- [28] E. Candès and T. Tao, “Decoding by Linear Programming,” *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1542412>
- [29] T. Cai, “Shifting Inequality and Recovery of Sparse Signals,” *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1300–1308, Mar. 2010. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5290058>
- [30] T. T. Cai, L. Wang, and G. Xu, “New Bounds for Restricted Isometry Constants,” *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4388–4394, Sep. 2010. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5550400>
- [31] E. J. Candès, “The restricted isometry property and its implications for compressed sensing,” *Comptes Rendus Mathematique*, vol. 346, no. 9-10, pp. 589–592, May 2008. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1631073X08000964>

- [32] S. Foucart and M.-J. Lai, “Sparsest solutions of underdetermined linear systems via  $\ell_q$ -minimization for  $0 < q \leq 1$ ,” *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 395–407, May 2009. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1063520308000882>
- [33] S. Foucart, “A note on guaranteed sparse recovery via  $\ell_1$ -minimization,” *Applied and Computational Harmonic Analysis*, vol. 29, no. 1, pp. 97–103, Jul. 2010. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1063520309001158>
- [34] Q. Mo and S. Li, “New bounds on the restricted isometry constant  $\delta_{2k}$ ,” *Applied and Computational Harmonic Analysis*, vol. 31, no. 3, pp. 460–468, Nov. 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1063520311000625>
- [35] M. E. Davies and R. Gribonval, “Restricted Isometry Constants Where  $\ell_p$  Sparse Recovery Can Fail for  $0 < p \neq 1$ ,” *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2203–2214, May 2009. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4839048>
- [36] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, “Uniform Uncertainty Principle for Bernoulli and Subgaussian Ensembles,” *Constructive Approximation*, vol. 28, no. 3, pp. 277–289, Feb. 2008. [Online]. Available: <http://link.springer.com/10.1007/s00365-007-9005-8>
- [37] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, “A Simple Proof of the Restricted Isometry Property for Random Matrices,” *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, Jan. 2008. [Online]. Available: <http://link.springer.com/10.1007/s00365-007-9003-x>
- [38] M. Rudelson and R. Vershynin, “On sparse reconstruction from Fourier and Gaussian measurements,” *Communications on Pure and Applied Mathematics*, vol. 61, no. 8, pp. 1025–1045, Aug. 2008. [Online]. Available: <http://doi.wiley.com/10.1002/cpa.20227>
- [39] H. Rauhut, “Compressive Sensing and Structured Random Matrices,” in *Theoretical Foundations and Numerical Methods for Sparse Recovery*, F. Massimo, Ed. DeGruyter, 2010, pp. 1–92.

- [40] H. Rauhut, J. Romberg, and J. A. Tropp, “Restricted isometries for partial random circulant matrices,” *Applied and Computational Harmonic Analysis*, vol. 32, no. 2, pp. 242–254, Mar. 2012. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1063520311000649>
- [41] J. Haupt, W. U. Bajwa, G. Raz, and R. Nowak, “Toeplitz Compressed Sensing Matrices With Applications to Sparse Channel Estimation,” *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5862–5875, Nov. 2010. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5605341>
- [42] R. Baraniuk and P. Steeghs, “Compressive Radar Imaging,” in *2007 IEEE Radar Conference*. IEEE, Apr. 2007, pp. 128–133. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4250297>
- [43] J. Romberg, “Compressive Sensing by Random Convolution,” *SIAM Journal on Imaging Sciences*, vol. 2, no. 4, pp. 1098–1128, Jan. 2009. [Online]. Available: <http://pubs.siam.org/doi/abs/10.1137/08072975X>
- [44] R. F. Marcia, Z. T. Harmany, and R. M. Willett, “Compressive coded aperture imaging,” in *Proc. SPIE 7246, Computational Imaging VII, 72460G*, C. A. Bouman, E. L. Miller, and I. Pollak, Eds. International Society for Optics and Photonics, Feb. 2009, pp. 72460G–72460G–13. [Online]. Available: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=812490>
- [45] W. Pultz and W. Hertl, “Perturbations in the infrared spectra of submicroscopic size, single crystal, fibrous SiC,” *Spectrochimica Acta*, vol. 22, no. 3, pp. 573–575, Mar. 1966. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0371195166800887>
- [46] A. G. Palmer and N. L. Thompson, “Fluorescence correlation spectroscopy for detecting submicroscopic clusters of fluorescent molecules in membranes,” *Chemistry and Physics of Lipids*, vol. 50, no. 3-4, pp. 253–270, Jun. 1989. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0009308489900534>
- [47] M. M. A. Elsayed and G. Cevc, “Turbidity spectroscopy for characterization of submicroscopic drug carriers, such as nanoparticles and lipid vesicles: size

- determination.” *Pharmaceutical research*, vol. 28, no. 9, pp. 2204–22, Sep. 2011. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21584845>
- [48] T. Pan and L. Sun, “Sub-microscopic phenomena of metallic corrosion studied by a combined Photoelectron Spectroscopy in Air (PESA) and Scanning Kelvin Probe Force Microscopy (SKPFM) approach,” *International Journal of Electrochemical Science*, vol. 7, no. 10, pp. 9325–9344, 2012.
- [49] D. Landgrebe, “Multispectral land sensing: where from, where to?” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 414–421, Mar. 2005. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1396315>
- [50] C. Creaser and A. Davies, *Analytical Applications of Spectroscopy II*. Royal Society of Chemistry, 1988.
- [51] B. E. Bayer, “Color Imaging Array,” 1976.
- [52] T. Wilson and R. Felt, “Hyperspectral remote sensing technology (HRST) program,” in *1998 IEEE Aerospace Conference Proceedings (Cat. No.98TH8339)*, vol. 5. IEEE, 1998, pp. 193–200. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=685819>
- [53] G. Healey, M. Prasad, and B. Tromberg, “Face recognition in hyperspectral images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1552–1560, Dec. 2003. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1251148>
- [54] A. Baldridge, S. Hook, C. Grove, and G. Rivera, “The ASTER spectral library version 2.0,” *Remote Sensing Environment*, vol. 113, pp. 711–715, 2009. [Online]. Available: <http://speclib.jpl.nasa.gov/downloads/2009-Baldridge.pdf>
- [55] S. Lundein, “AVIRIS Airborne Visible/Infrared Imaging Spectrometer,” 2015. [Online]. Available: <http://aviris.jpl.nasa.gov/>
- [56] G. Shaw and D. Manolakis, “Signal processing for hyperspectral image exploitation,” *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 12–16, 2002. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=974715>

- [57] S. Poger and E. Angelopoulou, "Selecting Components for Building Multispectral Sensors," *IEEE CVPR Technical Sketches*, 2001.
- [58] R. Schowengerdt, *Remote Sensing Models and Methods for Image Processing*, 3rd ed. Elsevier Inc., 2007.
- [59] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral Unmixing Overview: Geometrical, Statistical, and Sparse Regression-Based Approaches," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 354–379, Apr. 2012. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6200362>
- [60] M. Duarte, M. Davenport, D. Takhar, J. Laska, K. Kelly, and R. Baraniuk, "Single-Pixel Imaging via Compressive Sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, Mar. 2008. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4472247>
- [61] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. Kelly, and R. Baraniuk, "An Architecture for Compressive Imaging," in *2006 International Conference on Image Processing*. IEEE, 2006, pp. 1273–1276. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4106769>
- [62] T. Sun and K. Kelly, "Compressive Sensing Hyperspectral Imager," in *Frontiers in Optics 2009/Laser Science XXV/Fall 2009 OSA Optics & Photonics Technical Digest*. Washington, D.C.: OSA, Oct. 2009, p. CTuA5. [Online]. Available: <http://www.osapublishing.org/abstract.cfm?uri=COSI-2009-CTuA5>
- [63] J. E. Fowler, "Compressive pushbroom and whiskbroom sensing for hyperspectral remote-sensing imaging," in *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, Oct. 2014, pp. 684–688. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7025137>
- [64] G. Coluccia, S. K. Kuiteing, A. Abrardo, M. Barni, and E. Magli, "Progressive Compressed Sensing and Reconstruction of Multidimensional Signals Using Hybrid Transform/Prediction Sparsity Model," *IEEE Journal on Emerging and Selected*

- Topics in Circuits and Systems*, vol. 2, no. 3, pp. 340–352, Sep. 2012. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6331018>
- [65] M. E. Gehm, R. John, D. J. Brady, R. M. Willett, and T. J. Schulz, “Single-shot compressive spectral imaging with a dual-disperser architecture.” *Optics express*, vol. 15, no. 21, pp. 14 013–27, Oct. 2007. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/19550674>
- [66] A. Wagadarikar, R. John, R. Willett, and D. Brady, “Single disperser design for coded aperture snapshot spectral imaging.” *Applied optics*, vol. 47, no. 10, pp. B44–51, Apr. 2008. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/18382550>
- [67] Y. Wu, I. O. Mirza, G. R. Arce, and D. W. Prather, “Development of a digital-micromirror-device-based multishot snapshot spectral imaging system.” *Optics letters*, vol. 36, no. 14, pp. 2692–4, Jul. 2011. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21765511>
- [68] Y. August, C. Vachman, Y. Rivenson, and A. Stern, “Compressive hyperspectral imaging by random separable projections in both the spatial and the spectral domains.” *Applied optics*, vol. 52, no. 10, pp. D46–54, Apr. 2013. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/23545982>
- [69] H. Liu, Y. Li, S. Xiao, and C. Wu, “Distributed Compressive Hyperspectral Image Sensing,” in *2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*. IEEE, Oct. 2010, pp. 607–610. [Online]. Available: <http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=5636246>
- [70] A. Abrardo, M. Barni, C. Carretti, S. Kamdem, and E. Magli, “A compressive sampling scheme for iterative hyperspectral image reconstruction,” in *2011 19th European Signal Processing Conference*. Barcelona: IEEE, 2011, pp. 1120–1124.
- [71] C. Li, T. Sun, K. F. Kelly, and Y. Zhang, “A compressive sensing and unmixing scheme for hyperspectral data processing,” *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 1200–10, Mar. 2012. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21914570>

- [72] M. Golbabaei, S. Arberet, and P. Vandergheynst, “Compressive source separation: theory and methods for hyperspectral imaging.” *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5096–110, Dec. 2013. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/24043385>
- [73] A. A. Wagadarikar, N. P. Pitsianis, X. Sun, and D. J. Brady, “Spectral image estimation for coded aperture snapshot spectral imagers,” in *Optical Engineering + Applications*, P. J. Bones, M. A. Fiddy, and R. P. Millane, Eds. International Society for Optics and Photonics, Aug. 2008, p. 707602. [Online]. Available: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=794777>
- [74] M. F. Duarte and R. G. Baraniuk, “Kronecker compressive sensing,” *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 494–504, Feb. 2012. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21859622>
- [75] M. Golbabaei and P. Vandergheynst, “Joint trace/TV norm minimization: A new efficient approach for spectral compressive imaging,” in *2012 19th IEEE International Conference on Image Processing*. IEEE, Sep. 2012, pp. 933–936. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6467014>
- [76] S. Boyd, “Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers,” *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, Jan. 2010. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2185815.2185816>
- [77] J. Nocedal and S. Wright, “Penalty, Barrier, and Augmented Lagrangian Methods,” in *Numerical Optimization*, ser. Springer Series in Operations Research and Financial Engineering, J. Nocedal and S. J. Wright, Eds. New York: Springer-Verlag, 1999, pp. 511–520. [Online]. Available: <http://www.springerlink.com/index/10.1007/b98874>
- [78] J. Eckstein and D. P. Bertsekas, “On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators,” *Mathematical Programming*, vol. 55, no. 1-3, pp. 293–318, Apr. 1992. [Online]. Available: <http://link.springer.com/10.1007/BF01581204>

- [79] P. L. Combettes and J.-C. Pesquet, "Proximal Splitting Methods in Signal Processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. Russell Luke, and H. Wolkowicz, Eds. Springer New York, 2011, vol. 49, pp. 185–212. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.182.9488>
- [80] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems." *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 681–95, Mar. 2011. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/20840899>
- [81] D. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, May 1995. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=382009>
- [82] P. J. Davis, "Generalizations of Circulants: g-Circulants and Block Circulants," in *Circulant Matrices*, 1st ed. New York: John Wiley and Sons, 1979, pp. 184–186.
- [83] J.-F. Cai, E. J. Candès, and Z. Shen, "A Singular Value Thresholding Algorithm for Matrix Completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, Jan. 2010. [Online]. Available: <http://pubs.siam.org/doi/abs/10.1137/080738970>
- [84] R. Ward and F. Krahmer, "Stable and robust sampling strategies for compressive imaging." *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 612–622, Nov. 2013. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/24196861>
- [85] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, no. 3, pp. 947–968, Jun. 2007. [Online]. Available: <http://stacks.iop.org/0266-5611/23/i=3/a=007>
- [86] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1-4, pp. 259–268, Nov. 1992. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/016727899290242F>

- [87] A. Chambolle, “An Algorithm for Total Variation Minimization and Applications,” *Journal of Mathematical Imaging and Vision*, vol. 20, no. 1-2, pp. 89–97, 2004. [Online]. Available: <http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.160.5226>
- [88] S. H. Chan, R. Khoshabeh, K. B. Gibson, P. E. Gill, and T. Q. Nguyen, “An augmented Lagrangian method for total variation video restoration.” *IEEE Transactions on Image Processing*, vol. 20, no. 11, pp. 3097–111, Nov. 2011. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21632302>
- [89] W. Guo and W. Yin, “EdgeCS: edge guided compressive sensing reconstruction,” in *Visual Communications and Image Processing 2010*, P. Frossard, H. Li, F. Wu, B. Girod, S. Li, and G. Wei, Eds. International Society for Optics and Photonics, Jul. 2010, pp. 77440L–77440L–10. [Online]. Available: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1346856>
- [90] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, “Total Variation Spatial Regularization for Sparse Hyperspectral Unmixing,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 11, pp. 4484–4502, Nov. 2012. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6196219>
- [91] D. Needell and R. Ward, “Near-optimal compressed sensing guarantees for total variation minimization.” *IEEE Transactions on Image Processing*, vol. 22, no. 10, pp. 3941–9, Oct. 2013. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/23708808>
- [92] ——, “Stable Image Reconstruction Using Total Variation Minimization,” *SIAM Journal on Imaging Sciences*, vol. 6, no. 2, pp. 1035–1058, Jun. 2013. [Online]. Available: <http://pubs.siam.org/doi/abs/10.1137/120868281>
- [93] M. Golbabae and P. Vandergheynst, “Hyperspectral image compressed sensing via low-rank and joint-sparse matrix recovery,” in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, Mar. 2012, pp. 2741–2744. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6288484>

- [94] B. Recht, M. Fazel, and P. A. Parrilo, “Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization,” *SIAM Review*, vol. 52, no. 3, pp. 471–501, Jan. 2010. [Online]. Available: <http://pubs.siam.org/doi/abs/10.1137/070697835>
- [95] E. J. Candès and Y. Plan, “Tight Oracle Inequalities for Low-Rank Matrix Recovery From a Minimal Number of Noisy Random Measurements,” *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2342–2359, Apr. 2011. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5730578>
- [96] R. N. Clark, G. A. Swayze, R. A. Wise, K. E. Livo, T. M. Hoefen, R. F. Kokaly, and S. J. Sutley, “USGS Digital Spectral Library splib06a,” *Data Series*, 2007. [Online]. Available: <http://pubs.er.usgs.gov/publication/ds231>
- [97] M. E. Winter, “N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data,” in *SPIE’s International Symposium on Optical Science, Engineering, and Instrumentation*, M. R. Descour and S. S. Shen, Eds. International Society for Optics and Photonics, Oct. 1999, pp. 266–275. [Online]. Available: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=994814>
- [98] J. Nascimento and J. Dias, “Vertex component analysis: a fast algorithm to unmix hyperspectral data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, Apr. 2005. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1411995>
- [99] J. Li and J. M. Bioucas-Dias, “Minimum Volume Simplex Analysis: A Fast Algorithm to Unmix Hyperspectral Data,” in *IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium*, vol. 3. IEEE, 2008, pp. III – 250–III – 253. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4779330>
- [100] L. Miao and H. Qi, “Endmember Extraction From Highly Mixed Data Using Minimum Volume Constrained Nonnegative Matrix Factorization,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 3, pp. 765–777, Mar. 2007. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4106058>

- [101] M. Golbabaei, S. Arberet, and P. Vandergheynst, “Compressive source separation: theory and methods for hyperspectral imaging.” *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5096–110, Dec. 2013. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/24043385>
- [102] G. Martin, J. M. Bioucas-Dias, and A. Plaza, “HYCA: A New Technique for Hyperspectral Compressive Sensing,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2819–2831, May 2015. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6964803>
- [103] A. Zare and K. Ho, “Endmember Variability in Hyperspectral Analysis: Addressing Spectral Variability During Spectral Unmixing,” *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 95–104, Jan. 2014. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6678271>
- [104] M. Golbabaei, S. Arberet, and P. Vandergheynst, “Distributed compressed sensing of Hyperspectral images via blind source separation,” in *2010 Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers*. IEEE, Nov. 2010, pp. 196–198. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5757497>
- [105] Q. Zhang, R. Plemmons, D. Kittle, D. Brady, and S. Prasad, “Joint segmentation and reconstruction of hyperspectral data with compressed measurements.” *Applied optics*, vol. 50, no. 22, pp. 4417–35, Aug. 2011. [Online]. Available: <http://www.osapublishing.org/viewmedia.cfm?uri=ao-50-22-4417&seq=0&html=true>
- [106] D. T. Eason and M. Andrews, “Total variation regularization via continuation to recover compressed hyperspectral images.” *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 284–93, Jan. 2015. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/25438319>
- [107] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, “Compressed sensing with coherent and redundant dictionaries,” *Applied and Computational Harmonic Analysis*, vol. 31, no. 1, pp. 59–73, Jul. 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1063520310001156>

- [108] X.-L. Zhao, F. Wang, T.-Z. Huang, M. K. Ng, and R. J. Plemmons, “Deblurring and Sparse Unmixing for Hyperspectral Images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 7, pp. 4045–4058, Jul. 2013. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6423278>
- [109] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*. Society for Industrial and Applied Mathematics, 1974. [Online]. Available: <http://pubs.siam.org/doi/book/10.1137/1.9781611971217>
- [110] N. Ho, “Nonnegative Matrix Factorization Algorithms and Applications,” PhD Thesis, Université catholique de Louvain, 2008.
- [111] A. Cichocki and A.-H. Phan, “Fast Local Algorithms for Large Scale Nonnegative Matrix and Tensor Factorizations,” *IEICE Transactions on Fundamentals of Electronics*, vol. E92-A, no. 3, pp. 708–721, 2009. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.214.6398>
- [112] L. Li and Y.-J. Zhang, “FastNMF: highly efficient monotonic fixed-point nonnegative matrix factorization algorithm with good applicability,” *Journal of Electronic Imaging*, vol. 18, no. 3, p. 033004, Jul. 2009. [Online]. Available: <http://adsabs.harvard.edu/abs/2009JEI....18c3004L>
- [113] F. Kruse, A. Lefkoff, J. Boardman, K. Heidebrecht, A. Shapiro, P. Barloon, and A. Goetz, “The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data,” *Remote Sensing of Environment*, vol. 44, no. 2-3, pp. 145–163, May 1993. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/003442579390013N>
- [114] A. Beck and M. Teboulle, “A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems,” *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, Jan. 2009. [Online]. Available: <http://pubs.siam.org/doi/abs/10.1137/080716542>
- [115] D. P. Palomar and Y. C. Eldar, *Convex Optimization in Signal Processing and Communications*. Cambridge University Press, 2009.

- [116] Y. Nesterov, *Introductory Lectures on Convex Optimization*, ser. Applied Optimization. Boston, MA: Springer US, 2004, vol. 87. [Online]. Available: <http://link.springer.com/10.1007/978-1-4419-8853-9>
- [117] D. D. Lee and H. S. Seung, “Algorithms for Non-negative Matrix Factorization,” in *Advances in Neural Information Processing Systems*, 2001, pp. 556–562. [Online]. Available: <http://papers.nips.cc/paper/1861-algorithms-for-non-negative-matrix-factorization>
- [118] C.-J. Lin, “Projected gradient methods for nonnegative matrix factorization.” *Neural computation*, vol. 19, no. 10, pp. 2756–79, Oct. 2007. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/17716011>
- [119] A. Beck and M. Teboulle, “Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems.” *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2419–34, Nov. 2009. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/19635705>
- [120] A. Huck, M. Guillaume, and J. Blanc-Talon, “Minimum Dispersion Constrained Nonnegative Matrix Factorization to Unmix Hyperspectral Data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 6, pp. 2590–2602, Jun. 2010. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5422792>
- [121] Lu Gan, T. Do, and T. Tran, “Fast compressive imaging using scrambled block Hadamard ensemble,” pp. 1–5, 2008.
- [122] A. M. Tillmann and M. E. Pfetsch, “The Computational Complexity of the Restricted Isometry Property, the Nullspace Property, and Related Concepts in Compressed Sensing,” *IEEE Transactions on Information Theory*, vol. 60, no. 2, pp. 1248–1259, Feb. 2014. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6658871>
- [123] B. Adcock, A. C. Hansen, C. Poon, and B. Roman, “Breaking the coherence barrier: A new theory for compressed sensing,” Feb. 2013. [Online]. Available: <http://arxiv.org/abs/1302.0561>

- [124] A. Bastounis and A. C. Hansen, “On the absence of the RIP in real-world applications of compressed sensing and the RIP in levels,” Nov. 2014. [Online]. Available: <http://arxiv.org/abs/1411.4449>
- [125] R. Kueng and D. Gross, “RIPless compressed sensing from anisotropic measurements,” *Linear Algebra and its Applications*, vol. 441, pp. 110–123, Jan. 2014. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0024379513002802>
- [126] H. Calderón, J. F. Silva, J. M. Ortiz, and A. Egaña, “Reconstruction of channelized geological facies based on RIPless compressed sensing,” *Computers & Geosciences*, vol. 77, pp. 54–65, Apr. 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0098300415000072>
- [127] B. S. He, H. Yang, and S. L. Wang, “Alternating Direction Method with Self-Adaptive Penalty Parameters for Monotone Variational Inequalities,” *Journal of Optimization Theory and Applications*, vol. 106, no. 2, pp. 337–356, 2000. [Online]. Available: <http://link.springer.com/article/10.1023/A%3A1004603514434>
- [128] S. L. Wang and L. Z. Liao, “Decomposition Method with a Variable Parameter for a Class of Monotone Variational Inequality Problems,” *Journal of Optimization Theory and Applications*, vol. 109, no. 2, pp. 415–429, 2001. [Online]. Available: <http://link.springer.com/article/10.1023/A%3A1017522623963>
- [129] Z. Lin, R. Liu, and Z. Su, “Linearized Alternating Direction Method with Adaptive Penalty for Low-Rank Representation,” in *Advances in Neural Information Processing Systems*, 2011, pp. 612–620. [Online]. Available: <http://papers.nips.cc/paper/4434-linearized-alternating-direction-method-with-adaptive-penalty-for-low-rank-representation.pdf>
- [130] P. C. Hansen and D. P. O’Leary, “The Use of the L-Curve in the Regularization of Discrete Ill-Posed Problems,” *SIAM Journal on Scientific Computing*, vol. 14, no. 6, pp. 1487–1503, Nov. 1993. [Online]. Available: <http://pubs.siam.org/doi/abs/10.1137/0914086>
- [131] G. H. Golub, M. Heath, and G. Wahba, “Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter,” *Technometrics*, vol. 21, no. 2, pp. 215–223, Apr.

1979. [Online]. Available: <http://www.tandfonline.com/doi/abs/10.1080/00401706.1979.10489751#.VlS-4fmqpBc>