Inverse Problems in Hyperspectral Imaging

By

Sebastian Berisha
Doctor of Philosophy

Mathematics and Computer Science

James G. Nagy, PhD.
Advisor

Alessandro Veneziani, PhD.
Committee Member

Bree Ettinger, PhD.
Committee Member

Robert J. Plemmons, PhD.
Committee Member

Accepted:

Lisa A. Tedesco, PhD.
Dean of the James T. Laney School of Graduate Studies

Date

Inverse Problems in Hyperspectral Imaging

By

Sebastian Berisha
M.Sc., Wake Forest University, 2009

Advisor: James G. Nagy, PhD.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Mathematics and Computer Science

2014

Abstract

Inverse Problems in Hyperspectral Imaging

By
Sebastian Berisha

In hyperpsectral imaging, multiple images of the same scene are obtained over a contiguous range of wavelengths in the electromagnetic spectrum. Hyperspectral images represent observations of a scene at many different wavelengths and most importantly associate to each pixel in the imaged scene a full spectral vector or spectral signature. However, due to the presence of spectral mixtures (at different scales) in the scene and/or low spatial resolution of the hyperspectral sensor, the acquired spectral vectors of each pixel are actually a mixture of the spectra of the various materials present in the spatial coverage area of the corresponding pixel. Spectral unmixing is an important task in hyperspectral data analysis that aims to separate the measured pixel spectra into a set of spectral signatures of constituent materials (endmembers) and their corresponding fractional abundances, which represent the contribution percentage of the spectrum of each material present in the pixel. The blurring introduced by the hyperspectral imaging system and/or seeing conditions contributes furthermore to the overall mixing of the spectrum of different materials. Thus, hyperspectral unmixing and deblurring are necessary to be able to effectively perform hyperspectral image processing tasks.

This thesis presents a numerical approach for deblurring and sparse unmixing of ground-based astronomical images of space objects taken through the atmosphere at multiple wavelengths with narrow spectral channels. A major challenge for deblurring hyperspectral images is that of estimating the overall blurring operator, taking into account the fact that the blurring operator point spread function (PSF) can be wavelength dependent and depend on the imaging system as well as the effects of atmospheric turbulence on the arriving wavefront. Thus, we begin by developing an iterative approach to solve separable nonlinear least squares problems arising in the estimation of hyperspectral PSF parameters. A variable projection Gauss-Newton method is used to solve the nonlinear least squares problem. Our analysis shows that the Jacobian can be potentially very ill-conditioned. To deal with this ill-conditioning, we use a combination of subset selection and regularization. Experimental results related to hyperspectral PSF parameter identification and star spectrum reconstruction illustrate the effectiveness of the resulting numerical scheme. We then combine our hyperpsectral PSF estimation method with hyperpsectral deblurring and sparse unmixing by deriving a numerical scheme using multiple PSFs. Our approach is based on a preconditioned alternating direction method of multipliers. The effectiveness of our method is illustrated with numerical experiments on a commonly used test example, a simulated hyperspectral image of the Hubble Space Telescope satellite.

Inverse Problems in Hyperspectral Imaging

By

Sebastian Berisha
M.Sc., Wake Forest University, 2009

Advisor: James G. Nagy, PhD.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Mathematics and Computer Science

2014

Acknowledgements

Many people have helped, motivated, and inspired me both in my academic and personal life. I believe more in showing appreciation by good deeds rather than simple words. However, to adhere to the formalities, I would like to mention and thank some people who have helped me tremendously during my graduate school life.

I would like to thank my advisor Professor James Nagy for being my academic guidance throughout these years. He has inspired me with his outstanding passion and dedication for both teaching and research. Thanks for supporting me financially and for giving me the opportunity to travel and participate in conferences and trainings. Prof. Nagy is not only a great professor and researcher but also very personable and understanding. I consider myself lucky and privileged to have worked with you. Thank you for everything. Thanks to my other committee members, Professors Alessandro Veneziani, and Bree Ettinger for taking the time to provide me with constructive suggestions regarding my dissertation research.

I would like to thank Professor Robert Plemmons for his continuous support and motivation throughout the years. I have learned a lot and have been greatly inspired by his research. I am truly lucky and honored to have had and to continue to have the opportunity to work with you. Thank you also to Professor Plemmons' family and Professor Paul Pauca and his family for their continuous love and help.

Thanks to the graduate students in my batch for being supportive, helpful, and good friends over these years. I would like to thank Anita and Jim for their continuous support and friendship. Your house in the mountain is the best stress reliever for a graduate student. Thank you to Sussi for her friendship and help since I began graduate studies at Emory. I will miss your great food. Your kindness and careness have brought joy along this journey.

I would like to express my gratefulness to my family. My parents and my brother have always been unconditionally supportive and proud of any little accomplishment of mine. Thanks for teaching me not to give up.

I would like to end with a quote by Mother Theresa:

"We must know that we have been created for greater things, not just to be a number in the world, not just to go for diplomas and degrees, this work and that work. We have been created in order to love and to be loved."

Contents

List of Figures

List of Tables

Variable Definitions

N_m number of endmembers

N_w number of spectral bands or wavelengths

i pixel index

N_p number of pixels in an image

γ the ellipticity parameter in the elliptical Moffat function

Θ the rotation angle in the elliptical Moffat function

s_λ the unknown intensity of the star spectrum at wavelength λ

p the number of parameters used to define the hyperspectral PSF

N_m number of attenuating materials to be modeled

μ_1 sparsity regularization term

μ_2 total variation regularization term

$\boldsymbol{D}_v$ first order difference matrix in the vertical direction

$\boldsymbol{D}_h$ first order difference matrix in the horizontal direction

$\boldsymbol{e}_i$ the ith unit vector

$\boldsymbol{H}$ the overall hyperspectral blurring matrix

$\boldsymbol{N}$ the matrix representing the errors or noise

$\boldsymbol{X}$ the fractional abundances matrix of endmembers

$\boldsymbol{\phi}_\lambda$ vector of unknown parameters corresponding to wavelength λ

$\boldsymbol{h}_\lambda$ a vector representing the vectorized form of an exact original image of an isolated star corresponding to wavelength λ

$\boldsymbol{g}_\lambda$ a vector representing the vectorized form of an observed, blurred, and noisy image of an isolated star corresponding to wavelength λ

$\boldsymbol{G}$ the observed hyperspectral data matrix (each row contains the observed spectrum of a given pixel)

$\boldsymbol{M}$ spectral library which contains spectral signatures of endmembers

---

Chapter 1

Introduction

Information about the material composition of an object is contained most unequivocally in the spectral profiles of the brightness at the different surface pixels of the object. By acquiring the surface brightness distribution in narrow spectral channels, as in hyperspectral image (HSI) data cubes, and by performing spectral unmixing on such data cubes, one can infer the material identity as a function of position on the object surface [45]. A major advantage of hyperspectral imaging (HSI) over traditional imaging methods is its ability to perform object identification and classification in image scenes which contain objects that are too small to be detected by spatial imaging methods. A disadvantage is the fact that the improved ability to acquire HSI data over a large number of wavelengths poses data storage and data processing challenges.

A considerable amount of research has been done to develope hyperspectral image processing methods for HSI data compression, spectral signature identification of constituent materials, and determination of their corresponding fractional abundances [20, 35, 43, 53, 54, 71, 75, 90, 95, 116, 145, 147]. HSI is used in a wide range of applications, such as remote surveillance, astrophysics, medical imaging, geophysics, agriculture, and mineral exploration [94, 115, 131].

Knowledge of the data acquisition system is required to effectively perform HSI processing tasks. In particular, the data acquisition system introduces blurring which can be modeled by a convolution kernel or a point spread function (PSF). Moreover, the convolution of the PSF with the source data increases the mixing of the spectrum of different objects present in the imaged scene. A major challenge for deblurring HSIs consists in the estimation of the overall blurring operator $\boldsymbol{H}$, taking into account the fact that the blurring operator PSF can vary over the images in the HSI data cube. That is, the blurring can be wavelength dependent and depends on the imaging system as well as the effects of atmospheric turbulence on the arriving wavefront, see e.g., [87, 127, 132, 141]. In addition, the observing conditions change with the data and thus the PSFs cannot be obtained by a single calibration of the optical instrument, but instead have to be estimated for each HSI datacube.

In [133] the authors assume the model of the PSF is a linear combination of Gaussian functions and thus their blur identification process corresponds to finding the scalar weights for each of the Gaussian functions. More recently a hyperspectral Multi Unit Spectroscopic Explorer (MUSE) system has been installed on the Very Large Telescope (VLT) and deployed by the European Southern Observatory (ESO) at the Paranal Observatory in Chile. The MUSE system will collect up to $4,000$ bands, and research is ongoing to develop methods for estimation of the wavelength dependent PSFs for deblurring the resulting HSI datacube for ground-based astrophysical observations, see e.g. [127, 141]. In particular, Soules et al. [132] consider the restoration of hyperspectral astronomical data with spectrally varying blur, but assume that the spectrally varying PSF defining the spectrally varying convolution has already been provided by other means, and defer the PSF estimation to a later time.

An important method in hyperpsectral data analysis is hyperspectral unmixing [19, 143, 76, 75, 21, 74]. The goal of hypersectral unmixing is to find the un-

derlying materials present in each pixel of the observed scene. A widely used spectral mixing model is the linear mixture model which assumes that the measured mixed spectrum of a pixel can be expressed as a linear combination of the spectral signatures of "pure" materials, also called endmembers. The spectra of endmembers is measured in laboratory settings. Given the observed HSI data and a spectral library of endmembers, the aim becomes to estimate the fractional abundances of endmembers in each mixed pixel, i.e. determine the percentage contribution of endmembers in the coverage area of every pixel. Hyperspectral unmixing can be challenging due to low spatial resolution, spectral mixtures which can happen at different scales, blurring caused by the data acquisition system and/or atmospheric conditions, potentially large spectral libraries and/or the lack of *a priori* knowledge regarding the endmembers present in the scene. Since the number of endmembers present in a pixel is very small compared to the growing dimensionality of spectral libraries it is often desirable to use optimization techniques to enforce sparsity constraints in the solution of the hyperspectral unmixing problem.

Several methods have been developed for hyperspectral unmixing; for a comprehensive overview see [21]. Iordache et. al [75] also include spatial information in the formulation of the sparse hyperspectral unmixing problem. In particular, the authors exploit the spatial-contextual information present in HSIs by using total variation (TV) regularization. If we incorporate a blurring operator in the problem formulation for hyperspectral unmixing then we have to solve a hyperspectral unmixing and deblurring problem. Zhao et.al. [147] have studied TV in sparse hypersectral unmixing and deblurring by incorporating a blurring operator for dealing with blurring effects. However, they assume that the blurring is homogenous, i.e. it is the same at all wavelengths. Nevertheless, as previously mentioned, it is a known fact that hyperspectral PSFs are generally wavelength dependent.

The HSI problems discussed in this work can be modeled as inverse problems.

The hyperspectral PSF identification problem is the inverse problem of determining a set of parameters which define the overall hyperspectral PSF whereas hyperpsectral unmixing and deblurring is the inverse problem of determining the fractional abundances of the spectrally pure components present in the mixed spectra of each pixel.

Therefore, we begin this thesis by discussing some computational approaches to compute approximate solutions of inverse problems. We consider two common models: a linear model and a separable nonlinear model. Several regularization approaches are described in Section 1.1.1. In Section 1.1.3 we discuss some considerations for preconditioning of ill-posed problems. For a more extended survey of these methods that also includes image iterative restoration methods for constrained and unconstrained problems, see [13].

1.1   Inverse Problems

Inverse problems arise in many areas of science and engineering, such as astronomy, astrophysics, medical imaging, parameter identification, geophysics, inverse scattering, heat conduction, and signal processing [46, 67, 69, 142, 3, 108]. A classical example of an inverse problem in imaging is the image restoration problem. The goal of this problem is to reconstruct an approximation of the true image from the observed image that is blurred and noisy. Moreover, in many cases it is also necessary to estimate the parameters of the blurring operator from the observed data.

The general forward model of an inverse problem can take the form

$$g = K(\phi_{\text{true}})f_{\text{true}} + \eta \tag{1.1}$$

where $g \in \mathcal{R}^m$ is a known (observed data) vector, $\mathbf{K}(\phi_{\text{true}}) \in \mathcal{R}^{m \times n}$ is an operator defined by a parameter vector $\phi_{\text{true}} \in \mathcal{R}^p$, $f_{\text{true}} \in \mathcal{R}^n$ is a vector representing the

true data, and $\boldsymbol{\eta} \in \mathcal{R}^m$ is unknown noise and error present in the observed data. In general, some of the tasks associated with the inverse problems of the above form include

- Given $\boldsymbol{K}(\cdot)$ and $\boldsymbol{g}$, find an approximation to $\boldsymbol{f}_{\text{true}}$.

- Given $\boldsymbol{f}_{\text{true}}$ (or an approximation of it) and $\boldsymbol{g}$, find an approximation to $\boldsymbol{\phi}_{\text{true}}$.

- Given $\boldsymbol{g}$, find an approximation to $\boldsymbol{f}_{\text{true}}$ and to $\boldsymbol{\phi}_{\text{true}}$.

In the case when the operator $\boldsymbol{K}(\cdot)$ can be represented as a matrix then $\boldsymbol{K}$ is called the coefficient matrix. The numerical treatment of inverse problems becomes more difficult when the coefficient matrix has a very large condition number, i.e. the problem we are trying to solve is very ill-conditioned. Many linear inverse problems can be characterized as rank-deficient or ill-posed problems [67].

If the coefficient matrix in the formulation of the inverse problem has a cluster of small singular values and there is a well defined gap between large and small singular values then the inverse problem is (numerically, or approximately) rank-deficient. In these type of problems one or more columns or rows of the coefficient matrix are nearly linear combinations of some or all of the remaining columns or rows. An approach for the numerical solution of rank-deficient problems consists of extracting the linearly independent information in the coefficient matrix, i.e. transforming the rank-deficient coefficient matrix into a more well-conditioned matrix.

Ill-posed problems are characterized by having a coefficient matrix whose singular values decay gradually to zero, lacking a well-defined gap between large and small singular values. This implies that there is no notion of numerical rank for the coefficient matrix. The definition of well-posed and ill-posed inverse problems goes back to Hadamard [63]. Hadamard defined a problem as ill-posed if a solution to the problem does not exist, or is not unique, or does not depend continuously on the data; that is, small changes in the data can lead to large changes in the solution. In general,

when computing approximate solutions of ill-posed problems, the goal is to find a balance between the norm of the residual, the norm of the solution, and/or *a priori* expectations for the computed solution.

The typical approach to solving ill-posed systems is to use numerical regularization methods which stabilize the solution by including appropriate additional information in the problem formulation. The approaches to solving both rank-deficient and ill-posed problems, share many of the same regularization algorithms. Many inverse problems are large scale and computationally demanding and thus the numerical regularization methods used to solve them have to be both reliable and efficient.

1.1.1   Regularization

In this section, we discuss regularization approaches for linear discrete inverse problems. Here, we discuss regularization for linear inverse problems in image restoration. So, without loss of generality, for now we assume that $\boldsymbol{K} \equiv \boldsymbol{K}(\boldsymbol{\phi})$ is known exactly (we return to the case where $\boldsymbol{K}(\boldsymbol{\phi})$ needs to be estimated in Section 1.1.2). The discrete forward model for the digital image restoration problem can be written as

$$\boldsymbol{g} = \boldsymbol{K}\boldsymbol{f}_{\text{true}} + \boldsymbol{\eta}\,, \tag{1.2}$$

where $\boldsymbol{f}_{\text{true}} \in \mathcal{R}^{mn}$ is the true object, $\boldsymbol{g} \in \mathcal{R}^{mn}$ is the observed image, $\boldsymbol{\eta} \in \mathcal{R}^{mn}$ is additive noise, $\boldsymbol{K} \in \mathcal{R}^{mn \times mn}$ is a matrix defined by the PSF which models the blurring operation, and the images are assumed to have $m \times n$ pixels. In the case of spatially invariant blur the image restoration problem is often referred to as deconvolution. Here we assume that $\boldsymbol{K}$ and $\boldsymbol{g}$ are known and the goal is to compute an approximation of $\boldsymbol{f}_{\text{true}}$. A major challenge, that arises when we want to compute an approximation to the true solution, is that the matrix $\boldsymbol{K}$ can be severely ill-conditioned.

1.1.1.1 SVD Analysis

The Singular Value Decomposition (SVD) can be used to analyze the discrete inverse solution of problem (1.2). Let $\boldsymbol{g}_{\text{true}} = \boldsymbol{K}\boldsymbol{f}_{\text{true}}$ be the noise-free blurred image. Then the SVD of $\boldsymbol{K}$ is defined as

$$\boldsymbol{K} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top,$$

where $\boldsymbol{K} \in \mathcal{R}^{N \times N}$, $\boldsymbol{U} \in \mathcal{R}^{N \times N}$ (whose columns $\boldsymbol{u}_i$ are the left singular vectors of $\boldsymbol{K}$) and $\boldsymbol{V} \in \mathcal{R}^{N \times N}$ (whose columns $\boldsymbol{v}_i$ are the right singular vectors of $\boldsymbol{K}$) are orthogonal matrices, $\boldsymbol{\Sigma} \in \mathcal{R}^{N \times N}$ is a diagonal matrix with entries satisfying $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_N \geq 0$, and $N = mn$. Discrete problems of the form (1.2) inherit the following properties:

- The singular values, σ_i of the coefficient matrix $\boldsymbol{K}$ decay to, and cluster at 0, without a significant gap to indicate numerical rank.

- Smaller singular values σ_i correspond with more oscillation in the singular vectors $\boldsymbol{v}_i$.

- The spectral components $|\boldsymbol{u}_i{}^\top \boldsymbol{g}_{\text{true}}|$, decay on average faster than the singular values σ_i. In particular, for large σ_i, we expect $|\boldsymbol{u}_i{}^\top \boldsymbol{g}_{\text{true}}| > |\boldsymbol{u}_i{}^\top \boldsymbol{\eta}|$ and therefore $\boldsymbol{u}_i{}^\top \boldsymbol{g}$ dominates. For small σ_i, $|\boldsymbol{u}_i{}^\top \boldsymbol{g}_{\text{true}}| \to 0$ as $\sigma_i \to 0$ and so $\boldsymbol{u}_i{}^\top \boldsymbol{\eta}$ dominates. This is called the *discrete Picard condition* [67].

Notice that if $\boldsymbol{K}$ is nonsingular then

$$\boldsymbol{K}^{-1} = \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^\top = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_N \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_1} & & & \\ & \frac{1}{\sigma_2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_N} \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_1{}^\top \\ \boldsymbol{u}_2{}^\top \\ \vdots \\ \boldsymbol{u}_N{}^\top \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_N \end{bmatrix} \begin{bmatrix} \frac{\boldsymbol{u}_1^\top}{\sigma_1} \\ \frac{\boldsymbol{u}_2^\top}{\sigma_2} \\ \vdots \\ \frac{\boldsymbol{u}_N^\top}{\sigma_N} \end{bmatrix} = \sum_{i=1}^{N} \frac{1}{\sigma_i} \boldsymbol{v}_i \boldsymbol{u}_i^\top .$$

Now we can use the previously mentioned properties to see the effect of the error term $\boldsymbol{\eta}$ on the inverse solution

$$\begin{aligned} \boldsymbol{f}_{\text{inv}} &= \boldsymbol{K}^{-1} \boldsymbol{g} = \boldsymbol{V} \boldsymbol{\Sigma}^{-1} \boldsymbol{U}^\top \boldsymbol{g} = \sum_{i=1}^{N} \frac{1}{\sigma_i} \boldsymbol{v}_i \boldsymbol{u}_i^\top \boldsymbol{g} = \sum_{i=1}^{N} \frac{\boldsymbol{u}_i^\top \boldsymbol{g}}{\sigma_i} \boldsymbol{v}_i = \sum_{i=1}^{N} \frac{\boldsymbol{u}_i^\top (\boldsymbol{g}_{\text{true}} + \boldsymbol{\eta})}{\sigma_i} \boldsymbol{v}_i \\ &= \sum_{i=1}^{N} \frac{\boldsymbol{u}_i^\top \boldsymbol{g}_{\text{true}}}{\sigma_i} \boldsymbol{v}_i + \sum_{i=1}^{N} \frac{\boldsymbol{u}_i^\top \boldsymbol{\eta}}{\sigma_i} \boldsymbol{v}_i = \boldsymbol{f}_{\text{true}} + \text{error}. \end{aligned}$$

The above relation shows that division by small σ_i magnifies the noise and the high frequencies in the error term. As a result, $\boldsymbol{f}_{\text{inv}}$ can be a very poor approximation of the true solution, $\boldsymbol{f}_{\text{true}}$. This analysis indicates that the incorporation of some form of *regularization* is necessary in order to compute an accurate and meaningful approximation of $\boldsymbol{f}_{\text{true}}$. Some common and well-known techniques used to enforce regularization include: Tikhonov regularization, Wiener filtering, and/or the incorporation of nonnegativity constraints [46, 67, 68, 142]. In the rest of this subsection, we review some of the regularization methods that are used throughout this thesis.

1.1.1.2 Regularization by Filtering

Filtering is one class of regularization methods that can be formulated as a modification of the inverse solution [67]. In regularization by filtering, instead of computing the naive inverse solution

$$\boldsymbol{f}_{\text{inv}} = \sum_{i=1}^{N} \frac{\boldsymbol{u}_i^\top \boldsymbol{g}}{\sigma_i} \boldsymbol{v}_i , \tag{1.3}$$

we compute the filtered solution

$$\boldsymbol{f}_{\text{filt}} = \sum_{i=1}^{N} \omega_i \frac{\boldsymbol{u}_i^\top \boldsymbol{g}}{\sigma_i} \boldsymbol{v}_i \,, \tag{1.4}$$

where $0 \leq \omega_i \leq 1$.

In the above definition, a decision has to made about what is meant by "large" and "small" singular values. The filtered solution formulation implies that the filter factors are used to dampen the components corresponding to the small singular values whereas the components of the solution corresponding to the large singular values are reconstructed. There exist different choices for the filter factors, ω_i, and these choices lead to different regularization methods. Some of the popular choices are the truncated SVD, Tikhonov, and Wiener filters [67, 78, 142]. Some examples for the choice of filter factors ω_i include

- **Truncated SVD filters:** $\omega_i = \begin{cases} 1 & \text{if } \sigma_i > \text{tol} \\ 0 & \text{if } \sigma_i \leq \text{tol} \end{cases}$, where tol is predetermined.

- **Tikhonov filters:** $\omega_i = \frac{\sigma_i^2}{\sigma_i^2 + \alpha^2}$, for some $\alpha \in [\sigma_{min}, \sigma_{max}]$.

- **Exponential Filters:** $\omega_i = 1 - e^{-\frac{\sigma_i^2}{\alpha^2}}$, where α is typically in the interval $[\sigma_{min}, \sigma_{max}]$.

1.1.1.3 Spectral Value Decomposition

If $\boldsymbol{K}$ is a normal matrix, the SVD can be replaced with the spectral value decomposition,

$$\boldsymbol{K} = \boldsymbol{Q}^H \boldsymbol{\Lambda} \boldsymbol{Q} \,,$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix, and $\boldsymbol{Q}^H$ is the Hermitian transpose of $\boldsymbol{Q}$, with $\boldsymbol{Q}^H \boldsymbol{Q} = \boldsymbol{I}$. The columns of $\boldsymbol{Q}$ are the eigenvectors and the diagonal elements of $\boldsymbol{\Lambda}$ are the eigenvalues of $\boldsymbol{K}$. If $\boldsymbol{K}$ has a spectral factorization, then it can be used, in place of the

SVD, to implement the filtering methods described above. The advantage is that it is sometimes more computationally convenient to compute a spectral value decomposition than it is to compute a singular value decomposition. We use this decomposition in this thesis in the case of spatially invariant blur with periodic boundary conditions. In this case the eigenvalues are obtained by taking a Fourier transform of the PSF, the eigenvectors of $\boldsymbol{K}$ are the Fourier vectors, and fast Fourier transforms (FFT) can be used for operations involving $\boldsymbol{Q}$ and $\boldsymbol{Q}^H$.

1.1.1.4 Variational Regularization

If it is not computationally feasible to use a singular value or spectral value decomposition based filtering method then we can use another class of regularization methods for linear inverse problems, which is known as *variational* regularization and has the form

$$\min_{\boldsymbol{f}} \left\{ \|\boldsymbol{g} - \boldsymbol{K}\boldsymbol{f}\|_2^2 + \alpha^2 \mathcal{J}(\boldsymbol{f}) \right\} , \tag{1.5}$$

The regularization operator $\mathcal{J}$ and the regularization parameter α must be chosen and the choice depends on the application. Some choices for the regularization operator, $\mathcal{J}$, include:

- Tikhonov regularization [61, 136, 135, 137]: $\mathcal{J}(\boldsymbol{f}) = \|\boldsymbol{L}\boldsymbol{f}\|_2^2$

- Total variation (TV) [32, 123, 142]: $\mathcal{J}(\boldsymbol{f}) = \left\| \sqrt{(\boldsymbol{D}_{\mathrm{h}}\boldsymbol{f})^2 + (\boldsymbol{D}_{\mathrm{v}}\boldsymbol{f})^2} \right\|_1$

  or $\mathcal{J}(\boldsymbol{f}) = \left\| \sqrt{(\boldsymbol{D}_{\mathrm{h}}\boldsymbol{f})^2 + (\boldsymbol{D}_{\mathrm{v}}\boldsymbol{f})^2} \right\|_2$

- Sparsity constraints [29, 51, 140]: $\mathcal{J}(\boldsymbol{f}) = \|\boldsymbol{\Phi}\boldsymbol{f}\|_1.$

For Tikhonov regularization, $\boldsymbol{L}$ is typically chosen to be the identity matrix, or a discrete approximation to a derivative operator, such as the Laplacian. For TV, the matrices $\boldsymbol{D}_{\mathrm{h}}$ and $\boldsymbol{D}_{\mathrm{v}}$ represent discrete approximations of horizontal and vertical derivatives of the 2D image $\boldsymbol{f}$, and the approach extends to 3D images in a straight-

forward way. The 1-norm definition of TV is known as anisotropic TV whereas the 2-norm definition is referred to as isotropic TV. Implementing total variation in an efficient and stable way is nontrivial for some problems [32, 142]. For sparsity constraints, the matrix $\boldsymbol{\Phi}$ represents a basis in which the image, $\boldsymbol{f}$, is sparse and its choice depends on the structure of the image $\boldsymbol{f}$. There is a lot of ongoing research on the usage of sparsity constraints for regularization [51].

1.1.1.5   Iterative Regularization

Another class of regularization methods is based on iterative regularization. Iterative regularization is enforced using iterative methods which can be applied directly to the least squares problem

$$\min_{\boldsymbol{f}} \|\boldsymbol{g} - \boldsymbol{K}\boldsymbol{f}\|_2.$$

However, we have to be careful about the number of iterations used to run the iterative methods. If we apply an iterative method directly to the above least squares problem then the iteration will converge to the naive inverse solution, $\boldsymbol{f}_{\mathrm{inv}} = \boldsymbol{K}^{-1}\boldsymbol{g}$, which is typically a poor approximation of $\boldsymbol{f}_{\mathrm{true}}$. In cases where the inverse problem satisfies the Picard condition, many iterative methods exhibit a *semi-convergence* behavior with respect to the relative error,

$$\frac{\|\boldsymbol{f}_k - \boldsymbol{f}_{\mathrm{true}}\|_2}{\|\boldsymbol{f}_{\mathrm{true}}\|_2},$$

where $\boldsymbol{f}_k$ is the approximate solution at the kth iteration. The semi-convergence behavior occurs when in the early iterations the relative error begins to decrease and, after some "optimal" iteration, the error then begins to increase. The reason for this behavior is that at the early iterations the components of the solution corresponding to large singular values are reconstructed and then after the "optimal" iteration the components corresponding to small singular values are reconstructed. Thus, the com-

puted solution is dominated by noise and error. If we terminate the iteration when the error is low, then we obtain a regularized approximation of the solution.

Some of the advantages of using iterative methods include: efficiency (for spatially invariant as well as spatially variant blurs), the ability to incorporate a variety of regularization techniques and boundary conditions, and the ability to incorporate additional constraints, such as nonnegativity [10, 104]. Iterative regularization methods have the disadvantage that it might be difficult to determine an "optimal" stopping iteration, since the relative errors cannot be computed without knowing the true solution. However, it is possible to incorporate a regularization operator using the variational approach, and apply an iterative method to the resulting minimization problem. In this sense, iterative methods have many advantages over simple filtering techniques [14, 86, 88].

1.1.1.6 Regularization Parameter

In each of the regularization methods discussed above, we must chose a *regularization parameter*. In order to choose the "right" parameter, we often need to make assumptions about the statistical distribution of the noise. Some approaches assume additional prior information such as an estimate of $||\boldsymbol{\eta}||_2^2$. Other approaches attempt to extract noise information directly from the measured data.

In the variational form, one has to choose the scalar α, where as for iterative regularization methods the index where the iteration is terminated plays the role of the regularization parameter. In either case, it is a nontrivial matter to choose an "optimal" regularization parameter. Some methods for choosing regularization parameters include: generalized cross validation (GCV) [58], discrepancy principle [46], L-curve [67], L-ribbon [24], and many others [46, 67, 142]. However, in many cases it may be necessary to solve the problem for a variety of parameters, and use knowledge of the application to help decide which solution is best.

1.1.2 Separable Nonlinear Inverse Problems

In this section, we consider the inverse problem (1.1) when the matrix $\boldsymbol{K}(\boldsymbol{\phi})$ is not known exactly, and the goal is to compute approximations for both $\boldsymbol{\phi}$ and $\boldsymbol{f}$. We assume that it is possible to obtain initial guesses of the unknowns $\boldsymbol{\phi}$ and $\boldsymbol{f}$. For this problem, if we consider the variational form of the standard Tikhonov regularization then a nonlinear least squares problem can be formulated as

$$\min_{\boldsymbol{f},\boldsymbol{\phi}} \left\{ \|\boldsymbol{g} - \boldsymbol{K}\left(\boldsymbol{\phi}\right)\boldsymbol{f}\|_2^2 + \alpha^2 \|\boldsymbol{f}\|_2^2 \right\} = \min_{\boldsymbol{f},\boldsymbol{\phi}} \left\| \begin{bmatrix} \boldsymbol{g} \\ \boldsymbol{0} \end{bmatrix} - \begin{bmatrix} \boldsymbol{K}\left(\boldsymbol{\phi}\right) \\ \alpha\boldsymbol{I} \end{bmatrix} \boldsymbol{f} \right\|_2^2 \tag{1.6}$$

The above equation can be rewritten as

$$\min_{\boldsymbol{z}} \psi(\boldsymbol{z}) = \min_{\boldsymbol{z}} \|\zeta(\boldsymbol{z})\|_2^2 \tag{1.7}$$

where

$$\zeta(\boldsymbol{z}) = \zeta\left(\boldsymbol{f},\boldsymbol{\phi}\right) = \begin{bmatrix} \boldsymbol{g} \\ \boldsymbol{0} \end{bmatrix} - \begin{bmatrix} \boldsymbol{K}\left(\boldsymbol{\phi}\right) \\ \alpha\boldsymbol{I} \end{bmatrix} \boldsymbol{f}, \quad \boldsymbol{z} = \begin{bmatrix} \boldsymbol{f} \\ \boldsymbol{\phi} \end{bmatrix}$$

Some approaches for solving this problem include the fully coupled joint optimization approach, the partially coupled approach (variable projection), and the decoupled approach (the block coordinate descent). Here, we discuss the former two approaches in more detail and briefly touch upon the decoupled approach.

1.1.2.1 Solution Methods

Separable nonlinear inverse problems can be solved using an iterative method such as the Gauss-Newton method [41, 81, 109, 111]. In each iteration of the Gauss-Newton algorithm the solution vector is updated in the form

$$\boldsymbol{z}_{k+1} = \boldsymbol{z}_k + \boldsymbol{d}_k \,,$$

with the step direction given by

$$\boldsymbol{d}_k = - \left(\hat{\psi}''(\boldsymbol{z}_k) \right)^{-1} \psi'(\boldsymbol{z}_k)$$

where $\hat{\psi}''$ is an approximation of the Hessian ψ''.

The first derivative of the cost function $\psi(\cdot)$ is given by $\psi'(\boldsymbol{z}_k) = \boldsymbol{J}_\psi{}^\top \zeta$, and ψ'' is is approximated by $\hat{\psi}'' = \boldsymbol{J}^\top{}_\psi \boldsymbol{J}_\psi$, where

$$\boldsymbol{J}_\psi = \left[\frac{\partial \zeta\left(\boldsymbol{f}, \boldsymbol{\phi}\right)}{\partial \boldsymbol{f}} \quad \frac{\partial \zeta\left(\boldsymbol{f}, \boldsymbol{\phi}\right)}{\partial \boldsymbol{\phi}} \right],$$

is the Jacobian matrix. Let $\boldsymbol{r} = \begin{bmatrix} \boldsymbol{g} \\ \boldsymbol{0} \end{bmatrix} - \begin{bmatrix} \boldsymbol{K}\left(\boldsymbol{\phi}\right) \\ \alpha \boldsymbol{I} \end{bmatrix} \boldsymbol{f}$. Then the search direction $\boldsymbol{d}_k$ at each Gauss-Newton iteration is computed by solving a least squares problem of the form

$$\min_{\boldsymbol{d}} ||\boldsymbol{J}_\psi \boldsymbol{d} - \boldsymbol{r}||_2^2 \tag{1.8}$$

In general, a Gauss-Newton algorithm to solve problem (1.7) has the form:

General Gauss-Newton Algorithm

===

given: $\boldsymbol{g}$

choose: initial guess for $\boldsymbol{z}_0 = \begin{bmatrix} \boldsymbol{f}_0 \\ \boldsymbol{\phi}_0 \end{bmatrix}$

step length parameter τ

for $k = 1, 2, \cdots$, until stop

compute $\boldsymbol{r}_k = \begin{bmatrix} \boldsymbol{g} \\ \boldsymbol{0} \end{bmatrix} - \begin{bmatrix} \boldsymbol{K}(\boldsymbol{\phi}_k) \\ \alpha\boldsymbol{I} \end{bmatrix} \boldsymbol{f}_k$

solve

$$\boldsymbol{d}_k = \min_{\boldsymbol{d}} \left\|\boldsymbol{J}_\psi \boldsymbol{d} - \boldsymbol{r}_k\right\|_2^2$$

where

$$\boldsymbol{J}_\psi = \begin{bmatrix} \dfrac{\partial \zeta(\boldsymbol{f}_k, \boldsymbol{\phi}_k)}{\partial \boldsymbol{f}_k} & \dfrac{\partial \zeta(\boldsymbol{f}_k, \boldsymbol{\phi}_k)}{\partial \boldsymbol{\phi}_k} \end{bmatrix}$$

set $\boldsymbol{z}_{k+1} = \boldsymbol{z}_k + \tau_k \mathbf{d}_k$, where τ_k has to be chosen.

end for

A computational challenge associated with the general Gauss-Newton approach consists in constructing and solving systems of the form (1.8). This can be computationally expensive because of the size of $\boldsymbol{J}_\psi$ and it also might require some form of regularization when $\boldsymbol{J}_\psi$ is ill-conditioned or rank deficient. Moreover, effective preconditioners for (1.8) may be difficult to find and the regularization parameter α has to be either specified a priori or estimated within each iteration.

The general Gauss-Newton method is a fully coupled approach for solving nonlinear inverse problems. A key disadvantage of these type of approaches consists in not taking algorithmic advantage of the fact that the objective function is linear in

$\boldsymbol{f}$. These type of approaches can be very slow because the step size at each iteration can be very small due to the nonlinearity induced by $\boldsymbol{\phi}$.

Another approach for solving problem (1.6) is the block coordinate descent algorithm. This iterative algorithm fully decouples the problem by solving two minimization problems in each iteration. The first minimization problem consists of solving a linear least squares problem for $\boldsymbol{f}$ using a fixed initial guess for $\boldsymbol{\phi}$. In the second minimization problem, the variable $\boldsymbol{f}$ obtained in the first problem is fixed and a new $\boldsymbol{\phi}$ is obtained by solving a nonlinear least squares problem. Thus, the coordinate descent method explicitly separates the variables $\boldsymbol{f}$ and $\boldsymbol{\phi}$. A major drawback of this approach is slow convergence in problems with tightly coupled variables [109].

A third approach to solve the separable nonlinear least squares problem in Equation (1.6) is the variable projection method [56, 57, 80, 112, 124]. Notice that the nonlinear least squares problem (1.6) is linear in $\boldsymbol{f}$ and nonlinear in $\boldsymbol{\phi}$. Morever, in most cases $\boldsymbol{\phi}$ contains relatively few parameters compared to $\boldsymbol{f}$. The variable projection method exploits this structure of the problem by implicitly eliminating the linear parameters $\boldsymbol{f}$ and obtaining a reduced cost functional which is dependent only on the nonlinear term $\boldsymbol{\phi}$. The reduced cost functional is then minimized using the Gauss-Newton method. In particular, let $\boldsymbol{f}(\boldsymbol{\phi})$ be the solution of the least squares problem

$$\min_{\boldsymbol{f}} \psi\left(\boldsymbol{f}, \boldsymbol{\phi}\right) = \min_{\boldsymbol{f}} \left\| \begin{bmatrix} \boldsymbol{g} \\ \boldsymbol{0} \end{bmatrix} - \begin{bmatrix} \boldsymbol{K}\left(\boldsymbol{\phi}\right) \\ \alpha \boldsymbol{I} \end{bmatrix} \boldsymbol{f} \right\|_2^2 . \tag{1.9}$$

Now, consider the minimization problem for the reduced cost functional

$$\min_{\boldsymbol{\phi}} \upsilon(\boldsymbol{\phi}) = \min_{\boldsymbol{\phi}} \psi\left(\boldsymbol{f}(\boldsymbol{\phi}), \boldsymbol{\phi}\right)$$

To apply the Gauss-Newton algorithm to minimize the reduced cost functional we

need the first derivate of $\upsilon(\boldsymbol{\phi})$

$$\upsilon'(\boldsymbol{\phi}) = \frac{d\boldsymbol{f}(\boldsymbol{\phi})}{d\boldsymbol{\phi}} \frac{\partial \psi}{\partial \boldsymbol{f}(\boldsymbol{\phi})} + \frac{\partial \psi}{\partial \boldsymbol{\phi}}$$

Since $\boldsymbol{f}$ is the solution to the linear problem (1.9) then $\dfrac{\partial \psi}{\partial \boldsymbol{f}(\boldsymbol{\phi})} = 0$. Thus

$$\upsilon'(\boldsymbol{\phi}) = \frac{\partial \psi}{\partial \boldsymbol{\phi}} = \mathbf{J}_\upsilon{}^\top \zeta$$

The Jacobian of the reduced cost functional can be written as

$$\boldsymbol{J}_\upsilon = \frac{\partial}{\partial \boldsymbol{\phi}} [\boldsymbol{K}(\boldsymbol{\phi})\boldsymbol{f}] .$$

The Jacobian matrix $\boldsymbol{J}_\upsilon$ has to be evaluated analytically or it can be approximated by a finite-difference approach.

In general, the variable projection Gauss-Newton algorithm has the form:

<div style="border:1px solid black; padding:1em;">

Variable Projection Gauss-Newton Algorithm

given: $\boldsymbol{g}$

choose: initial guess $\boldsymbol{f}_0$

        step length parameter τ

        initial guess for parameter vector $\boldsymbol{\phi}_0$

for $k = 1, 2, \cdots$,   until stop

     choose α_k

     compute $\boldsymbol{r}_k = \begin{bmatrix} \boldsymbol{g} \\ \boldsymbol{0} \end{bmatrix} - \begin{bmatrix} \boldsymbol{K}(\boldsymbol{\phi}_k) \\ \alpha \boldsymbol{I} \end{bmatrix} \boldsymbol{f}_k$

     solve

$$\boldsymbol{f}_k = \arg\min_{\boldsymbol{f}} \left\| \begin{bmatrix} \boldsymbol{g} \\ \boldsymbol{0} \end{bmatrix} - \begin{bmatrix} \boldsymbol{K}(\boldsymbol{\phi}_k) \\ \alpha_k \boldsymbol{I} \end{bmatrix} \boldsymbol{f} \right\|_2$$

$$\boldsymbol{d}_k = \arg\min_{\boldsymbol{d}} \left\| \boldsymbol{J}_v \boldsymbol{d} - \boldsymbol{r}_k \right\|_2$$

     where

$$\boldsymbol{J}_v = \frac{\partial}{\partial \boldsymbol{\phi}} [\boldsymbol{K}(\boldsymbol{\phi}) \boldsymbol{f}_k]$$

     set $\boldsymbol{\phi}_{k+1} = \boldsymbol{\phi}_k + \tau_k \boldsymbol{d}_k$, where τ_k has to be chosen.

end for

</div>

It should be noted that the variable projection method is usually the best method for solving separable nonlinear least squares problems. However, a drawback is that the derivates may be difficult to compute.

1.1.3 Preconditioning

Preconditioning techniques are used in many areas of scientific computation to accelerate convergence and improve the reliability of iterative methods. Much effort has been put in the development of effective preconditionining techniques for the solution of well-posed problems [12, 60, 125]. In the case of the preconditioning of ill-posed inverse problems special considerations have to be taken into account to avoid convergence to a poor approximate solution. Preconditioned iterative regularization schemes for overcoming the difficulties associated with ill-posed problems in image deblurring have been proposed: for applications where the blurring operator is known [65, 66, 106], for image restoration problems with spatially variant blurs [102, 103, 105], and multi-frame problems [30]. In general, the additional cost associated with constructing and using preconditioners in iterative methods is acceptable and justified if the overall number of iterations is reduced substantially and/or the preconditioner can be reused or updated easily throughout the iterations.

Preconditioning is a technique that attempts to improve certain spectral properties of the matrix $\boldsymbol{K}$, which in general dictate the speed of convergence of iterative methods. Preconditioning is often presented in the context of solving linear systems of the form

$$\boldsymbol{K}\boldsymbol{f} = \boldsymbol{g}.$$

Preconditioning refers to the process of transforming the above linear systems into other systems with more favorable spectral properties for accelerating convergence of the iterative solution. Suppose $\boldsymbol{R}$ is a nonsingular matrix. Then instead of applying the iterative method to the above linear system, we apply it to a modified system $\hat{\boldsymbol{K}}\hat{\boldsymbol{f}} = \hat{\boldsymbol{g}}$, where

$$\text{Right preconditioning:} \quad \hat{\boldsymbol{K}} = \boldsymbol{K}\boldsymbol{R}^{-1}, \quad \hat{\boldsymbol{f}} = \boldsymbol{R}\boldsymbol{f}, \quad \hat{\boldsymbol{g}} = \boldsymbol{g}$$

$$\text{Left preconditioning:} \quad \hat{\boldsymbol{K}} = \boldsymbol{R}^{-\top}\boldsymbol{K}, \quad \hat{\boldsymbol{f}} = \boldsymbol{f}, \quad \hat{\boldsymbol{g}} = \boldsymbol{R}^{-\top}\boldsymbol{g}$$

In the implementation of iterative methods, the most common operations involved are matrix-vector multiplications with $\hat{\boldsymbol{K}}$. Hence we would like to construct a matrix, $\boldsymbol{R}$, that satisfies the following properties:

- It should be relatively inexpensive to construct $\boldsymbol{R}$. In general, this is a one time cost since it is common that the coefficient matrix is the same or varies slowly throughout the iterations.

- It should be relatively inexpensive to solve linear systems of the form $\boldsymbol{R}\boldsymbol{z} = \boldsymbol{w}$ and $\boldsymbol{R}^{\top}\boldsymbol{z} = \boldsymbol{w}$. Solving linear system with $\boldsymbol{R}$ and $\boldsymbol{R}^{\top}$ is usually required at each iteration.

- The preconditioned system should satisfy $\hat{\boldsymbol{K}} = \boldsymbol{K}\boldsymbol{R}^{-1} \approx \boldsymbol{I}$ or $\hat{\boldsymbol{K}} = \boldsymbol{R}^{-\top}\boldsymbol{K} \approx \boldsymbol{I}$. This property usually determines the speed of convergence.

1.1.3.1 Considerations for Ill-posed Inverse Problems

In general, we want to design a preconditioner such that $\hat{\boldsymbol{K}} \approx \boldsymbol{I}$, i.e. the singular values of $\hat{\boldsymbol{K}}$ are clustered at 1. Note that this implies that $\boldsymbol{R} \approx \boldsymbol{K}$ or $\boldsymbol{R}^{\top} \approx \boldsymbol{K}^{\top}$. Solving linear systems with $\boldsymbol{R}$ and $\boldsymbol{R}^{\top}$ can magnify the noise. Thus, this approach does not work well for ill-posed problems. In particular, if we apply an iterative method directly to the unregularized least squares problems

$$\min_{\boldsymbol{f}} \|\boldsymbol{g} - \boldsymbol{K}\boldsymbol{f}\|_2 \,,$$

as in the case of iterative regularization, then clustering all singular values around 1 will likely lead to very poor reconstructions. Recall that the SVD analysis outlined in Section 1.1.1 suggests that the large singular values correspond to signal information we want to reconstruct, while small singular values correspond to noise information that we do not want to reconstruct. Clustering all singular values around 1 mixes the

signal and noise information together. Thus, it is impossible for the iterative method to distinguish between signal information and noise information. Hence we get fast convergence to the noisy inverse solution. We can apply two possible remedies to avoid this problem:

- Apply the iterative method to the Tikhonov regularized least squares problem

$$\min_{\boldsymbol{f}} \left\| \begin{bmatrix} \boldsymbol{g} \\ \boldsymbol{0} \end{bmatrix} - \begin{bmatrix} \boldsymbol{K} \\ \alpha \boldsymbol{L} \end{bmatrix} \boldsymbol{f} \right\|_2$$

  Here, we first have to choose α and $\boldsymbol{L}$ and then find a preconditioner for $\begin{bmatrix} \boldsymbol{K} \\ \alpha \boldsymbol{L} \end{bmatrix}$.

- Choose $\boldsymbol{R}$ so that only the large singular values of $\hat{\boldsymbol{K}}$ are clustered at 1. If we denote the singular values of $\hat{\boldsymbol{K}}$ by $\hat{\sigma}_i$ and the singular values of $\boldsymbol{K}$ by σ_i then in this case

$$\hat{\sigma}_i = \begin{cases} 1 & \text{for } i = 1, \cdots, l \\ \sigma_i & \text{for } i = l+1, \cdots, N \end{cases}.$$

Another consideration we have to take into account when applying a preconditioner to an ill-posed inverse problem is the statistical characteristics of the problem. Consider the inverse problem

$$\boldsymbol{g} = \boldsymbol{K}\boldsymbol{f} + \boldsymbol{\eta}.$$

Then, after applying left preconditioning, the problem is modified as

$$\boldsymbol{R}^{-\top}\boldsymbol{g} = \boldsymbol{R}^{-\top}\boldsymbol{K}\boldsymbol{f} + \boldsymbol{R}^{-\top}\boldsymbol{\eta}.$$

From the above relation, we can see that the statistical characteristics of the data and noise of the original problem are modified after applying the left preconditioner, $\boldsymbol{R}$.

Therefore, care must be taken especially if the iterative method assumes particular statistical distribution of data and noise.

1.2 Outline of Work

In the rest of this work, we use some of the approaches discussed in this chapter to solve inverse problems in HSI. In particular, we use a regularized separable nonlinear least squares framework and a conjugate gradient preconditioner for our work on hyperpsectral imaging. We begin by giving a brief overview of hyperpsectral imaging in Chapter 2.

In Chapter 3, we consider a hyperspectral PSF parameter identification problem. We explain in detail our formulation of the problem and show our approaches for finding the PSF parameters. Several regularization methods are used to solve the Jacobian system within the Gauss-Newton iteration. In particular, we solve the nonlinear system using a variable projection Gauss-Newton method and we solve the Jacobian using different regularization schemes such as subset selection [59], Tikhonov regularization, and truncated SVD.

Chapter 4 considers a hyperspectral unmixing and deblurring problem using wavelength varying PSFs. We use the estimated hyperpsectral PSFs from Chapter 3 in our numerical scheme for hyperspectral unmixing and deblurring with multiple PSFs. We present our formulation of the problem and we solve the problem using an Alternating Direction Method of Multipliers (ADMM) with sparsity constraints and TV regularization. One of the subproblems in the ADMM method for the multiple PSF case is solved using a preconditioned conjugate gradient method.

In Chapter 5 we show the numerical results of our approaches for solving the hyperspectral PSF parameter identification problem and the sparse hyperspectral

unmixing and deblurring problem. Some concluding remarks are provided in Chapter 6.

1.3   Contributions

The work presented in this dissertation describes a number of contributions as follows:

- Hyperspectral PSF parameter estimation:

  - We have used a regularized variable projection Gauss-Newton method for solving the hyperspectral PSF parameter identification problem. Our approach allows for a **joint estimation approach without prior estimation** of the noise variance and star spectrum.

  - We have applied **subset selection with Tikhonov regularization** to improve the conditioning of the Jacobian system for finding the hyperspectral parameters for the elliptical Moffat PSF.

  - We have provided a **theoretical analysis** for the conditioning of the Jacobian matrix.

  - We have implemented our approach in MATLAB.

- Hyperspectral unmixing and deblurring:

  - We have considered the case when ground-based HSIs of objects are taken through the atmosphere at **multiple wavelengths** with narrow spectral channels.

  - We have **exploited properties of Kronecker product** to formulate and solve the problem for the multiple PSF case.

  - We have shown that the number of iterations can be significantly reduced by using a simple **conjugate gradient preconditioner** to solve an important subproblem in the ADMM method.

- – We have verified the accuracy and effectiveness of our method with numerical experiments on a simulated hyperspectral image of the Hubble Space Telescope satellite.

- Other work we have done, which is not explicitly described in this thesis, relates to software development for the material described in Chapter 1 (see [13] for a full detailed discussion):

  - – We have described a variety of iterative methods that can be used for image restoration.

  - – We have compared their performance on a variety of test problems.

  - – We have described preconditioning techniques that can be used to accelerate convergence.

  - – We have provided MATLAB software implementing image restoration methods. The software can be found at http://www.mathcs.emory.edu/~nagy/RestoreTools.

  - – We have implemented the iterative methods in C++ utilizing the Trilinos framework.

Chapter 2

Hyperspectral Imaging

Hyperspectral remote sensing is a technology that extracts information from objects on the Earth's surface, or in orbit around the Earth, based on the object's measured radiance at many narrow spectral bands. As a result of recording the radiance of the scene at different wavelengths of light, multiple images of the same scene are created over continuous narrow spectral bands. The product is the so-called hyperspectral image (HSI). Hyperspectral remote sensing has been developed significantly and, thus, the availability of hyperspectral data has greatly increased in the last decade. The increased ability of hyperspectral sensors to collect data at hundreds or even thousands of wavelengths has dramatically increased precision in image classification tasks. Hyperspectral imaging is used in a broad range of commercial and government related applications, such as remote monitoring of the environment, food safety, remote urban monitoring, pharmaceutical process monitoring, agriculture, mineralogy, chemical and oils spills, remote surveillance, astrophysics, medicine, and many other security and defense areas [45, 121, 33, 94, 73].

In this chapter we explain briefly some basic concepts in hyperspectral imaging (HSI). We start by discussing the electromagnetic spectrum and wavelengths in Section 2.1. The structure of the HSI data is described in Section 2.2. In Section 2.3

we provide a brief explanation of how the HSI data are collected by imaging spectrometers. The key feature of HSI data, the spectral reflectance or spectral signature, is explained in Section 2.4. Some factors that contribute to the mixing of the observed spectra are given in Section 2.5. Hyperpsectral mixing models are discussed in Section 2.6.

2.1 Electromagnetic Spectrum

Hyperspectral remote sensing devices measure the interaction of objects with electromagnetic energy. According to wave theory, electromagnetic radiation is energy that propagates in the form of electromagnetic waves. Electromagnetic waves are vibrations of magnetic and electric fields that are both perpendicular to the direction of propagation and travel at the speed of light. Electromagnetic waves can be characterized in terms of wavelength, frequency, and energy. A wavelength is the distance between any two points of the electromagnetic wave before the shape of the wave repeats. Wavelengths are inversely proportional to frequencies. Another theory, namely particle theory, defines the electromagnetic radiation as energy composed of discrete units called photons or quanta.
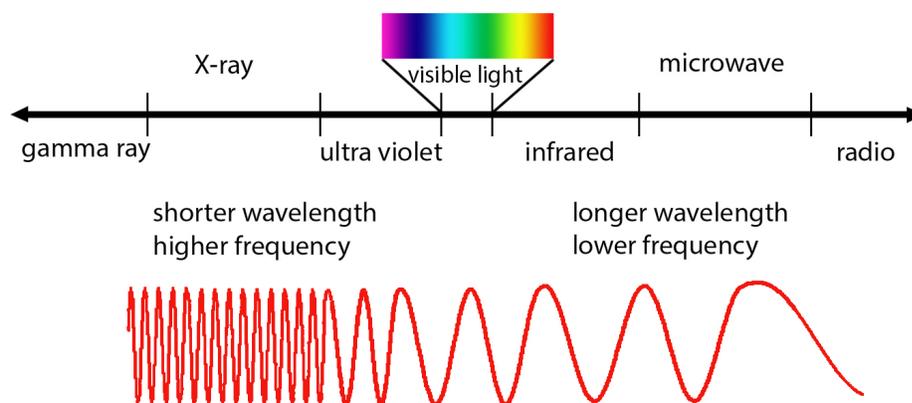
Figure 2.1: The electromagnetic spectrum.

The energy of a photon is directly proportional to its frequency and inversely proportional to its wavelength. The particle theory gives more insight into the interaction of electromagnetic radiation with the matter [1]. The electromagnetic spectrum represents the entire family of electromagnetic radiation. It can be generally classified into cosmic rays, gamma-rays, X-rays, ultraviolet light, visible light, infrared light, microwaves, radio waves, and terahertz radiation. In principle, the range of the electromagnetic spectrum is infinite and continuous. Figures 2.1 and 2.2 illustrate the electromagnetic spectrum and wavelength.

one wavelength

$$\text{Speed} = \text{Frequency} \times \text{Wavelength}$$
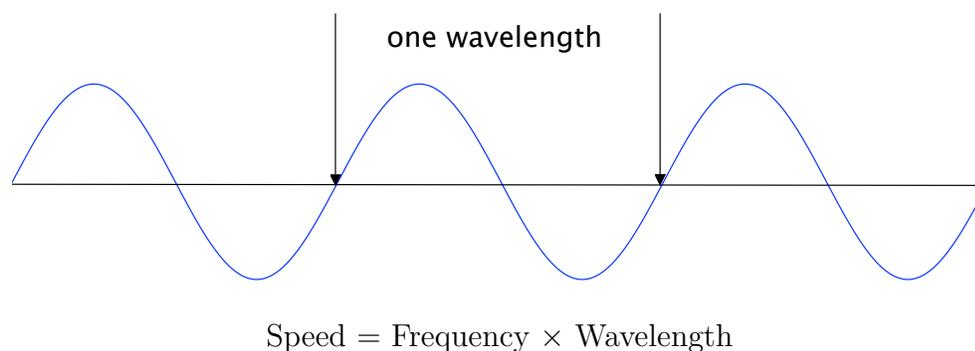
Figure 2.2: Wavelength illustration.

The visible wavelength region is the first *significant window* in which the energy can significantly pass through the atmosphere. This is the light which is detected by our eyes. The primary colors of the visible wavelength region are red, green, and blue. The blue wavelengths are substantially attenuated by atmospheric scattering. The atmosphere and glass lenses also cause significant absorption and/or scattering of the very short wavelengths such as ultraviolet. There is much less interference by the atmosphere for the green, red, and near-infrared wavelengths.

Different objects interact in different ways with electromagnetic radiation and thus an object's response to electromagnetic radiation can provide significant clues about

its chemical and physical composition. For example, the near-infrared wavelengths provide important clues for the structure of plant leaves. In geological applications, the middle infrared wavelengths are very useful. The thermal region of the electromagnetic spectrum has been proven to be useful in monitoring: the spatial distribution of heat from industrial activity, fires, and soil moisture conditions. The microwave region of the spectrum is very useful in environmental remote sensing, specifically in active radar imaging. This region of the spectrum is also usable at night and it is able to penetrate through the clouds. The radio region of the spectrum is used in meteorology and commercial broadcasting. For a more detailed discussion see [44, 1]. Thus, combination of the information gained at different wavelengths significantly increases the ability to discriminate different materials.

2.2   Structure of HSI Data

Traditional imaging systems collect imaging data at one panchromatic band or at three color bands (red, green, blue) covering the visible electromagnetic spectrum [128]. In spectral imaging the data is collected at many spectral bands or wavelengths across the entire electromagnetic spectrum. Multispectral imaging systems acquire image data in a few and relatively broad wavelength bands [131]. In HSI, the images are acquired at hundreds or even thousands of narrow adjacent spectral bands. The most important characteristic that distinguishes HSI from other imaging systems consists of the fact that it associates a *continuous* spectrum to each pixel.

Spectral imaging data represents measured intensities of the electromagnetic radiation reflected from the materials present in an imaged scene. The recorded reflectance intensities vary with respect to different wavelengths across the spectral range. Spectral data can be visualized as a cube of images, where each slice of the cube consists of an image of the same scene corresponding to a particular wavelength.
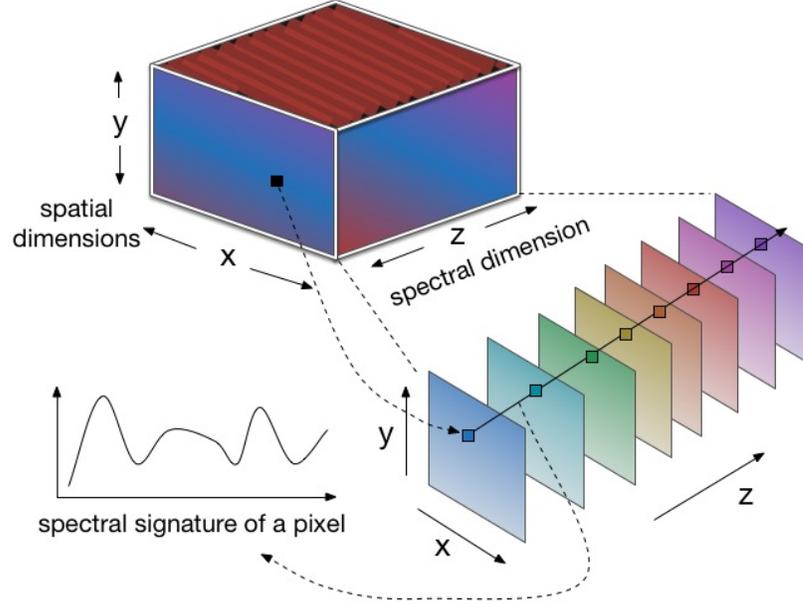
Figure 2.3: Structure of a hyperspectral data cube and spectral signature of a pixel. The measured images at different wavelengths are stacked to form a 3D hyperpsectral data cube with spatial information in the x and y dimensions and spectral information across the z dimension. The spectral vectors along the z axis represent the spectral signatures of each pixel.

Let $\boldsymbol{S}$ denote the spectral cube. Then $\boldsymbol{S}_{ijk}$ is the measured reflection at the kth wavelength of the pixel at the (i, j) spatial position. Thus, each vertical slice of the spectral cube along the z axis contains the reflection intensities of all pixels in the image at a particular wavelength, while each horizontal slice of the spectral cube along the y axis contains the reflections of a row of pixels of the image at all wavelengths.

If we denote the reflection intensities of a pixel (i, j) at N_w wavelengths by a vector $\boldsymbol{s}_{(ij)} = \begin{bmatrix} \boldsymbol{s}_{ij}(\lambda_1) & \boldsymbol{s}_{ij}(\lambda_2) & \cdots & \boldsymbol{s}_{ij}(\lambda_{N_w}) \end{bmatrix}^T$ then the vector $\boldsymbol{s}_{(ij)}$ is often referred to as the spectral signature of the pixel corresponding to the (i, j) spatial position. Thus, spectral images not only represent observations of a scene at different wavelengths but also associate to each pixel of the imaged scene a full spectral vector or spectral signature. For comprehensive discussions see [21, 115, 129] and the references therein. Figure 2.3 illustrates the concept of a hyperpsectral data cube and the spectral signature of a pixel. Figure 2.4 shows an example of a hyperspectral data cube. For

HSI data cube   Image at a single wavelength

Spectra for a single pixel
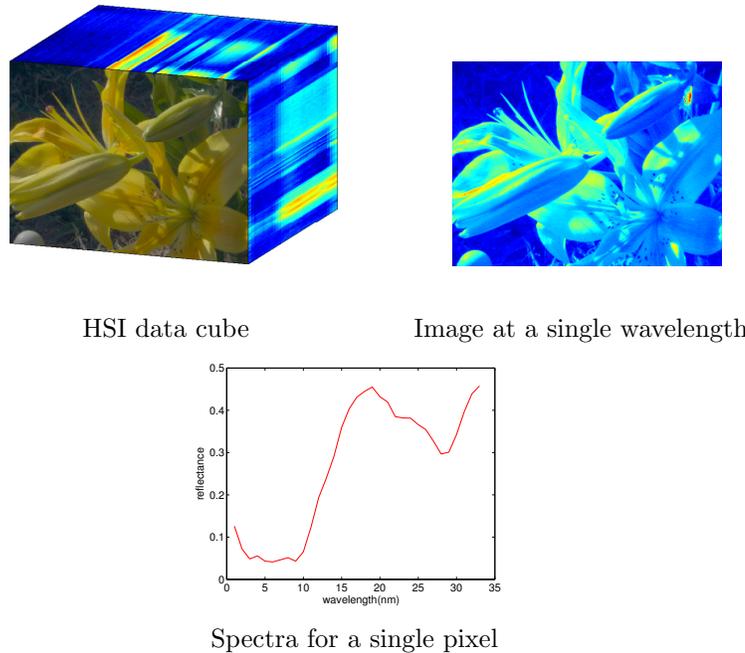
Figure 2.4: Visual depiction of an HSI datacube. The images and the HSI data in this figure were obtained from the online tutorial: `http://personalpages.manchester.ac.uk/staff/david.foster/Tutorial_HSI2RGB/Tutorial_HSI2RGB.html`.

illustration purposes the front image of the datacube is a false color image created by transforming hyperspectral image reflectances into an RGB color image.

2.3   Measuring HSI Data

Hyperspectral data are measured by imaging sensors called spectrometers. Spectrometers have been developed as a result of the combination of two technologies: spectroscopy and remote sensing. Spectroscopy studies light which is emitted by or reflected from materials as a function of wavelength [131]. Optical remote sensing is an imaging technology in which information about an object or a scene is acquired without coming into physical contact with the object or scene [128]. A remote sensing system consists of three main components: the scene, the sensor, and the processing

algorithms. In this section we focus on the sensors used to measure spectral imaging data.
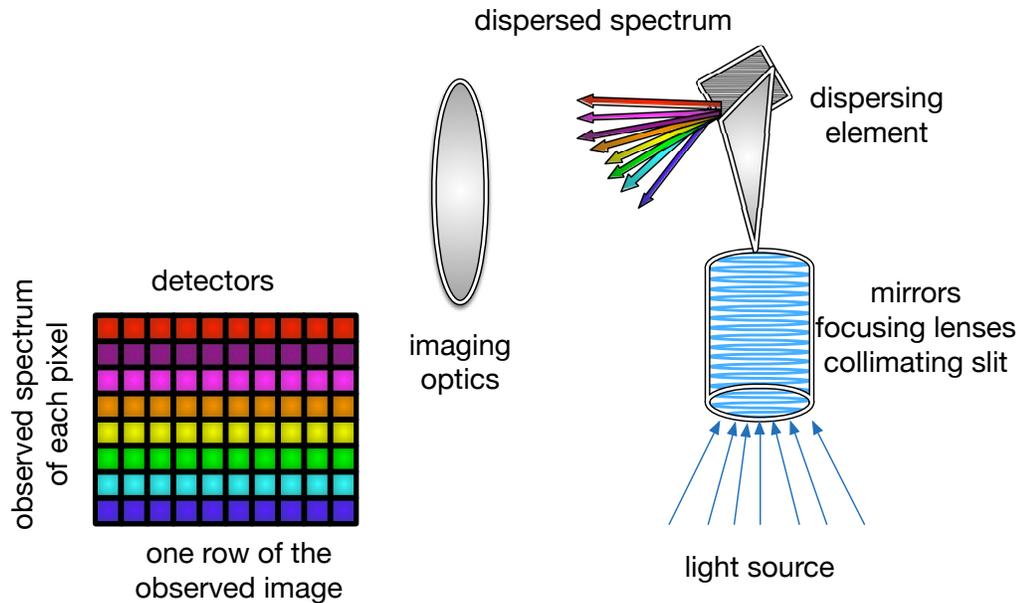
Figure 2.5: Visual depiction of the process of recording spectral data by an imaging spectrometer. A spectrometer disperses light onto a two-dimensional array of detectors. The result is spectral vectors, which contain hundreds or thousands of elements, for each pixel in the scene.

A spectrometer measures the spectral reflectance of each material in the observed scene. [128]. The basic elements of an imaging spectrometer consist of scan mirrors, focusing lenses, collimating slit, dispersing element (grating or a prism), imaging optics, and detectors. The incoming light first goes through scanning mirrors and other optics. Then, using a dispersing element such as grating or prism, the light is split into many narrow adjacent wavelength or spectral bands. The dispersed light goes through additional imaging optics and then the energy in each band is measured by separate detectors. The number of detectors used depends in the design of the instrument and can vary from hundreds to even thousands. The range of the wavelengths can vary also and can be as wide as from 400nm to 2,500nm. Figure 2.5 illustrates the concept of recording hyperspectral data using an imaging spectrometer.

A variety of platforms have been developed for capturing multispectral and hyperspectral data. Two of the most used airborne imaging spectrometer instruments are the Hyperspectral Digital Imagery Collection Experiment (HYDICE) instrument [11] and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)[1]. The spaceborne imaging spectrometer instruments in use include HYPERION[2], the Environmental Mapping and Analysis Program (ENMAP)[3], PRecursore IperSpettrale of the application mission (PRISMA)[4], Compact High Resolution Imaging Spectrometer (CHRIS)[5], and the Infrared Atmospheric Sounding Interferometer (IASI)[6]. More recently a new hyperspectral instrument, namely the hyperspectral Multi Unit Spectroscopic Explorer (MUSE)[7], has been developed and installed on the Very Large Telescope (VLT) at the Paranal Observatory in Chile. The MUSE system will be used to study the formation and evolution of distant galaxies and will be able collect hyperspectral data up to 4,000 spectral bands.

2.4   Spectral Reflectance

In general, there are three types of interactions between electromagnetic energy and a material: reflection, absorption, and/or transmission [100]. The portion that is returned to the sensor system is the reflected spectrum. The amount of reflected light varies and depends on the chemical and physical characteristics of the material as well as the number and the range of wavelengths measured by the sensor. As a result of measuring the reflected spectrum over a range of wavelengths we obtain a spectral response/reflectance pattern which is also called the spectral signature. If we plot the reflectance intensities versus the wavelength then we obtain a spectral

---

[1]http://aviris.jpl.nasa.gov
[2]http://eo1.usgs.gov/sensors/hyperion
[3]http://www.enmap.org
[4]http://www.asi.it/en/activity/earth_observation/prisma_
[5]https://earth.esa.int/web/guest/missions/esa-operational-eo-missions/proba
[6]http://smsc.cnes.fr/IASI
[7]http://muse.univ-lyon1.fr/?lang=en

reflectance curve that shows the degree to which electromagnetic energy is reflected in the different regions of the spectrum.
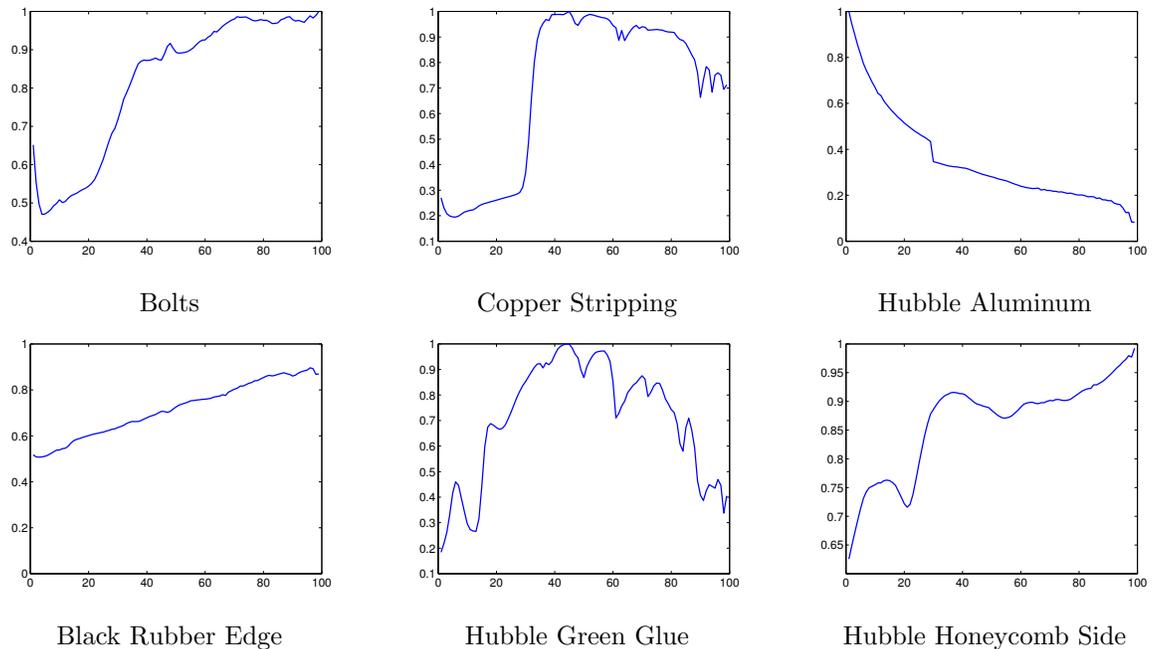
Figure 2.6: Spectral signatures of different materials commonly present in the Hubble Space Telescope.

In HSI measurements, the spectral reflectance or spectral signature is a fundamental property and it is obtained for each pixel [1]. The spectral reflectance is defined as the ratio of the reflected energy to incident energy as a function of wavelength

$$s(\lambda) = 100 \frac{E_R(\lambda)}{E_I(\lambda)} \, ,$$

where $s(\lambda)$ is the spectral reflectance at a particular wavelength, $E_R(\lambda)$ is the energy of wavelength reflected by an object, and $E_I(\lambda)$ is the energy of the wavelength incident upon an object.

The light at different wavelengths is scattered or absorbed to different degrees. Thus the reflectance intensities for most materials vary as a function of wavelength and as a consequence different materials possess fairly unique spectral properties.

Since a spectral signature consists of a set of reflectance values for each spectral channel then it can also be interpreted as a vector in a spectral space $\mathcal{R}^{N_w}$, where N_w is the number of spectral channels.

If we compare only a few spectral reflectance curves then the spectral plots provide a convenient way to visualize the difference in the spectral properties of different materials. Figure 2.6 shows typical reflectance curves for common materials present in the Hubble Space Telescope. The sharp downward deflections of the spectral reflectance curves at specific wavelengths indicate that the material selectively absorbs the incident light at those wavelengths. The bands at which the material absorbs the incident light are commonly referred to as absorption bands. The comparison of the overall shape of spectral signatures and the position and strength of absorption bands allows us to identify and discriminate different materials that are present in the observed scene. For a comprehensive discussion see [131].

The "pure" materials whose spectral reflectance at different wavelengths are measured in laboratory settings are called endmembers. Many hyperspectral signature libraries have been created in lab conditions [44]. These libraries contain spectral signatures of natural and man-made materials. [128]. Some of the most used and publicly available spectral libraries include the ASTER Spectral Library[8] and the USGS Spectral Library[9]. ASTER is made available by NASA and it contains almost 2000 spectra of minerals, rocks, soils, man-made materials, water, and snow with a spectral range of 0.4 to 14μm. The USGS Spectral Library is provided by the United States Geological Survey Spectroscopy Lab and it consists of 500 spectra of minerals and plants with a spectral range of 0.2 to 3μm. The reference spectra of endmembers' libraries can be used in the interpretation and analysis of hyperspectral images [131, 119]. The increasing number of publicly available endmember libraries has helped to increase the precision in matching endmember signatures to surface ma-

---

[8]http://speclib.jpl.nasa.gov
[9]http://speclab.cr.usgs.gov/spectral-lib.html

terials. However, in real life applications hyperspectral sensors record rather mixed spectra of different materials present in the scene. Environmental conditions, natural variations in materials, atmospheric and sensor effects contribute to the mixing of the spectra and make the task of analyzing HSIs more challenging.

2.5 Mixed Spectra

HSI sensors record raw data which represent the radiance of the observed scene [128]. The raw measured spectral radiance data by a sensor is composed not only of the surface reflectance but is also affected by the the spectrum of the source illumination, illumination geometry, shadowing, atmospheric effects, and the characteristics of the sensor system. Thus, to obtain ground-leaving spectral reflectance values, which can be used for further analysis and post processing algorithms, the sensor has to correct for these factors so that spectral reflectance can be derived from measured radiance.
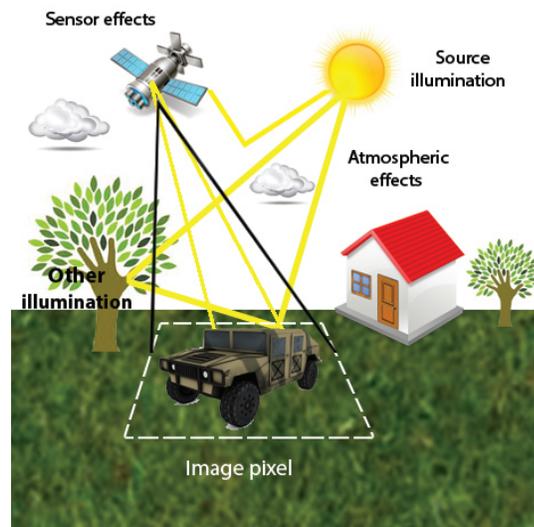
Figure 2.7: The light that reaches the hyperpsectral sensor is a combination of the light emitted by the source of illumination and reflected by the observed target, light which is scattered by aerosols and atmospheric gases and then reflected by the target, light reflected by the surrounding objects and then reflected by the target, and the light scattered by the aerosols and atmospheric gases without contact with the target. In addition, the geometry or viewing angle of the source of illumination and sensor noise have also a contaminating effect on the recorded radiance.

The pre-processing operations are generally grouped as radiometric calibration, geometric corrections, and atmospheric corrections. The aim of pre-processing techniques is to create a more faithful representation of the original scene. Radiometric calibration is an operation that corrects the data for sensor irregularities and unwanted sensor or atmospheric noise. The geometric correction operation corrects geometric distortions due to sensor-Earth geometry variations, and it converts data to coordinates, such as latitude and longitude. Moreover, as the reflected light passes through the atmosphere it gets absorbed and scattered by atmospheric gases and aerosols. The interactions of light through the atmosphere are very complex and challenging to model. For a detailed discussion on atmospheric correction see [101]. Figure 2.7 shows some possible factors that can affect the measured radiance of a pixel.

Let us ignore for a moment the atmospheric effects, the sensor noise, and the illumination geometry. Even if we assume that the sensor has corrected for all these effects the observed spectrum of a given pixel in the scene is still not pure, i.e. it does not contain only the spectrum of a single material. The reason is that the light intensity of a given pixel, which is recorded by the sensor at different wavelengths, is a mixture of the light scattered by the materials that constitute the pixel coverage area. Thus, the measured spectrum of a pixel is a result of contribution of the spectra of more than one material. This is especially the case when the size of the scene area is large, i.e. when the spatial resolution of the sensor is low.

The spatial resolution of a sensor is given by its Field of View (FOV) and it generally depends on the sensor design and the height of the sensor above the surface. The spectral resolution of a sensor is determined by the bandwidth of the spectral bands. The spectrum measured by a sensor is the average of the materials' reflectances within the sensor's FOV [128]. Hyperspectral imaging sensors acquire spectral data at many spectral bands with a narrow bandwidth and thus hyperspectral images

in general are spectrally smooth. Moreover, the reflectance values in neighboring locations and wavelengths are highly correlated.

As a result of the mixing of the spectra, it is not possible to identify the materials present in a given pixel by simply looking at the recorded spectral signature of the pixel [82, 128]. The ability to discriminate materials based on their spectral responses is fundamental to performing hyperspectral imaging data analysis and exploitation tasks. Thus, before we discuss methods to solve this problem, we first have to consider the formulation of spectral mixing models.

2.6   Hyperspectral Mixing Models

There are two commonly used hyperspectral mixing models: the linear mixing model (LMM) [147, 36] and the nonlinear mixing model(NLMM) [35, 42]. The LMM assumes that the recorded light from a scene is linearly mixed within the sensor [21, 82]. The general reason for this mixture is the insufficient spatial resolution of the sensor. Let $\boldsymbol{s} = \begin{bmatrix} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \cdots & \boldsymbol{s}_{N_w} \end{bmatrix}$ denote the measured spectral vector of a pixel. Then the LMM for a pixel can be written as

$$\boldsymbol{s} = \sum_{i=1}^{N_w} x_i \boldsymbol{m}_i + \boldsymbol{\eta},$$

where $\boldsymbol{m}_i \in \mathcal{R}^{N_w}$ is the spectral signature of the ith endmember, $x_i \geq 0$ is a scalar representing the percentage that the ith material occupies inside the pixel (also known as the fractional abundance), $\boldsymbol{\eta} \in \mathcal{R}^{N_w}$ represents additive noise and other perturbations, and N_w is the number of spectral bands or wavelengths. Since each x_i represents percentages they should be nonnegative. The other constraint that is usually imposed on the fractional abundances is that the sum of abundances for a given pixel should be 1. The goal in hyperspectral unmixing problems is to estimate the endmembers $\boldsymbol{m}_i$ present in mixed pixels and their respective fractional abundances, α_i.
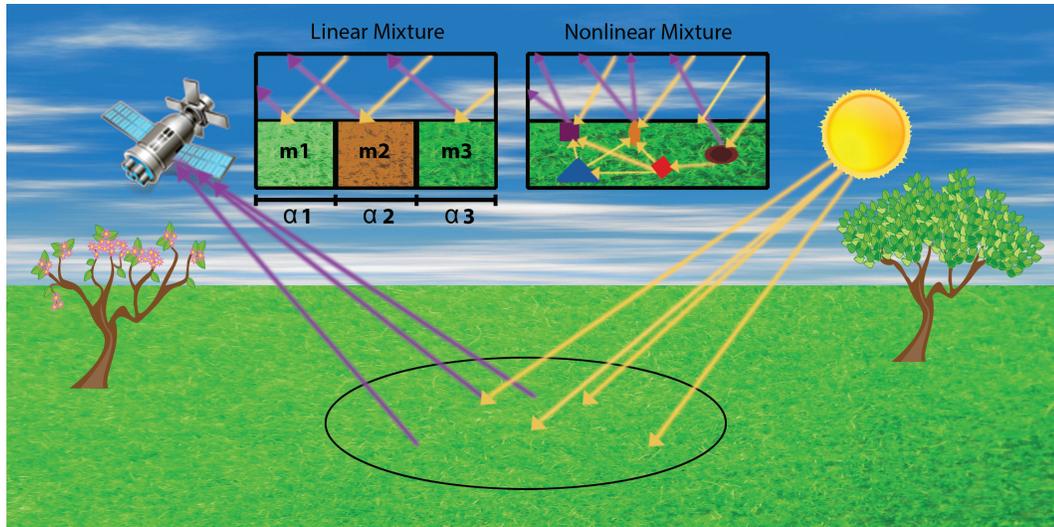
Figure 2.8: Illustration of linear and nonlinear mixture models for a pixel. The linear mixture model assumes minimal secondary reflection and minimal multiple scattering effects in a pixel's coverage area. The nonlinear mixture model assumes that the observed spectrum of a given pixel is a result of multiple scattering effects and microscopic mixing scales.

The LMM is a simple model that has been widely used in hyperspectral unmixing problems. The LMM holds true especially in the case of macroscopic mixing scales and when the incident light interacts with only one material [130, 70, 93]. Moreover, it assumes minimal secondary reflections and/or multiple scattering effects. Despite its apparent simplicity, the LMM is still an acceptable approximation model for many real world scenarios. A major practical advantage of the LMM is its ease of implementation. Optimization techniques, which usually incorporate sparsity constraints and other regularization methods, are used to solve hyperspectral unmixing problems using the LMM and potentially very large spectral libraries of endmembers.

The NLMM assumes that the source radiation is scattered multiple times before it reaches the sensor [85, 144]. Using the NLMM the measured spectral vector of a pixel can be formulated as

$$s = f(x, M) + \eta \,,$$

where $x \in \mathcal{R}^{N_w}$ is a vector containing the fractional abundances for a given pixel,

$\boldsymbol{M} \in \mathcal{R}^{N_w \times N_m}$ is a library of spectral signatures of endmembers, N_m is the number of endmembers, and f is a nonlinear function which defines the interaction between the spectral vectors of endmembers and the fractional abundances. The nonlinear function f can be estimated using kernel-based learning theory [34], physics-based and statistical models [42, 2], and multi-layer perceptron (MLP) networks [85]. The NLMM is a much more complicated model than the LMM and due to its complexity has been explored less. Figure 2.8 illustrates visually the linear and nonlinear mixtures of a pixel in an observed scene. For a detailed discussion of the LMM, the NLMM, and spectral unmixing algorithms, see e.g. [82].

Chapter 3

Estimation of Hyperspectral PSF Parameters

In [141] the authors model the PSFs at different wavelengths by Moffat functions, and they have proposed a method to estimate the PSF parameters from the HSI data of an isolated star. The first step of their method consists of estimating the star spectrum and the noise variance for each pixel in the HSI datacube by assuming a local average along the wavelength dimension. After obtaining the star spectrum and the noise variance the authors then estimate the PSF parameters. In this thesis, we use the same PSF model described in [141]. We pose the hyperspectral PSF estimation and star spectrum identification reconstruction problem in a nonlinear least squares framework, which allows us to use a joint estimation approach. We consider an iterative approach to solve the nonlinear least squares problem using all the noisy measurements of an isolated star at all wavelengths.

We set up the mathematical framework in Section 3.1. Specifically, we describe the data formation model, the PSF model used, and the general problem formulation for multispectral wavelengths. In Section 3.2 we discuss the approach we use, based on solving a separable nonlinear least squares problem, to jointly estimate the hyper-

spectral PSF parameters and to reconstruct the star spectrum. The structure of the Jacobian matrix is discussed in detail in Section 3.2.2.

3.1 Mathematical Framework

In this section we present the star image formation model, and describe the PSF models often used in astronomical imaging. In particular, we consider a circular Moffat function that can be used in cases when the blur is assumed to be spatially invariant, and an elliptical model can be used for spatially variant blurs.

3.1.1 PSF Star Image Formation Model

We follow the noisy data model used in [141], where the image formation process for an isolated star, often called a guide star in general ground-based astronomical imaging, e.g. [122], observed at a particular wavelength can be written as:

$$\boldsymbol{g}_\lambda = \boldsymbol{h}_\lambda s_\lambda + \boldsymbol{\eta}_\lambda$$

where $\boldsymbol{g}_\lambda$ is a vector representing the vectorized form of an observed, blurred, and noisy image of an isolated star corresponding to wavelength λ, $\boldsymbol{h}_\lambda$ is a vector representing the vectorized form of an exact original image of the isolated star corresponding to wavelength λ, and s_λ is a scalar representing the unknown intensity of the star spectrum at wavelength λ. In this model we assume that the sky background has been subtracted from the observed isolated star images. Note that $\boldsymbol{h}_\lambda$ also represents the hyperspectral PSF, which is, in general, unknown. However, it is reasonable to assume that a parametric model of $\boldsymbol{h}_\lambda$ is known, where the parameters defining $\boldsymbol{h}_\lambda$ are wavelength dependent. That is, by assuming a parametrized formula for the PSF,

the image formation model becomes

$$\boldsymbol{g}_\lambda = \boldsymbol{h}(\boldsymbol{\phi}_\lambda)s_\lambda + \boldsymbol{\eta}_\lambda$$

where $\boldsymbol{\phi}_\lambda$ is a vector of unknown parameters corresponding to wavelength λ. The aim, then, is given measured data $\boldsymbol{g}_{\lambda_i}$ at known wavelengths $\lambda_1, \lambda_2, \ldots, \lambda_{N_w}$, jointly compute estimates of the unknowns $\boldsymbol{\phi}_{\lambda_i}$ and s_{λ_i}. Our approach will be to do this by solving a nonlinear least squares problem; however, before we can describe this approach, we need to first discuss parametric models for the PSFs $\boldsymbol{h}(\boldsymbol{\phi}_\lambda)$.

3.1.2   Circular Moffat

In [127] the authors have shown that a circular Moffat function can be used to approximate the PSF at any spatial and spectral position of a Multi Unit Spectroscopic Explorer (MUSE) datacube. In particular, the circular variant of the Moffat function has been used to approximate the PSF when MUSE is operated without adaptive optics (AO) corrections.

The circular Moffat function is defined by a positive scale factor α and a shape parameter β. In this case, $\boldsymbol{\phi} = \begin{bmatrix} \alpha & \beta \end{bmatrix}^T$, and the PSF has the form:

$$\boldsymbol{h}(\boldsymbol{\phi})_{i_x,j_y} = \boldsymbol{h}(\alpha, \beta)_{i_x,j_y} = \left(1 + \frac{i_x^2 + j_y^2}{\alpha^2}\right)^{-\beta} \tag{3.1}$$

The flux of the Moffat function is $\iint \boldsymbol{h}_{i_x,j_y}(\alpha, \beta)\, di_x\, dj_y = \frac{\pi\alpha^2}{\beta-1}$. A scaling factor corresponding to the inverse of the flux of the Moffat function is introduced so that the PSF entries sum to 1. The PSF normalization imposes boundary conditions on the shape parameter, specifically $1 < \beta < \infty$.

It has been shown in [139] that the analytical approximation in (3.1) provides the best fit to the PSF predicted from atmospheric turbulence theory when $\beta \approx 4.765$. In
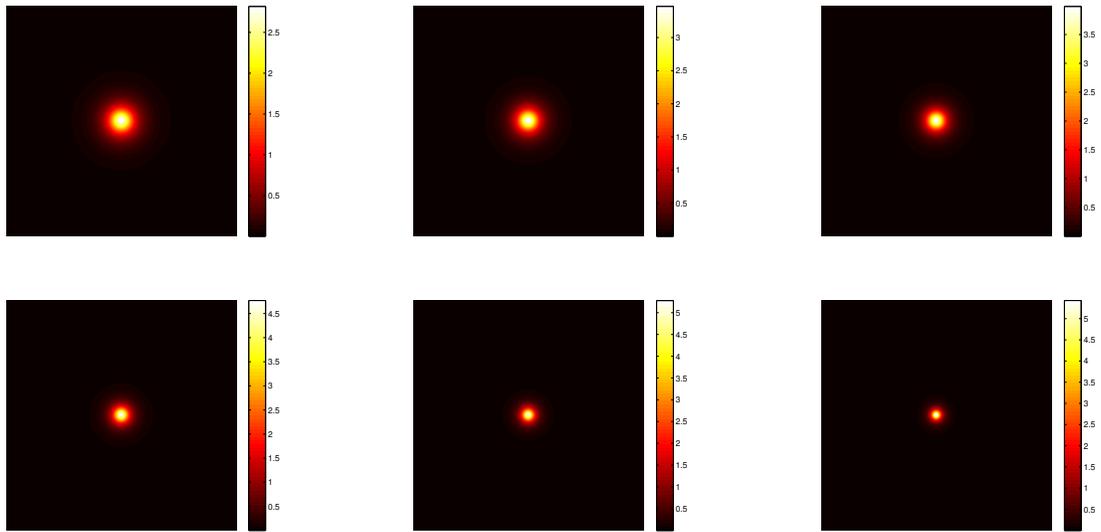
Figure 3.1: Noise-free images of an isolated star modeled by the circular Moffat PSF corresponding to wavelengths $\lambda = 450, 550, 650, 750, 850, 950$nm (top left to bottom right, respectively).
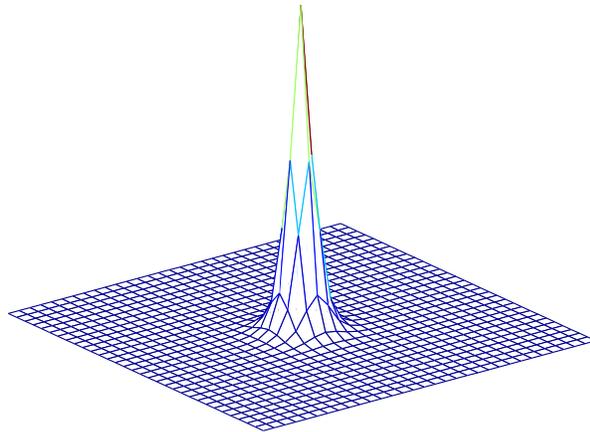
Figure 3.2: A Moffat PSF as given in equation (4.5) with $\beta = 3.31$, $\alpha_0 = 2.51$, $\alpha_1 = -7.2 \cdot 10^{-4}$, and $\lambda = 465$nm.

hyperspectral imaging the PSF is system dependent and it varies with wavelength. In [141] the authors have shown that the variation of the PSF with respect to λ can be modeled by fixing the shape parameter β as a constant, $\beta(\lambda) = \beta_0$, and by varying the scale parameter α linearly with respect to the wavelength, $\alpha(\lambda) = \alpha_0 + \alpha_1\lambda$. Using this model the parameter vector becomes $\boldsymbol{\phi} = \begin{bmatrix} \alpha_0 & \alpha_1 & \beta \end{bmatrix}^T$ and the normalized wavelength varying PSF takes the form:

$$\boldsymbol{h}(\boldsymbol{\phi})_{i_x,j_y,\lambda} = \boldsymbol{h}(\alpha_0, \alpha_1, \beta)_{i_x,j_y,\lambda} = \frac{\beta - 1}{\pi(\alpha_0 + \alpha_1\lambda)}\left(1 + \frac{i_x^2 + j_y^2}{(\alpha_0 + \alpha_1\lambda)^2}\right)^{-\beta} \quad (3.2)$$

Modeling the hyperspectral PSF at a particular wavelength by a circular Moffat function provides a simple model involving only 3 parameters. In Figure 3.1 example hyperspectral PSFs corresponding to different wavelengths are shown. A mesh plot of the Moffat PSF corresponding to $\lambda = 465$nm and defined by the parameters $\boldsymbol{\phi} = \begin{bmatrix} \alpha_0 & \alpha_1 & \beta \end{bmatrix}^T = \begin{bmatrix} 2.51 & -7.2 \cdot 10^{-4} & 3.31 \end{bmatrix}^T$ is shown in Figure 3.2.

3.1.3 Elliptical Moffat

Although the circular Moffat function may be a good model for certain spatially invariant blurs, it cannot be used in the more challenging situation when the blur is spatially variant. For spatially variant blurs, we need to use an elliptical Moffat function

$$\boldsymbol{h}(\boldsymbol{\phi})_{i_x,j_y} = \boldsymbol{h}(\alpha, \beta, \gamma, \Theta)_{i_x,j_y} = \left[1 + \frac{1}{\alpha^2}\left(i_r^2 + \frac{j_r^2}{\gamma^2}\right)\right]^{-\beta} \quad (3.3)$$

where γ is the ellipticity parameter, Θ is the rotation angle, and

$$\begin{bmatrix} i_r \\ j_r \end{bmatrix} = \begin{bmatrix} \cos(\Theta) & \sin(\Theta) \\ -\sin(\Theta) & \cos(\Theta) \end{bmatrix} \begin{bmatrix} i_x \\ j_y \end{bmatrix}.$$

The property $\boldsymbol{h}(\alpha, \beta, \gamma, \Theta)_{i_x,j_y} = \boldsymbol{h}(\alpha, \beta, \frac{1}{\gamma}, \Theta - \frac{\pi}{2})_{i_x,j_y}$ imposes boundary condi-
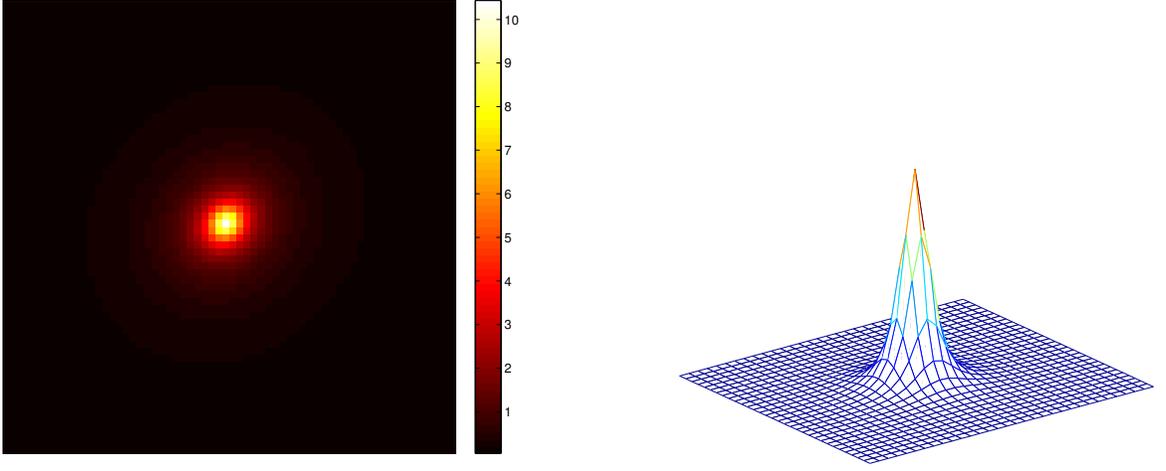
Figure 3.3: Image, and corresponding mesh plot, of an isolated star modeled by the elliptical Moffat PSF with wavelength $\lambda = 465$nm and parameters: $\alpha_0 = 3.75, \alpha_1 = -2.99 \cdot 10^{-3}, \alpha_2 = -4.31 \cdot 10^{-3}, \alpha_3 = 1.98 \cdot 10^{-6}, \beta = 1.74, \gamma_0 = 6.86 \cdot 10^{-4}, \gamma_1 = 2.17 \cdot 10^{-6}$.

tions on the ellipticity and rotation angle parameters, specifically: $1 \leq \gamma < \infty$ and $0 \leq \Theta < \frac{\pi}{2}$. In [127] the authors have proposed a model for the variation of the elliptical Moffat PSF with respect to λ and the polar coordinates (ρ, θ) in the field of view. We use the same model for the variation of the hyperspectral PSF parameters.

The ellipticity parameter, γ, is modeled with a bilinear variation, $\gamma(\lambda, \rho) = 1 + (\gamma_0 + \gamma_1 \lambda)\rho$. This implies that the hyperspectral PSFs will be more elliptical towards the edges of the field of view and also more elliptical for red than for blue light. The shape parameter, β, is kept as a constant while the variations of the scale parameter α are modeled as a linear function of ρ and a quadratic function of λ, specifically: $\alpha(\lambda, \rho) = \alpha_0 + \alpha_1 \rho + \alpha_2 \lambda + \alpha_3 \lambda^2$.

The orientation parameter, Θ, is modeled as a decreasing function of θ in the field of view, specifically $\Theta = \frac{\pi}{2} - \theta$. Figure 3.3 shows an example of an elliptical Moffat PSF corresponding to wavelength $\lambda = 465$nm. Elliptical Moffat PSFs corresponding to different locations in the field of view are shown in Figure 3.4.

Modeling the hyperspectral PSF at a particular wavelength by an elliptical Moffat
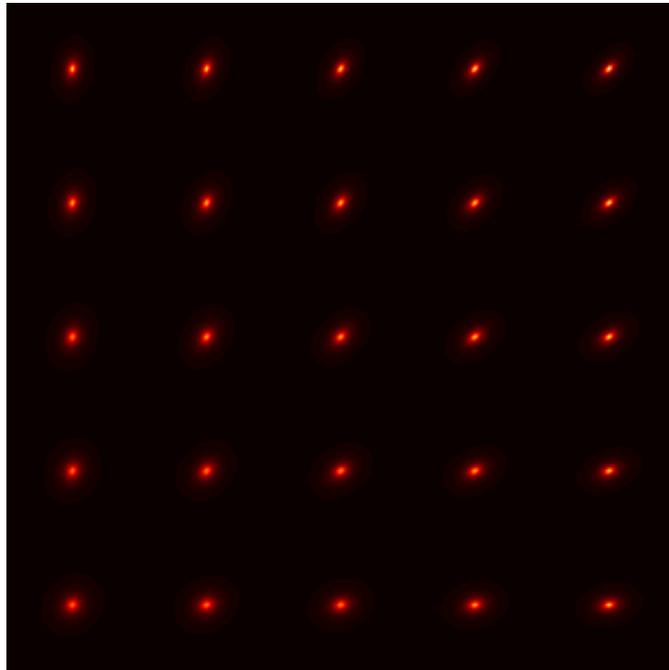
Figure 3.4: The variations of the elliptical Moffat PSF (λ = 465nm) in the first quadrant of the field of view.

function provides a model involving 7 parameters. In particular the parameter vector becomes $\phi = \begin{bmatrix} \alpha_0 & \alpha_1 & \alpha_2 & \alpha_3 & \beta & \gamma_0 & \gamma_1 \end{bmatrix}$. In the elliptical Moffat model the PSF parameters vary with respect to the location of the star in the field of view and with respect to the wavelength.

3.2 Optimization Problem

As described in the previous section, the hyperspectral imaging problem consists of obtaining multiple images of the same object over a large number of wavelengths. Suppose that each observed star image contains $n \times n$ pixels, which are stacked as vectors of length $N_p = n^2$. Then the set of observed isolated star images at N_w wavelengths can be written as:

$$\boldsymbol{g}_{\lambda_i} = \boldsymbol{h}_{\lambda_i} s_{\lambda_i} + \boldsymbol{\eta}_{\lambda_i}, \quad i = 1, 2, \cdots, N_w$$

where $\boldsymbol{g}_{\lambda_i} \in \mathcal{R}^{N_p}, \boldsymbol{h}_{\lambda_i} \in \mathcal{R}^{N_p}, \boldsymbol{\eta}_{\lambda_i} \in \mathcal{R}^{N_p}$, and s_{λ_i} is a scalar representing the reflectance of the star at a particular wavelength λ_i.

Since there is a one-to-one correspondence between λ_i and the index i, without loss of generality, we will use the notation:

$$\boldsymbol{b}_i = \boldsymbol{g}_{\lambda_i}, \quad \boldsymbol{h}_{\lambda_i} = \boldsymbol{h}_i, \quad s_{\lambda_i} = s_i, \quad \boldsymbol{\eta}_{\lambda_i} = \boldsymbol{\eta}_i.$$

By stacking all observations, we obtain the overall image formation model as:

$$\boldsymbol{b} = \boldsymbol{H}(\boldsymbol{\phi})\boldsymbol{s} + \boldsymbol{\eta}$$

where

$$\boldsymbol{s} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_{N_w} \end{bmatrix}$$

and in the case of a circular Moffat PSF,

$$\boldsymbol{b} = \begin{bmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \\ \vdots \\ \boldsymbol{b}_{N_w} \end{bmatrix}, \quad \boldsymbol{H}(\boldsymbol{\phi}) = \begin{bmatrix} \boldsymbol{h}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{h}_2 & & \vdots \\ \vdots & & \ddots & \\ \boldsymbol{0} & \cdots & & \boldsymbol{h}_{N_w} \end{bmatrix}, \quad \boldsymbol{\phi} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \beta \end{bmatrix}.$$

In the case of elliptical Moffat PSF, with N_w wavelengths and N_o orientations (i.e.,

N_o polar coordinates (ρ_ℓ, θ_ℓ), $\ell = 1, 2, \ldots, N_o$) in the field of view, we have

$$
\boldsymbol{b} = \begin{bmatrix} \boldsymbol{b}_1^{(1)} \\ \vdots \\ \boldsymbol{b}_1^{(N_o)} \\ \boldsymbol{b}_2^{(1)} \\ \vdots \\ \boldsymbol{b}_2^{(N_o)} \\ \vdots \\ \boldsymbol{b}_{N_w}^{(1)} \\ \vdots \\ \boldsymbol{b}_{N_w}^{(N_o)} \end{bmatrix}, \quad
\boldsymbol{H}(\boldsymbol{\phi}) = \begin{bmatrix} \boldsymbol{h}_1^{(1)} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \cdots & \vdots \\ \boldsymbol{h}_1^{(N_o)} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{h}_2^{(1)} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \cdots & \vdots \\ \boldsymbol{0} & \boldsymbol{h}_2^{(N_o)} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & & \ddots & \vdots \\ \boldsymbol{0} & \cdots & & \boldsymbol{h}_{N_w}^{(1)} \\ \vdots & \ddots & & \vdots \\ \boldsymbol{0} & \cdots & & \boldsymbol{h}_{N_w}^{(N_o)} \end{bmatrix}, \quad
\boldsymbol{\phi} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta \\ \gamma_0 \\ \gamma_1 \end{bmatrix}.
$$

In [141] the authors have proposed an estimation scheme consisting of two steps. In the first step the star spectrum, $\boldsymbol{s}$, and the noise, $\boldsymbol{\eta}$, are estimated from the hyperspectral data. To estimate the star spectrum, the authors compute a local average (spectral binning) of the HSI data cube using L wavelengths around a particular wavelength. It is assumed that the PSF is constant for such wavelengths. The authors use the data before subtraction of the sky spectrum as a rough estimator of the noise variance and then compute a local average of this estimator along L wavelengths. The estimated quantities $\boldsymbol{s}$ and $\boldsymbol{\eta}$ are then replaced in a quadratic criterion and the PSF parameters are approximated by using a maximum likelihood estimator.

Here, we formulate the PSF parameter estimation and star spectrum reconstruction problem in a nonlinear least squares framework:

$$
\min_{\boldsymbol{\phi}, \boldsymbol{s}} \left(f(\boldsymbol{\phi}, \boldsymbol{s}) = \|\boldsymbol{b} - \boldsymbol{H}(\boldsymbol{\phi})\boldsymbol{s}\|_2^2 \right)
$$

This approach allows us to jointly estimate the PSF parameters and the star spectrum without prior estimation of the noise variance. We do not assume the PSF to be constant in any wavelength range. Moreover the tunning of the number L of wavelengths used for spectral binning is avoided. We show that using this approach the PSF parameters and star spectrum can be estimated with low relative errors even in the presence of high noise levels.

Variable projection [57, 80, 124] is applied to the objective function, $f(\boldsymbol{\phi}, \boldsymbol{s})$, to obtain a nonlinear reduced cost functional. The variable projection method exploits the separability of the nonlinear least squares problem by eliminating the linear term $\boldsymbol{s}$ and optimizing only over the nonlinear term $\boldsymbol{\phi}$.

3.2.1 Variable Projection

We would like to find the PSF parameters $\boldsymbol{\phi}$ and an approximation of the true star spectrum $\boldsymbol{s}$ to minimize the function

$$f(\boldsymbol{\phi}, \boldsymbol{s}) = \|\boldsymbol{b} - \boldsymbol{H}(\boldsymbol{\phi})\boldsymbol{s}\|_2^2$$

It can be observed that f depends nonlinearly on $\boldsymbol{\phi}$ and linearly on $\boldsymbol{s}$. We apply the variable projection method to eliminate the linear variable $\boldsymbol{s}$. That is, first consider the linear least squares problem

$$\min_{\boldsymbol{s}} f(\boldsymbol{\phi}, \boldsymbol{s}) = \min_{\boldsymbol{s}} \|\boldsymbol{b} - \boldsymbol{H}(\boldsymbol{\phi})\boldsymbol{s}\|_2^2,$$

whose minimum is attained at

$$\hat{\boldsymbol{s}} = \boldsymbol{H}(\boldsymbol{\phi})^\dagger \boldsymbol{b}$$

where $\boldsymbol{H}(\boldsymbol{\phi})^\dagger$ is the pseudoinverse of $\boldsymbol{H}(\boldsymbol{\phi})$. Using this mathematical representation for $\boldsymbol{s}$, we obtain a reduced cost functional that depends only on $\boldsymbol{\phi}$:

$$\tilde{f}(\boldsymbol{\phi}) = \|\boldsymbol{b} - \boldsymbol{H}(\boldsymbol{\phi})\boldsymbol{H}(\boldsymbol{\phi})^\dagger \boldsymbol{b}\|_2^2 = \|(\boldsymbol{I} - \boldsymbol{H}(\boldsymbol{\phi})\boldsymbol{H}(\boldsymbol{\phi})^\dagger)\boldsymbol{b}\|_2^2 \tag{3.4}$$

For notational convenience, we drop "$(\boldsymbol{\phi})$" in the following equations. Note that,

$$\boldsymbol{H}^T \boldsymbol{H} = \begin{bmatrix} \boldsymbol{h}_1^T & & & \\ & \boldsymbol{h}_2^T & & \\ & & \ddots & \\ & & & \boldsymbol{h}_N^T \end{bmatrix} \begin{bmatrix} \boldsymbol{h}_1 & & & \\ & \boldsymbol{h}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{h}_N \end{bmatrix} = \begin{bmatrix} \|\boldsymbol{h}_1\|_2^2 & & & \\ & \|\boldsymbol{h}_2\|_2^2 & & \\ & & \ddots & \\ & & & \|\boldsymbol{h}_N\|_2^2 \end{bmatrix}.$$

Thus,

$$\boldsymbol{H}\boldsymbol{H}^\dagger = \boldsymbol{H}\left(\boldsymbol{H}^T \boldsymbol{H}\right)^{-1} \boldsymbol{H}^T = \begin{bmatrix} \frac{\boldsymbol{h}_1 \boldsymbol{h}_1^T}{\|\boldsymbol{h}_1\|_2^2} & & & \\ & \frac{\boldsymbol{h}_2 \boldsymbol{h}_2}{\|\boldsymbol{h}_2^T\|_2^2} & & \\ & & \ddots & \\ & & & \frac{\boldsymbol{h}_N \boldsymbol{h}_N^T}{\|\boldsymbol{h}_N\|_2^2} \end{bmatrix}$$

and

$$\boldsymbol{P} = \boldsymbol{I} - \boldsymbol{H}\boldsymbol{H}^\dagger = \begin{bmatrix} \boldsymbol{I} - \frac{\boldsymbol{h}_1 \boldsymbol{h}_1^T}{\|\boldsymbol{h}_1\|_2^2} & & & \\ & \boldsymbol{I} - \frac{\boldsymbol{h}_2 \boldsymbol{h}_2}{\|\boldsymbol{h}_2\|_2^2} & & \\ & & \ddots & \\ & & & \boldsymbol{I} - \frac{\boldsymbol{h}_N \boldsymbol{h}_N^T}{\|\boldsymbol{h}_N\|_2^2} \end{bmatrix},$$

where $\boldsymbol{P}$ is the projector onto the orthogonal complement of the range of $\boldsymbol{H}$. With this notation the minimization problem in (3.4) can be reposed as a nonlinear least squares problem:

$$\min_{\boldsymbol{\phi}} \left(\tilde{f}(\boldsymbol{\phi}) = \|\boldsymbol{P}(\boldsymbol{\phi})\boldsymbol{b}\|_2^2\right).$$

We use the Gauss-Newton algorithm [81, 109, 111] to solve this problem.

Algorithm: Gauss-Newton

given: $\boldsymbol{b}$

choose: initial guess for $\boldsymbol{\phi}$

        step length parameter τ

`for` $i = 1, 2, \cdots ,$ `until stop`

        `solve`

$$\boldsymbol{J}(\boldsymbol{\phi})^T \boldsymbol{J}(\boldsymbol{\phi}) \boldsymbol{d} = -\boldsymbol{J}(\boldsymbol{\phi})^T \boldsymbol{r}$$

        `where`

$$\boldsymbol{r} = \boldsymbol{b} - \boldsymbol{H}(\boldsymbol{\phi})\boldsymbol{s} = \boldsymbol{P}(\boldsymbol{\phi})\boldsymbol{b}$$

$$\boldsymbol{J} = \nabla \left(\boldsymbol{P}(\boldsymbol{\phi})\boldsymbol{b} \right) = \nabla \boldsymbol{P}(\boldsymbol{\phi})\boldsymbol{b}$$

        Set $\boldsymbol{\phi} = \boldsymbol{\phi} + \tau \boldsymbol{d}$, where τ is chosen using line search.

`end for`

Set $\boldsymbol{s} = \boldsymbol{H}(\boldsymbol{\phi})^{\dagger} \boldsymbol{b}$.

The behavior of the Gauss-Newton method depends on the conditioning of the Jacobian matrix, e.g. [109].

3.2.2   Jacobian Matrix

In this section we consider the structure and conditioning of the Jacobian matrix. For separable nonlinear least squares problems, Golub and Pereyra [57] show that it is convenient to calculate individual columns of the Jacobian, using derivatives of $\boldsymbol{H}(\boldsymbol{\phi})$. Specifically (we use similar notation as O'Leary and Rust [113]), for each variable ϕ_k, define derivative matrices $\boldsymbol{D}_k$ to have (i, j) entries

$$[\boldsymbol{D}_k]_{i,j} = \left[\frac{\partial \boldsymbol{H}_j}{\partial \phi_k} \right]_i ,$$

where $\boldsymbol{H}_j$ is the jth column of $\boldsymbol{H}(\boldsymbol{\phi})$. With this definition, the Jacobian can be written as

$$\boldsymbol{J} = -(\boldsymbol{C} + \boldsymbol{F}) \tag{3.5}$$

where the kth column of $\boldsymbol{C}$ is

$$\boldsymbol{c}_k = \boldsymbol{P}\boldsymbol{D}_k\boldsymbol{H}^\dagger\boldsymbol{b} = \boldsymbol{P}\boldsymbol{D}_k\boldsymbol{s}, \tag{3.6}$$

and the kth column of $\boldsymbol{F}$ is

$$\boldsymbol{f}_k = \left(\boldsymbol{P}\boldsymbol{D}_k\boldsymbol{H}^\dagger\right)^T \boldsymbol{b} = \left(\boldsymbol{H}^\dagger\right)^T \boldsymbol{D}_k^T\boldsymbol{P}^T\boldsymbol{b} = \left(\boldsymbol{H}^\dagger\right)^T \boldsymbol{D}_k^T\boldsymbol{r}. \tag{3.7}$$

The elliptical Moffat model is the more important case and thus in the rest of this section we consider only the elliptical Moffat PSF. In this case, the ith entry of the PSF corresponding to wavelength λ_j and ℓth orientation (ρ_ℓ, θ_ℓ) is

$$\left[\boldsymbol{h}_j^{(\ell)}\right]_i = \left[1 + \frac{\left(i_r^2 + \frac{j_r^2}{(1+(\gamma_0+\gamma_1\lambda_j)\rho_\ell)^2}\right)}{(\alpha_0 + \alpha_1\rho_\ell + \alpha_2\lambda_j + \alpha_3\lambda_j^2)^2}\right]^{-\beta}.$$

The derivatives with respect to α_0, α_1, α_2, α_3, β, γ_0, and γ_1 are:

$$\frac{\partial\left[\boldsymbol{h}_j^{(\ell)}\right]_i}{\partial\alpha_0} = 2\beta\frac{\left(i_r^2 + \frac{j_r^2}{(1+(\gamma_0+\gamma_1\lambda_j)\rho_\ell)^2}\right)}{(\alpha_0 + \alpha_1\rho_\ell + \alpha_2\lambda_j + \alpha_3\lambda_j^2)^3}\left(\left[\boldsymbol{h}_j^{(\ell)}\right]_i\right)^{-(\beta+1)}$$

$$\frac{\partial\left[\boldsymbol{h}_j^{(\ell)}\right]_i}{\partial\alpha_1} = \rho_l 2\beta\frac{\left(i_r^2 + \frac{j_r^2}{(1+(\gamma_0+\gamma_1\lambda)\rho_l)^2}\right)}{(\alpha_0 + \alpha_1\rho_l + \alpha_2\lambda_j + \alpha_3\lambda_j^2)^3}\left(\left[\boldsymbol{h}_j^{(\ell)}\right]_i\right)^{-(\beta+1)}$$

$$\frac{\partial\left[\boldsymbol{h}_j^{(\ell)}\right]_i}{\partial\alpha_2} = \lambda_j 2\beta\frac{\left(i_r^2 + \frac{j_r^2}{(1+(\gamma_0+\gamma_1\lambda)\rho_l)^2}\right)}{(\alpha_0 + \alpha_1\rho_l + \alpha_2\lambda_j + \alpha_3\lambda_j^2)^3}\left(\left[\boldsymbol{h}_j^{(\ell)}\right]_i\right)^{-(\beta+1)}$$

$$\frac{\partial \left[\boldsymbol{h}_j^{(\ell)} \right]_i}{\partial \alpha_3} = \lambda_j^2 2\beta \frac{\left(i_r^2 + \frac{j_r^2}{(1+(\gamma_0+\gamma_1\lambda)\rho_l)^2} \right)}{(\alpha_0 + \alpha_1\rho_l + \alpha_2\lambda_j + \alpha_3\lambda_j^2)^3} \left(\left[\boldsymbol{h}_j^{(\ell)} \right]_i \right)^{-(\beta+1)}$$

$$\frac{\partial \left[\boldsymbol{h}_j^{(\ell)} \right]_i}{\partial \beta} = \left(\left[\boldsymbol{h}_j^{(\ell)} \right]_i \right)^{-\beta} \ln \left(\left(\left[\boldsymbol{h}_j^{(\ell)} \right]_i \right)^{-1} \right)$$

$$\frac{\partial \left[\boldsymbol{h}_j^{(\ell)} \right]_i}{\partial \gamma_0} = 2\rho_\ell\beta \frac{\frac{j_r^2}{(1+(\gamma_0+\gamma_1\lambda_j)\rho\ell)^3}}{(\alpha_0 + \alpha_1\rho_\ell + \alpha_2\lambda_j + \alpha_3\lambda_j^2)^2} \left(\left[\boldsymbol{h}_j^{(\ell)} \right]_i \right)^{-(\beta+1)}$$

$$\frac{\partial \left[\boldsymbol{h}_j^{(\ell)} \right]_i}{\partial \gamma_1} = \lambda_j 2\rho_l\beta \frac{\frac{j_r^2}{(1+(\gamma_0+\gamma_1\lambda)\rho_l)^3}}{(\alpha_0 + \alpha_1\rho_l + \alpha_2\lambda_j + \alpha_3\lambda_j^2)^2} \left(\left[\boldsymbol{h}_j^{(\ell)} \right]_i \right)^{-(\beta+1)}$$

By identifying variables ϕ_k as

$$\boldsymbol{\phi}^T = \left[\begin{array}{ccccccc} \phi_1 & \phi_2 & \phi_3 & \phi_4 & \phi_5 & \phi_6 & \phi_7 \end{array} \right] = \left[\begin{array}{ccccccc} \alpha_0 & \alpha_1 & \alpha_2 & \alpha_3 & \beta & \gamma_0 & \gamma_1 \end{array} \right],$$

the above derivatives define the entries of matrices $\boldsymbol{D}_1$, $\boldsymbol{D}_2$, $\boldsymbol{D}_3$, $\boldsymbol{D}_4$, $\boldsymbol{D}_5$, $\boldsymbol{D}_6$, and $\boldsymbol{D}_7$. Observe that the derivatives satisfy the relations

$$\frac{\partial \left[\boldsymbol{h}_j^{(\ell)} \right]_i}{\partial \alpha_1} = \rho_\ell \frac{\partial \left[\boldsymbol{h}_j^{(\ell)} \right]_i}{\partial \alpha_0},$$

$$\frac{\partial \left[\boldsymbol{h}_j^{(\ell)} \right]_i}{\partial \alpha_2} = \lambda_j \frac{\partial \left[\boldsymbol{h}_j^{(\ell)} \right]_i}{\partial \alpha_0},$$

$$\frac{\partial \left[\boldsymbol{h}_j^{(\ell)} \right]_i}{\partial \alpha_3} = \lambda_j^2 \frac{\partial \left[\boldsymbol{h}_j^{(\ell)} \right]_i}{\partial \alpha_0},$$

$$\frac{\partial \left[\boldsymbol{h}_j^{(\ell)} \right]_i}{\partial \gamma_1} = \lambda_j \frac{\partial \left[\boldsymbol{h}_j^{(\ell)} \right]_i}{\partial \gamma_0}.$$

Now,

$$\boldsymbol{D}_1 = \begin{bmatrix} \dfrac{\partial\left[\boldsymbol{h}_1^{(1)}\right]}{\partial\alpha_0} & & & \\[2em] \dfrac{\partial\left[\boldsymbol{h}_1^{(2)}\right]}{\partial\alpha_0} & & & \\[1em] \vdots & & & \\[1em] \dfrac{\partial\left[\boldsymbol{h}_1^{(N_o)}\right]}{\partial\alpha_0} & & & \\[2em] & \dfrac{\partial\left[\boldsymbol{h}_2^{(1)}\right]}{\partial\alpha_0} & & \\[2em] & \dfrac{\partial\left[\boldsymbol{h}_2^{(2)}\right]}{\partial\alpha_0} & & \\[1em] & \vdots & & \\[1em] & \dfrac{\partial\left[\boldsymbol{h}_2^{(N_o)}\right]}{\partial\alpha_0} & & \\[2em] & & \ddots & \\[2em] & & & \dfrac{\partial\left[\boldsymbol{h}_{N_w}^{(1)}\right]}{\partial\alpha_0} \\[2em] & & & \dfrac{\partial\left[\boldsymbol{h}_{N_w}^{(2)}\right]}{\partial\alpha_0} \\[1em] & & & \vdots \\[1em] & & & \dfrac{\partial\left[\boldsymbol{h}_{N_w}^{(N_o)}\right]}{\partial\alpha_0} \end{bmatrix}$$

From the derivative relations it follows that

$$
\boldsymbol{D}_2 =
\begin{bmatrix}
\dfrac{\partial\left[\boldsymbol{h}_1^{(1)}\right]}{\partial\alpha_1} & & & \\[2ex]
\dfrac{\partial\left[\boldsymbol{h}_1^{(2)}\right]}{\partial\alpha_1} & & & \\[2ex]
\vdots & & & \\[2ex]
\dfrac{\partial\left[\boldsymbol{h}_1^{(N_o)}\right]}{\partial\alpha_1} & & & \\[2ex]
& \dfrac{\partial\left[\boldsymbol{h}_2^{(1)}\right]}{\partial\alpha_1} & & \\[2ex]
& \dfrac{\partial\left[\boldsymbol{h}_2^{(2)}\right]}{\partial\alpha_1} & & \\[2ex]
& \vdots & & \\[2ex]
& \dfrac{\partial\left[\boldsymbol{h}_2^{(N_o)}\right]}{\partial\alpha_1} & & \\[2ex]
& & \ddots & \\[2ex]
& & & \dfrac{\partial\left[\boldsymbol{h}_{N_w}^{(1)}\right]}{\partial\alpha_1} \\[2ex]
& & & \dfrac{\partial\left[\boldsymbol{h}_{N_w}^{(2)}\right]}{\partial\alpha_1} \\[2ex]
& & & \dfrac{\partial\left[\boldsymbol{h}_{N_w}^{(N_o)}\right]}{\partial\alpha_1}
\end{bmatrix}
=
\begin{bmatrix}
\rho_1\dfrac{\partial\left[\boldsymbol{h}_1^{(1)}\right]}{\partial\alpha_1} & & & \\[2ex]
\rho_2\dfrac{\partial\left[\boldsymbol{h}_1^{(2)}\right]}{\partial\alpha_1} & & & \\[2ex]
\vdots & & & \\[2ex]
\rho_{N_o}\dfrac{\partial\left[\boldsymbol{h}_1^{(N_o)}\right]}{\partial\alpha_1} & & & \\[2ex]
& \rho_1\dfrac{\partial\left[\boldsymbol{h}_2^{(1)}\right]}{\partial\alpha_1} & & \\[2ex]
& \rho_2\dfrac{\partial\left[\boldsymbol{h}_2^{(2)}\right]}{\partial\alpha_1} & & \\[2ex]
& \vdots & & \\[2ex]
& \rho_{N_o}\dfrac{\partial\left[\boldsymbol{h}_2^{(N_o)}\right]}{\partial\alpha_1} & & \\[2ex]
& & \ddots & \\[2ex]
& & & \rho_1\dfrac{\partial\left[\boldsymbol{h}_{N_w}^{(1)}\right]}{\partial\alpha_1} \\[2ex]
& & & \rho_2\dfrac{\partial\left[\boldsymbol{h}_{N_w}^{(2)}\right]}{\partial\alpha_1} \\[2ex]
& & & \rho_{N_o}\dfrac{\partial\left[\boldsymbol{h}_{N_w}^{(N_o)}\right]}{\partial\alpha_1}
\end{bmatrix}
$$

Hence, the following relations hold

$$
\begin{aligned}
\boldsymbol{D}_2 &= \boldsymbol{R}\boldsymbol{D}_1 \\[1ex]
\boldsymbol{D}_3 &= \boldsymbol{D}_1\boldsymbol{\Lambda} \\[1ex]
\boldsymbol{D}_4 &= \boldsymbol{D}_1\boldsymbol{\Lambda}^2 \\[1ex]
\boldsymbol{D}_7 &= \boldsymbol{D}_6\boldsymbol{\Lambda}
\end{aligned}
$$

where

$$
\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{N_w}) \quad \text{and} \quad \boldsymbol{R} = \boldsymbol{I} \otimes \mathrm{diag}(\rho_1, \rho_2, \ldots, \rho_{N_o}).
$$

These relations for $\boldsymbol{D}_k$ can be used, according to equations (3.5-3.7), to compute the Jacobian. It is difficult to determine a precise analytical result regarding the conditioning of the Jacobian. However, it is interesting to notice that since $\boldsymbol{\Lambda}$ is a diagonal matrix, we can write

$$\boldsymbol{\Lambda}\boldsymbol{s} = \mathrm{diag}(\boldsymbol{s})\boldsymbol{\lambda}, \quad \boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 & \cdots & \lambda_{N_w} \end{bmatrix}^T.$$

We can similarly write $\boldsymbol{s} = \mathrm{diag}(\boldsymbol{s})\boldsymbol{1}$, where $\boldsymbol{1}$ is a vector of all ones. Using these observations, we can write a subset of the columns of $\boldsymbol{C}$ as

$$\begin{bmatrix} \boldsymbol{c}_1 & \boldsymbol{c}_3 & \boldsymbol{c}_4 \end{bmatrix} = \boldsymbol{P}\boldsymbol{D}_1\mathrm{diag}(\boldsymbol{s}) \begin{bmatrix} \boldsymbol{1} & \boldsymbol{\lambda} & \boldsymbol{\lambda}^2 \end{bmatrix},$$

where $\boldsymbol{\lambda}^2 = \begin{bmatrix} \lambda_1^2 & \cdots & \lambda_{N_w}^2 \end{bmatrix}^T$. A similar observation can be made for the same subset of columns of $\boldsymbol{F}$, namely

$$\begin{bmatrix} \boldsymbol{f}_1 & \boldsymbol{f}_3 & \boldsymbol{f}_4 \end{bmatrix} = (\boldsymbol{H}^\dagger)^T\mathrm{diag}(\boldsymbol{D}_1^T\boldsymbol{r}) \begin{bmatrix} \boldsymbol{1} & \boldsymbol{\lambda} & \boldsymbol{\lambda}^2 \end{bmatrix}.$$

The Vandermonde structure of these corresponding columns of $\boldsymbol{J}$ indicate the potential of ill-conditioning; our numerical experiments (see Chapter 5) verify that this is the case. This structure also indicates that the conditioning of the Jacobian might be improved by rescaling the wavelengths, but this was not the case for our numerical experiments. We emphasize that the nonlinear dependency of the parameters is difficult to analyze, but it is important that we implement some safeguards against potential parameter dependency and ill-conditioning.

3.2.3 Subset Selection

It has been shown in [77] that subset selection can be applied for accurate and efficient parameter estimation. In particular the authors show that subset selection is more

numerically stable for use in nonlinear least squares problems than the approach of the truncation of the Jacobian by singular value decomposition.

Let $\boldsymbol{J}(\boldsymbol{\phi}) \in R^{M \times p}$, where M is the number of wavelengths multiplied by the size of the vector representing a star image, and p is the number of parameters used to define the hyperspectral PSF. Let $1 \leq k \leq p$ be a sampling parameter. We use subset selection to choose k linearly independent columns of the Jacobian matrix $\boldsymbol{J}(\boldsymbol{\phi})$. Subset selection applied to the Jacobian matrix $\boldsymbol{J}(\boldsymbol{\phi})$ produces a permutation matrix Π such that

$$\boldsymbol{J}(\boldsymbol{\phi})\Pi = \begin{bmatrix} \boldsymbol{J}_1 & \boldsymbol{J}_2 \end{bmatrix},$$

where $\boldsymbol{J}_1$ contains k columns. The goal is to bring the k linearly independent columns of $\boldsymbol{J}(\boldsymbol{\phi})$ to the front. This results in identifying the "best" k parameters of the original parameter set $\boldsymbol{\phi}$ while keeping the rest of the $p - k$ parameters unchanged. Thus we solve the new Jacobian system

$$\boldsymbol{J}_1^T \boldsymbol{J}_1 \boldsymbol{d} = -\boldsymbol{J}_1^T \boldsymbol{r}.$$

This guarantees that the nonlinear least squares problem we are trying to solve has a full rank Jacobian. We use the Chan-Foster high-rank-revealing QR algorithm [31, 52] for performing subset selection.

Chapter 4

Deblurring and Sparse Unmixing of Hyperspectral Images using Multiple PSFs

Hyperspectral unmixing involves the computation of the fractional contribution of elementary spectra, called endmembers. By assuming the measured spectrum of each mixed pixel in an HSI is a linear combination of spectral signatures of endmembers, the underlying image model can be formulated as a linear mixture of endmembers with nonnegative and sparse coefficients

$$G = XM + N,$$

where $M \in \mathcal{R}^{N_m \times N_w}$ represents a spectral library containing N_m spectral signatures of endmembers with N_w spectral bands or wavelengths, $G \in \mathcal{R}^{N_p \times N_w}$ is the observed data matrix (each row contains the observed spectrum of a given pixel), and $X \in \mathcal{R}^{N_p \times N_m}$ contains the fractional abundances of endmembers (each column contains the fractional abundances of a given endmember). Here, we assume $X \in \mathcal{R}^{N_p \times N_m}$ is a sparse matrix that contains the nonnegative fractional abundances of endmembers

(each column contains the fractional abundances of a given endmember), and $\boldsymbol{N} \in \mathcal{R}^{N_p \times N_w}$ is the matrix collecting the errors or noise affecting the measurements at each spectral band or wavelength, e.g. [45, 147]. If we assume that the data at each wavelength has been degraded by a blurring operator, $\boldsymbol{H}$, then the problem takes the form

$$\boldsymbol{G} = \boldsymbol{H}\boldsymbol{X}\boldsymbol{M} + \boldsymbol{N}.$$

Given a spectral library of the endmembers, $\boldsymbol{M}$, and assuming that we have computed *a priori* the parameters defining the blurring operator $\boldsymbol{H}$, the goal becomes to compute the nonnegative and sparse matrix of fractional abundances, $\boldsymbol{X}$. This is known as the hyperspectral unmixing and deblurring problem.

The work presented in this chapter is concerned with deblurring and spectrally analyzing ground-based astronomical images of space objects. A numerical approach is provided for deblurring and sparse unmixing of ground-based hyperspectral images of objects taken through the atmosphere at multiple wavelengths with narrow spectral channels. Here, the problem is quite challenging since the PSF depends on the imaging system as well as the seeing conditions and is generally wavelength dependent. We assume the hyperpsectral PSF has been estimated using numerical methods presented in the previous chapter. We then derive a numerical scheme for deblurring and unmixing of the HSI datacube using the estimated PSFs. Our approach is based on a preconditioned alternating direction method of multipliers.

In Section 4.1 we review a numerical approach for deblurring and sparse unmixing of HSI datacubes for the special case of a homogeneous PSF across the wavelengths, based on work by Zhao et al. [147]. Our approach for hyperpsectral unmixing and deblurring with multiple PSFs is presented in Section 4.2.

4.1 Numerical Scheme for the Single PSF Case

In this section we describe and expand upon the numerical scheme used in [147] for solving the hyperspectral image deblurring and unmixing problem by using the Alternating Direction Method of Multipliers (ADMM) in the single PSF case. Here, it is assumed that the blurring operator $\boldsymbol{H}$ is defined by a single PSF and each column of $\boldsymbol{XM}$ is blurred by the same $\boldsymbol{H}$. The authors in [147] have presented a total variation (TV) regularization method for solving the deblurring and sparse hyperspectral unmixing of the form

$$\min_{\boldsymbol{X} \geq 0} \frac{1}{2}||\boldsymbol{HXM} - \boldsymbol{G}||_F^2 + \mu_1||\boldsymbol{X}||_1 + \mu_2 TV(\boldsymbol{X})$$

where $\boldsymbol{H} \in \mathcal{R}^{N_p \times N_p}$ is is a blurring matrix constructed from a single Gaussian function assuming periodic boundary conditions, and μ_1, μ_2 are two regularization terms used to control the importance of the sparsity and the total variation terms. They present numerical schemes for both isotropic and anisotropic total variation. For the isotropic TV the above problem can be rewritten as

$$\min \frac{1}{2}||\boldsymbol{HXM} - \boldsymbol{G}||_F^2 + \mu_1||\boldsymbol{V}||_1 + \mu_2 \sum_{i=1}^{N_p} \sum_{j=1}^{N_m} ||\boldsymbol{W}_{ij}||_2 \tag{4.1}$$

subject to

$$\boldsymbol{D}_h\boldsymbol{X} = \boldsymbol{W}^{(1)}, \boldsymbol{D}_v\boldsymbol{X} = \boldsymbol{W}^{(2)}, \boldsymbol{V} = \boldsymbol{X}, \boldsymbol{V} \in \mathcal{K} = \left\{\boldsymbol{V} \in R^{N_p \times N_m}, \boldsymbol{V} \geq 0\right\},$$

where

$$\boldsymbol{W}_{i,j} = \left[\boldsymbol{W}_{i,j}^{(1)}, \boldsymbol{W}_{i,j}^{(2)}\right] \in R^{1 \times 2},$$

$$\boldsymbol{W}_{i,j}^{(1)} = \boldsymbol{D}_{i,h}\boldsymbol{x}_j, \quad \boldsymbol{W}_{i,j}^{(2)} = \boldsymbol{D}_{i,v}\boldsymbol{x}_j, \quad 1 \leq i \leq N_p, \quad 1 \leq j \leq N_m.$$

Here, the matrices $\boldsymbol{D}_h$ and $\boldsymbol{D}_v$ represent the first order difference matrices in the vertical and horizontal direction. The authors in [147] solve the above minimization problem using an alternating direction method. The problem in (4.1) can be decoupled by letting

$$f_1(\boldsymbol{X}) = \frac{1}{2}||\boldsymbol{HXM} - \boldsymbol{G}||_F^2$$

and

$$f_2(\boldsymbol{Z}) = \mathcal{X}_\mathcal{K}(\boldsymbol{V}) + \mu_2 \sum_{i=1}^{N_p} \sum_{j=1}^{N_m} ||\boldsymbol{W}_{ij}||_2 + \mu_1||\boldsymbol{V}||_1$$

where

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{W}^{(1)} & \boldsymbol{W}^{(2)} & \boldsymbol{V} \end{bmatrix}, \quad \mathcal{X}_\mathcal{K} = \begin{cases} \boldsymbol{0} & \text{if } \boldsymbol{V} \in \mathcal{K} \\ \infty & \text{otherwise} \end{cases}$$

The constraints are expressed as

$$\boldsymbol{BX} + \boldsymbol{CZ} = \begin{bmatrix} \boldsymbol{D}_h \\ \boldsymbol{D}_v \\ \boldsymbol{I}_{N_p \times N_p} \end{bmatrix} \boldsymbol{X} - \boldsymbol{I}_{3N_p \times 3N_p} \begin{bmatrix} \boldsymbol{W}^{(1)} \\ \boldsymbol{W}^{(2)} \\ \boldsymbol{V} \end{bmatrix} = \boldsymbol{0}_{3N_p \times N_m}$$

Furthermore, by attaching the Lagrange multiplier to the linear constraints the augmented Lagrangian function of (4.1) is written as

$$L(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\Lambda}) = f_1(\boldsymbol{X}) + f_2(\boldsymbol{Z}) + <\boldsymbol{\Lambda}, \boldsymbol{BX} + \boldsymbol{CZ}> + \frac{\beta}{2}||\boldsymbol{BX} + \boldsymbol{CZ}||_F^2 \,,$$

where

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}^{(1)} \\ \boldsymbol{\Lambda}^{(2)} \\ \boldsymbol{\Lambda}^{(3)} \end{bmatrix} \in R^{3N_p \times N_m} \,,$$

$\beta > 0$ is the penalty parameter for violating the linear constraints, and $< \cdot, \cdot >$ is the sum of the entries of the Hadamard product. With this formulation the hyperspectral

unmixing and deblurring problem is solved using an alternating direction method consisting of solving 3 subproblems at each iteration k:

$$\begin{cases} \text{Step 1:} & \boldsymbol{X}_{k+1} \leftarrow \arg\min L(\boldsymbol{X}, \boldsymbol{Z}_k, \boldsymbol{\Gamma}_k) \\ \text{Step 2:} & \boldsymbol{Z}_{k+1} \leftarrow \arg\min L(\boldsymbol{X}_{k+1}, \boldsymbol{Z}, \boldsymbol{\Gamma}_k) \\ \text{Step 3:} & \boldsymbol{\Gamma}_{k+1} \leftarrow \boldsymbol{\Gamma}_k + \beta(\boldsymbol{B}\boldsymbol{X}_{k+1} + \boldsymbol{C}\boldsymbol{Z}_{k+1}) \end{cases}$$

The $\boldsymbol{X}$-subproblem or Step 1 consists of solving

$$\boldsymbol{X}_{k+1} \leftarrow \arg\min \left\{ \frac{1}{2}||\boldsymbol{H}\boldsymbol{X}\boldsymbol{M} - \boldsymbol{G}||_F^2 + <\boldsymbol{\Lambda}, \boldsymbol{B}\boldsymbol{X} + \boldsymbol{C}\boldsymbol{Z}> + \frac{\beta}{2}||\boldsymbol{B}\boldsymbol{X} + \boldsymbol{C}\boldsymbol{Z}||_F^2 \right\} \tag{4.2}$$

The above subproblem is the solution of the classical Sylvester matrix equation

$$\boldsymbol{H}^\top\boldsymbol{H}\boldsymbol{X}\boldsymbol{M}\boldsymbol{M}^\top + \beta\boldsymbol{B}^\top\boldsymbol{B}\boldsymbol{X} = \boldsymbol{H}^\top\boldsymbol{G}\boldsymbol{M}^\top - \beta\boldsymbol{B}^\top\boldsymbol{C}\boldsymbol{Z}_k\boldsymbol{B}^\top\boldsymbol{\Gamma}_k. \tag{4.3}$$

Similar alternating minimization schemes have been used for solving the hyperspectral unmixing problem in [75] and [90]. However, the key step of the alternating minimization scheme presented in [147] consists in transforming the matrix equation (4.3) to a linear system which has a closed-form solution. In particular the authors in [147] reformulate the Sylvester matrix equation in (4.3) as

$$(\boldsymbol{M}\boldsymbol{M}^\top \otimes \boldsymbol{H}^\top\boldsymbol{H} + \beta\boldsymbol{I} \otimes \boldsymbol{B}^\top\boldsymbol{B})\boldsymbol{x} = \hat{\boldsymbol{g}}$$

where $\boldsymbol{x} = vec(\boldsymbol{X})$ and $\hat{\boldsymbol{g}} = vec(\boldsymbol{H}^\top\boldsymbol{G}\boldsymbol{M}^\top - \beta\boldsymbol{B}^\top\boldsymbol{C}\boldsymbol{Z}_k\boldsymbol{B}^\top\boldsymbol{\Gamma}_k)$. Let $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ be the singular value decomposition of $\boldsymbol{M}$, and let $\boldsymbol{H} = \boldsymbol{F}^*\boldsymbol{\Gamma}\boldsymbol{F}$ and $\boldsymbol{B}^\top\boldsymbol{B} = \boldsymbol{F}^*\boldsymbol{\Psi}\boldsymbol{F}$ be the Fourier decomposition of $\boldsymbol{H}$. Here, we assume spatially invariant blur with periodic boundary conditions. Now, the above linear system takes the form

$$(\boldsymbol{U} \otimes \boldsymbol{F}^*)(\boldsymbol{\Sigma}^2 \otimes \boldsymbol{\Gamma}^2 + \boldsymbol{I} \otimes \boldsymbol{\Psi}^2)(\boldsymbol{U}^\top \otimes \boldsymbol{F})\boldsymbol{x} = \hat{\boldsymbol{g}}$$

and thus a direct solution is given by

$$\boldsymbol{x} = (\boldsymbol{U}^\top \otimes \boldsymbol{F})(\boldsymbol{\Sigma}^2 \otimes \boldsymbol{\Gamma}^2 + \boldsymbol{I} \otimes \boldsymbol{\Psi}^2)^{-1}(\boldsymbol{U} \otimes \boldsymbol{F}^*)\hat{\boldsymbol{g}}$$

For a detailed description of the solution of steps 2 and 3 of the alternating minimization approach see [147].

4.2 Numerical Scheme for the Multiple PSF Case

In this section we provide the problem formulation for the general case where each column of the matrix $\boldsymbol{XM}$ is generally blurred by a different blurring operator. In particular, the deblurring and hyperspectral unmixing problem with multiple PSFs takes the form

$$\begin{bmatrix} \boldsymbol{H}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{H}_2 & \cdots & \boldsymbol{0} \\ \vdots & \boldsymbol{0} & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{H}_{N_w} \end{bmatrix} \begin{bmatrix} \boldsymbol{XMe}_1 \\ \boldsymbol{XMe}_2 \\ \vdots \\ \boldsymbol{XMe}_{N_w} \end{bmatrix} = \begin{bmatrix} \boldsymbol{g}_1 \\ \boldsymbol{g}_2 \\ \vdots \\ \boldsymbol{g}_{N_w} \end{bmatrix} \tag{4.4}$$

where $\boldsymbol{e}_i \in \mathcal{R}^{N_w}$ is the ith unit vector, $\boldsymbol{g}_i$ is the ith column of the observed matrix $\boldsymbol{G}$, and each blurring matrix $\boldsymbol{H}_i$ is constructed using different circular Moffat functions whose parameters vary with wavelength. In particular, a blurring operator $\boldsymbol{H}$ for a particular wavelength, λ, is represented by a circular Moffat function

$$\boldsymbol{H}(\alpha_0, \alpha_1, \beta, \lambda) = \frac{\beta - 1}{\pi(\alpha_0 + \alpha_1\lambda)} \left(1 + \frac{i^2 + j^2}{(\alpha_0 + \alpha_1\lambda)^2}\right)^{-\beta} \tag{4.5}$$

Moffat functions are widely used to paramaterize PSFs in astronomical imaging. The parameters α_0, α_1 and β are the Moffat function shape parameters for the associated

PSF and we assume that they have been estimated using the numerical approaches presented in Chapter 3.

Notice that problem (4.4) can be rewritten as

$$
\begin{bmatrix}
\boldsymbol{H}_1 \boldsymbol{X} \boldsymbol{M} \boldsymbol{e}_1 \\
\boldsymbol{H}_2 \boldsymbol{X} \boldsymbol{M} \boldsymbol{e}_2 \\
\vdots \\
\boldsymbol{H}_{N_w} \boldsymbol{X} \boldsymbol{M} \boldsymbol{e}_{N_w}
\end{bmatrix}
=
\begin{bmatrix}
\boldsymbol{g}_1 \\
\boldsymbol{g}_2 \\
\vdots \\
\boldsymbol{g}_{N_w}
\end{bmatrix}
$$

By utilizing Kronecker product properties the above equation can be reformulated as

$$
\begin{bmatrix}
(\boldsymbol{e}_1{}^\top \boldsymbol{M}^\top \otimes \boldsymbol{H}_1)\boldsymbol{x} \\
(\boldsymbol{e}_2{}^\top \boldsymbol{M}^\top \otimes \boldsymbol{H}_2)\boldsymbol{x} \\
\vdots \\
(\boldsymbol{e}_{N_w}{}^\top \boldsymbol{M}^\top \otimes \boldsymbol{H}_{N_w})\boldsymbol{x}
\end{bmatrix}
=
\begin{bmatrix}
\boldsymbol{g}_1 \\
\boldsymbol{g}_2 \\
\vdots \\
\boldsymbol{g}_{N_w}
\end{bmatrix}
$$

where $\boldsymbol{x} = vec(\boldsymbol{X})$. Thus, the multiple PSF hyperspectral image deblurring and unmixing problem can be formulated as

$$
\mathcal{H}\boldsymbol{x} = \boldsymbol{g}
$$

where

$$
\mathcal{H} =
\begin{bmatrix}
\boldsymbol{m}_1{}^\top \otimes \boldsymbol{H}_1 \\
\boldsymbol{m}_2{}^\top \otimes \boldsymbol{H}_2 \\
\vdots \\
\boldsymbol{m}_{N_w}{}^\top \otimes \boldsymbol{H}_{N_w}
\end{bmatrix},
\quad \boldsymbol{x} = vec(\boldsymbol{X}), \quad
\boldsymbol{g} =
\begin{bmatrix}
\boldsymbol{g}_1 \\
\boldsymbol{g}_2 \\
\vdots \\
\boldsymbol{g}_{N_w}
\end{bmatrix}
$$

Hence, the $\boldsymbol{X}$-subproblem (4.2) for the multiple PSF formulation takes the form

$$
\boldsymbol{X}_{k+1} \in \arg\min \left\{ \frac{1}{2} ||\mathcal{H}\boldsymbol{x} - \boldsymbol{g}||_F^2 + < \boldsymbol{\Lambda}, \boldsymbol{B}\boldsymbol{X} + \boldsymbol{C}\boldsymbol{Z} > + \frac{\beta}{2} ||\boldsymbol{B}\boldsymbol{X} + \boldsymbol{C}\boldsymbol{Z}||_F^2 \right\}
$$

If we set the gradient of the augmented Lagrangian for the $\boldsymbol{X}$-subproblem to 0 then we obtain

$$\mathcal{H}^\top \mathcal{H} \boldsymbol{x} + \beta \boldsymbol{B}^\top \boldsymbol{B} \boldsymbol{X} = \mathcal{H}^\top \boldsymbol{g} - \beta \boldsymbol{B}^\top \boldsymbol{C} \boldsymbol{Z}_k \boldsymbol{B}^\top \boldsymbol{\Gamma}_k$$

The above equation can be rewritten as

$$(\mathcal{H}^\top \mathcal{H} + \beta \boldsymbol{I} \otimes \boldsymbol{B}^\top \boldsymbol{B}) \boldsymbol{x} = \hat{\boldsymbol{g}}, \tag{4.6}$$

where $\hat{\boldsymbol{g}} = \text{vec}(\boldsymbol{H}_1{}^\top (\boldsymbol{g}_1 \boldsymbol{m}_1{}^\top) + \cdots + \boldsymbol{H}_{N_w}{}^\top (\boldsymbol{g}_1 \boldsymbol{m}_{N_w}{}^\top) - \beta \boldsymbol{B}^\top \boldsymbol{C} \boldsymbol{Z}_k - \boldsymbol{B}^\top \boldsymbol{\Gamma}_k)$. Notice that

$$\mathcal{H}^\top \mathcal{H} = \begin{bmatrix} \boldsymbol{m}_1 \otimes \boldsymbol{H}_1{}^\top & \boldsymbol{m}_2 \otimes \boldsymbol{H}_2{}^\top & \cdots & \boldsymbol{m}_{N_w} \otimes \boldsymbol{H}_{N_w}{}^\top \end{bmatrix} \begin{bmatrix} \boldsymbol{m}_1{}^\top \otimes \boldsymbol{H}_1 \\ \boldsymbol{m}_2{}^\top \otimes \boldsymbol{H}_2 \\ \vdots \\ \boldsymbol{m}_{N_w}{}^\top \otimes \boldsymbol{H}_{N_w} \end{bmatrix}$$

$$= \boldsymbol{m}_1 \boldsymbol{m}_1{}^\top \otimes \boldsymbol{H}_1{}^\top \boldsymbol{H}_1 + \boldsymbol{m}_2 \boldsymbol{m}_2{}^\top \otimes \boldsymbol{H}_2{}^\top \boldsymbol{H}_2 + \cdots + \boldsymbol{m}_{N_w} \boldsymbol{m}_{N_w}{}^\top \otimes \boldsymbol{H}_{N_w}{}^\top \boldsymbol{H}_{N_w}$$

Using the decompositions $\boldsymbol{M}\boldsymbol{M}^\top = \boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{U}^\top$, $\boldsymbol{H}_i{}^\top \boldsymbol{H}_i = \boldsymbol{F}^*\boldsymbol{\Gamma}_i^2\boldsymbol{F}$, and $\boldsymbol{B}^\top \boldsymbol{B} = \boldsymbol{F}^*\boldsymbol{\Psi}^2\boldsymbol{F}$ equation (4.6) takes the form

$$(\boldsymbol{m}_1 \boldsymbol{m}_1{}^\top \otimes \boldsymbol{F}^*\boldsymbol{\Gamma}_1^2\boldsymbol{F} + \boldsymbol{m}_2 \boldsymbol{m}_2{}^\top \otimes \boldsymbol{F}^*\boldsymbol{\Gamma}_2^2\boldsymbol{F} + \cdots + \boldsymbol{m}_{N_w} \boldsymbol{m}_{N_w}{}^\top \otimes \boldsymbol{F}^*\boldsymbol{\Gamma}_{N_w}^2\boldsymbol{F} + \boldsymbol{I} \otimes \boldsymbol{F}^*\boldsymbol{\Psi}^2\boldsymbol{F})\boldsymbol{x} = \hat{\boldsymbol{g}}.$$

Thus, the $\boldsymbol{X}$-subproblem to be solved for the multiple PSF case is

$$(\boldsymbol{I} \otimes \boldsymbol{F}^*)(\boldsymbol{m}_1 \boldsymbol{m}_1{}^\top \otimes \boldsymbol{\Gamma}_1^2 + \boldsymbol{m}_2 \boldsymbol{m}_2{}^\top \otimes \boldsymbol{\Gamma}_2^2 + \cdots + \boldsymbol{m}_{N_w} \boldsymbol{m}_{N_w}{}^\top \otimes \boldsymbol{\Gamma}_{N_w}^2 + \boldsymbol{I} \otimes \boldsymbol{\Psi}^2)(\boldsymbol{I} \otimes \boldsymbol{F})\boldsymbol{x} = \hat{\boldsymbol{g}}.$$

Note that the middle part of the coefficient matrix in the above linear system is not diagonal as in the single PSF case, and thus there is not an explicit solution of

the $\boldsymbol{X}$-subproblem for multiple PFSs. However, the coefficient matrix for multiple PSFs is symmetric positive definite and thus we use the conjugate gradient method to solve the $\boldsymbol{X}$-subproblem for hyperspectral unmixing and deblurring using multiple PSFs. Construction of a preconditioner is described next.

4.3   Conjugate Gradient Preconditioner

In the single PSF case there is only one PSF matrix and therefore only one matrix of eigenvalues associated with its Fourier decomposition. In the multiple PSF case the eigenvalues change from one PSF to another, according to wavelength, and thus we cannot apply the same technique of direct inversion for solving the $\boldsymbol{X}$-subproblem. However, though the PSF eigenvalues change with wavelengths, the eigenvalues of PSFs corresponding to adjacent wavelengths have similar values. Thus, we approximate each $\boldsymbol{\Gamma}_i$ with a matrix that contains the average of all the eigenvalues corresponding to different PSFs, which we denote by $\boldsymbol{\Gamma}_{avg}$. We use $\boldsymbol{\Gamma}_{avg}$ to construct a preconditioner. In this case we can rewrite the coefficient matrix for the $\boldsymbol{X}$-subproblem with multiple PSFs as

$$\boldsymbol{m}_1\boldsymbol{m}_1{}^\top \otimes \boldsymbol{F}^*\boldsymbol{\Gamma}_{avg}^2\boldsymbol{F} + \boldsymbol{m}_2\boldsymbol{m}_2{}^\top \otimes \boldsymbol{F}^*\boldsymbol{\Gamma}_{avg}^2\boldsymbol{F} + \cdots + \boldsymbol{m}_{N_w}\boldsymbol{m}_{N_w}{}^\top \otimes \boldsymbol{F}^*\boldsymbol{\Gamma}_{avg}^2\boldsymbol{F} + \boldsymbol{I} \otimes \boldsymbol{F}^*\boldsymbol{\Psi}^2\boldsymbol{F}$$

$$= (\boldsymbol{m}_1\boldsymbol{m}_1{}^\top + \boldsymbol{m}_2\boldsymbol{m}_2{}^\top + \cdots + \boldsymbol{m}_{N_w}\boldsymbol{m}_{N_w}{}^\top) \otimes \boldsymbol{F}^*\boldsymbol{\Gamma}_{avg}^2\boldsymbol{F} + \boldsymbol{I} \otimes \boldsymbol{F}^*\boldsymbol{\Psi}^2\boldsymbol{F}$$

$$= \boldsymbol{M}\boldsymbol{M}^\top \otimes \boldsymbol{F}^*\boldsymbol{\Gamma}_{avg}^2\boldsymbol{F} + \boldsymbol{I} \otimes \boldsymbol{F}^*\boldsymbol{\Psi}^2\boldsymbol{F}$$

$$= \boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{U}^\top \otimes \boldsymbol{F}^*\boldsymbol{\Gamma}_{avg}^2\boldsymbol{F} + \boldsymbol{I} \otimes \boldsymbol{F}^*\boldsymbol{\Psi}^2\boldsymbol{F}.$$

This leads to the preconditioner

$$\boldsymbol{P} = (\boldsymbol{U} \otimes \boldsymbol{F}^*)(\boldsymbol{\Sigma}^2 \otimes \boldsymbol{\Gamma}_{avg}^2 + \boldsymbol{I} \otimes \boldsymbol{\Psi}^2)(\boldsymbol{U}^\top \otimes \boldsymbol{F}).$$

We remark that the preconditioner $\boldsymbol{P}$ is computed once using the matrix of the average of eigenvalues for all the PSFs. Thus, we do not need to consider the computational costs of computing the Fourier decompositions of PSFs at each iteration. Also, $\boldsymbol{\Psi}$ and the SVD of $\boldsymbol{M}$ are precomputed once and stored. Furthermore, the middle term of $\boldsymbol{P}$ is now diagonal, the left and right terms are unitary, and thus $\boldsymbol{P}$ can be easily inverted. Hence, it is easy to solve linear systems with $\boldsymbol{P}$ and $\boldsymbol{P}^{-1}$. We use the preconditioned conjugate gradient method to solve the hyperspectral unmixing and deblurring problem for the multiple PSF case.

4.3.1 Approximation Quality of the Preconditioner

In this section, we consider the approximation quality of the preconditioner by looking at the singular values of the original coefficient matrix and the preconditioned system. Specifically, we consider the singular values of the matrix, $\mathcal{H}^{\top}\mathcal{H} + \beta\boldsymbol{I} \otimes \boldsymbol{B}^{\top}\boldsymbol{B}$, as well as the singular values of the preconditioned matrix,$(\mathcal{H}^{\top}\mathcal{H} + \beta\boldsymbol{I} \otimes \boldsymbol{B}^{\top}\boldsymbol{B})\boldsymbol{P}^{-1}$.
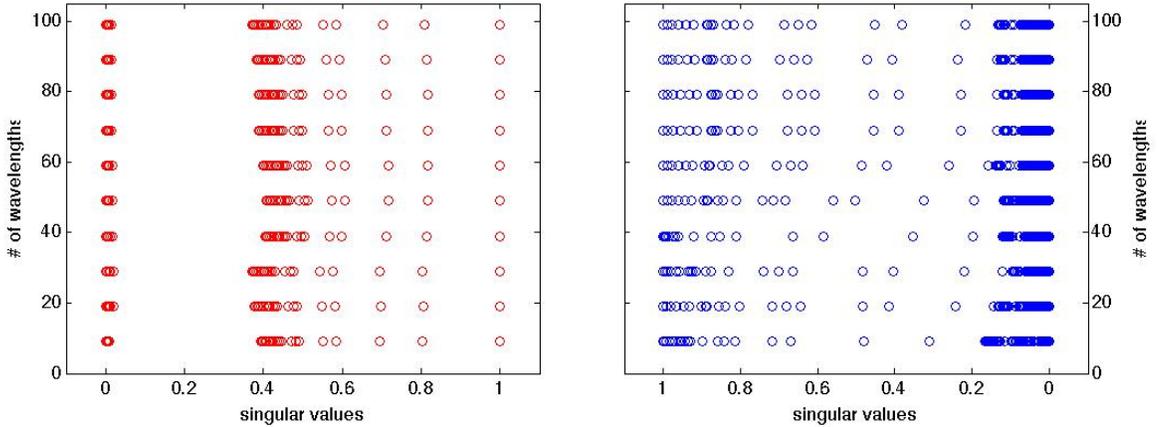
Figure 4.1: Singular values for the non-preconditoned (left) and preconditioned (right) systems with varying number of wavelengths.

Figure 4.1 shows the singular values for 10 preconditioned and non-preconditioned systems. We used PSFs of size 10×10 and vary the number of wavelengths from 9 to

99 with a 10nm step size. Note that the singular values of the preconditioned system tend to cluster more towards to 1 compared to the non-preconditioned system. We also noticed that the singular values of the preconditioned system show a tendency to move away from 0. The behavior of the singular values is similar as the number of wavelengths varies. Even though there is not a tight clustering of the singular values around 1, as one would expect from an effective preconditioner, our numerical results show that our simple preconditioner significantly reduces the number of necessary iterations for the convergence of the conjugate gradient method.

Chapter 5

Numerical Results

We now presents some results illustrating the use of effectiveness of our approaches on hyperspectral PSF parameter estimation and hyperspectral unmixing and deblurring problems. In Section 5.1 we show our results for the hyperspectal PSF parameter estimation problem. In particular, we test our methods proposed in Chapter 3 for the circular and elliptical Moffat PSF models. Section 5.1.3 shows our results for hyperspectal unmixing and deblurring using multiple PSFs.

5.1 Hyperspectral PSF Parameter Estimation Results

In this section we illustrate the use of our approach on a parameter identification and star spectrum reconstruction problem arising in hyperpsectral imaging. Different levels of Gaussian white noise are added to the original star images. An example of an isolated star observed at wavelength $\lambda = 465$nm with different levels of noise is shown in Figure 5.1.

1% noise   5% noise   10% noise
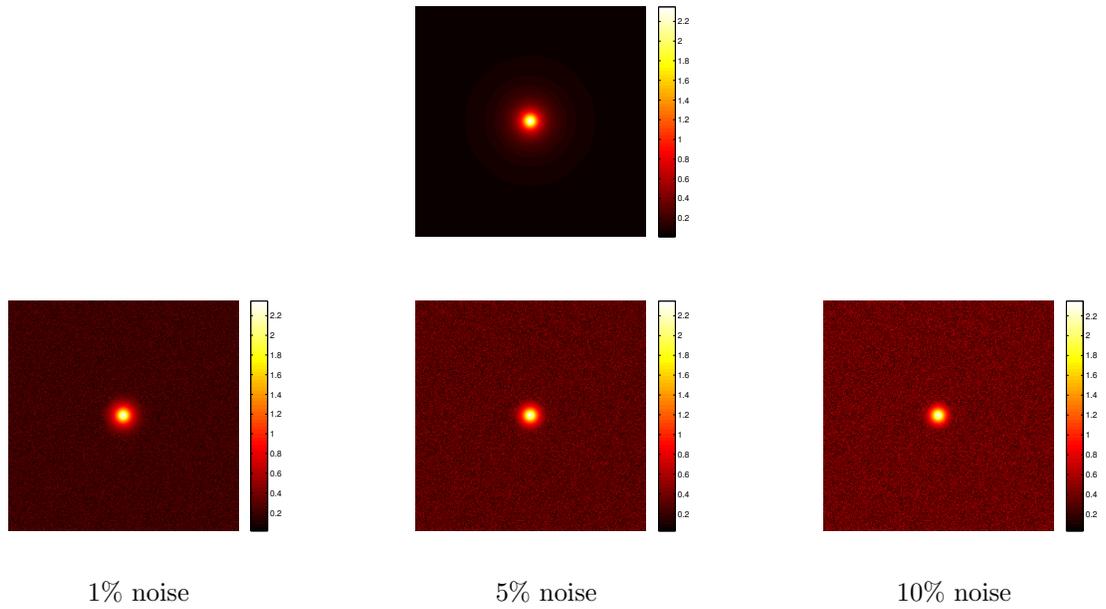
Figure 5.1: Image of an isolated star observed at wavelength $\lambda = 465$nm. The top row is the noise-free image. The bottom row shows the star image with different levels of noise.
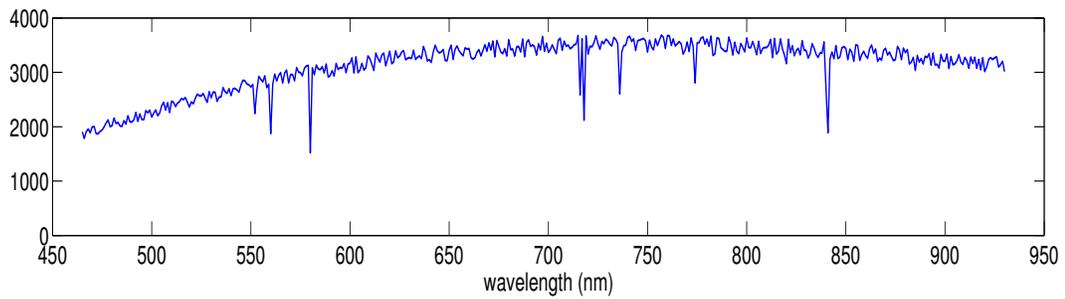
Figure 5.2: The simulated true star spectrum, $\boldsymbol{s}_{true}$, for the wavelength range 465-930nm.

Recall from Chapter 3, we are minimizing:

$$f(\boldsymbol{\phi}, \boldsymbol{s}) = \|\boldsymbol{b} - \boldsymbol{H}(\boldsymbol{\phi})\boldsymbol{s}\|_2^2$$

Our goal is to find the parameters defining the hyperspectral PSF (e.g. the top row of Figure 5.1) and to reconstruct the star spectrum (e.g. Figure 5.2), given the noisy measurements (e.g. the bottom row of Figure 5.1). The relative error at the kth iteration is defined as

$$\frac{\|\boldsymbol{\phi}_k - \boldsymbol{\phi}_{true}\|_2}{\|\boldsymbol{\phi}_{true}\|_2},$$

for the hyperspectral PSF parameters, and

$$\frac{\|\boldsymbol{s}_k - \boldsymbol{s}_{true}\|_2}{\|\boldsymbol{s}_{true}\|_2},$$

for the star spectrum.

5.1.1   Circular Moffat Results

For the numerical experiments corresponding to the circular Moffat model we use 466 simulated noisy (10% Gaussian noise) PSFs corresponding to 466 wavelengths from $\lambda = 465$nm to $\lambda = 930$nm with a step size of 1nm. For the real MUSE data up to 3700 wavelengths may be available [127]. The simulated noisy hyperspectral PSFs are of size 64×64. We use $\boldsymbol{\phi}_{true} = \begin{bmatrix} \alpha_0 & \alpha_1 & \beta \end{bmatrix}^\top = \begin{bmatrix} 2.42 & -0.001 & 2.66 \end{bmatrix}^\top$ to generate the data and $\boldsymbol{\phi}_0 = \begin{bmatrix} 4.61 & -0.0009 & 4.3 \end{bmatrix}^\top$ as an initial guess.

Figure 5.3 shows a plot of the singular values of the Jacobian matrix associated with the circular Moffat model. Note that the Jacobian matrix is ill-conditioned (with a condition number of $\approx 10^5$) and there is no clear gap between the second and third singular values. We plot the objective function (Figure 5.4), the relative errors for the PSF parameters (Figure 5.5), and the approximated star spectrum for the
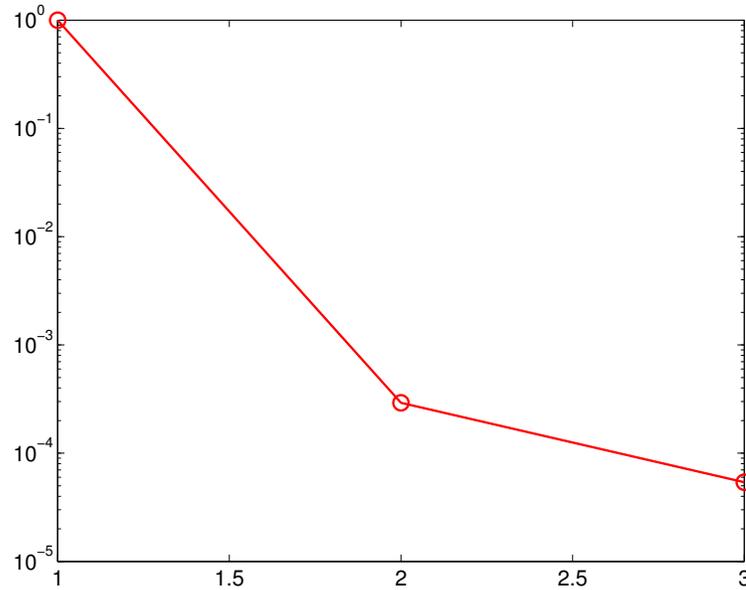
Figure 5.3: Singular values of the Jacobian matrix for the circular Moffat model.

first 100 wavelengths (Figure 5.6) for the variable projection Gauss-Newton method where all three singular values are kept. We can observe that the objective function decreases for a few iterations and then stagnates while the norm of the gradient becomes significantly smaller.

The computed PSF parameters after running the Gauss-Newton algorithm for 20 iterations are $\boldsymbol{\phi}_{GN} = \begin{bmatrix} 2.4155 & -0.001 & 2.6533 \end{bmatrix}^{\top}$ corresponding to a relative error of 0.0022. The estimated star spectrum is a very close approximation of the true star spectrum (Figure 5.6).
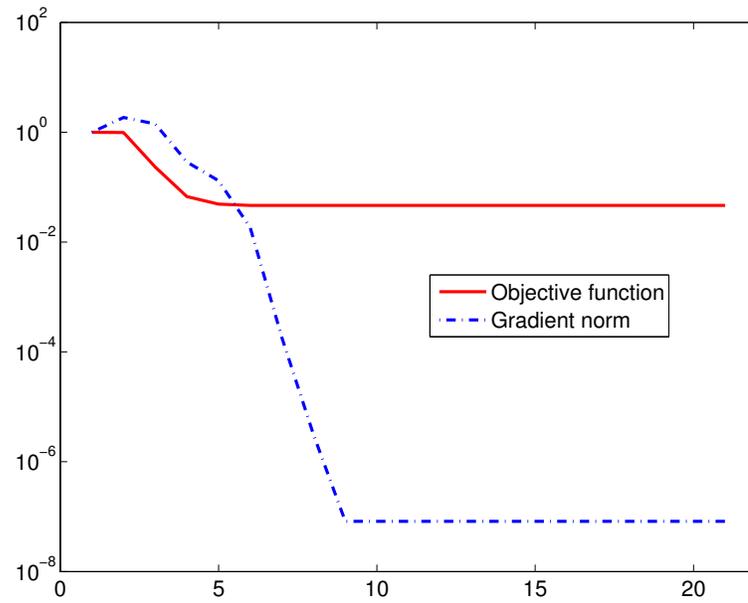
Figure 5.4: Iteration history of the objective function and the norm of gradient for the circular Moffat model.
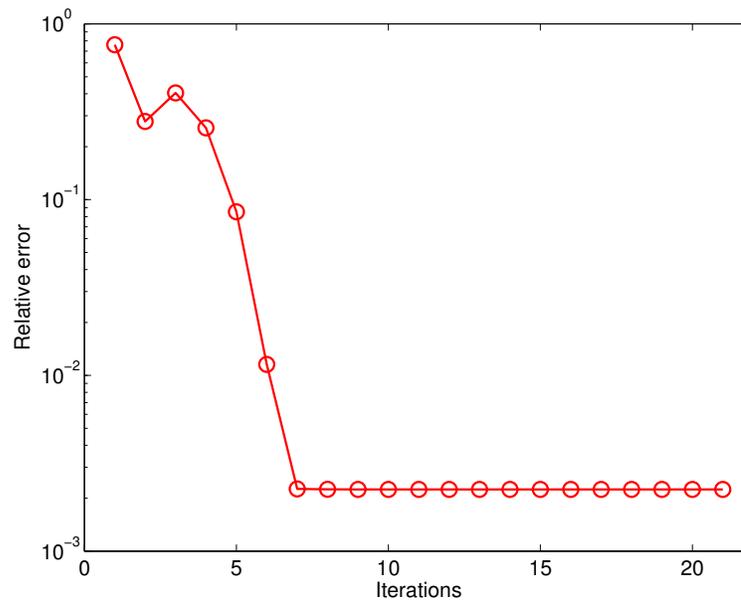
Figure 5.5: PSF parameter relative errors for the circular Moffat model.
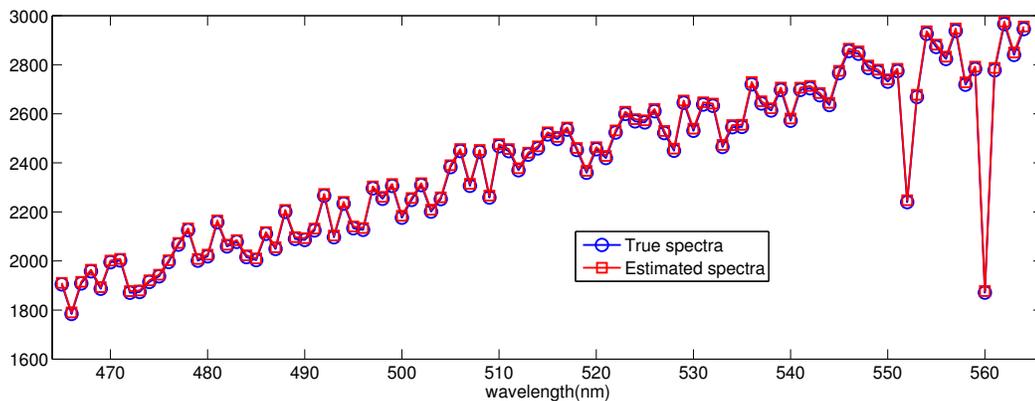
Figure 5.6: True and computed spectra for the first 100 wavelengths (from $\lambda = 465$ to $\lambda = 564$nm with a step size of 1nm between wavelengths) corresponding to the circular Moffat model.

5.1.2   Elliptical Moffat Model

For the numerical experiments with the elliptical Moffat model we generate the data using

$$\boldsymbol{\phi}_{true} = \begin{bmatrix} 3.75 & -2.99 \cdot 10^{-3} & -4.31 \cdot 10^{-3} & 1.98 \cdot 10^{-6} & 1.74 & 6.86 \cdot 10^{-4} & 2.17 \cdot 10^{-6} \end{bmatrix}^{\top}$$

as the exact parameter vector. However, to insure that all parameters have the same order of magnitude we rewrite the vector of unknown parameters as

$$\boldsymbol{\phi} = \begin{bmatrix} \alpha_0 & \alpha_1' \cdot 10^{-3} & \alpha_2' \cdot 10^{-3} & \alpha_3' \cdot 10^{-6} & \beta & \gamma_0' \cdot 10^{-4} & \gamma_1' \cdot 10^{-6} \end{bmatrix}^{\top}$$

Thus we seek to find the parameters $\boldsymbol{\phi} = \begin{bmatrix} \alpha_0 & \alpha_1' & \alpha_2' & \alpha_3' & \beta & \gamma_0' & \gamma_1' \end{bmatrix}^{\top}$ which best approximate the data. In this example the vector of true parameters becomes

$$\boldsymbol{\phi}_{true} = \begin{bmatrix} 3.75 & -2.99 & -4.31 & 1.98 & 1.74 & 6.86 & 2.17 \end{bmatrix}^{\top}$$

and we use $\boldsymbol{\phi}_0 = \begin{bmatrix} 6.41 & -1.07 & -3.15 & 3.28 & 2.42 & 13.18 & 3.81 \end{bmatrix}^{\top}$ as an initial guess.
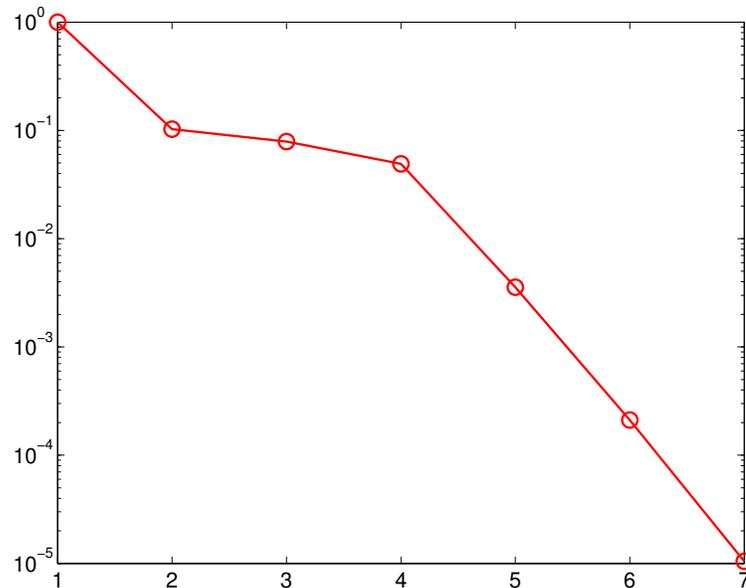
Figure 5.7: Singular values of the Jacobian matrix corresponding to the elliptical Moffat model with 20 wavelengths and 25 orientations in the field of view.

We use 20 wavelengths with 25 orientations in the first quadrant of the field of view for each wavelength.

Figure 5.7 shows a plot of the singular values of the Jacobian matrix associated with the elliptical Moffat model. Notice that the singular values decay gradually without a well-determined gap between them, and that the Jacobian matrix is ill-conditioned ($\mathrm{cond}(\boldsymbol{J}) = 10^5$). Thus the exact Gauss-Newton algorithm has the potential to compute an incorrect step direction vector when solving the system $\boldsymbol{J}^\top\boldsymbol{J}\boldsymbol{d} = -\boldsymbol{J}^\top\boldsymbol{r}$. To improve the step direction, as previously discussed, we attempt to use subset selection, truncated SVD, and Tikhonov regularization.

For the computation using subset selection we use 6 most linearly independent columns of the Jacobian matrix (note that in this example $\boldsymbol{J}$ has 7 columns). Subset selection picks α_3 to be the problematic parameter and thus we fix α_3 at its initial guess and we do not update it throughout the iterations. In the truncated SVD approach we keep 6 singular values, i.e. we truncate only the smallest singular value of the Jacobian matrix.
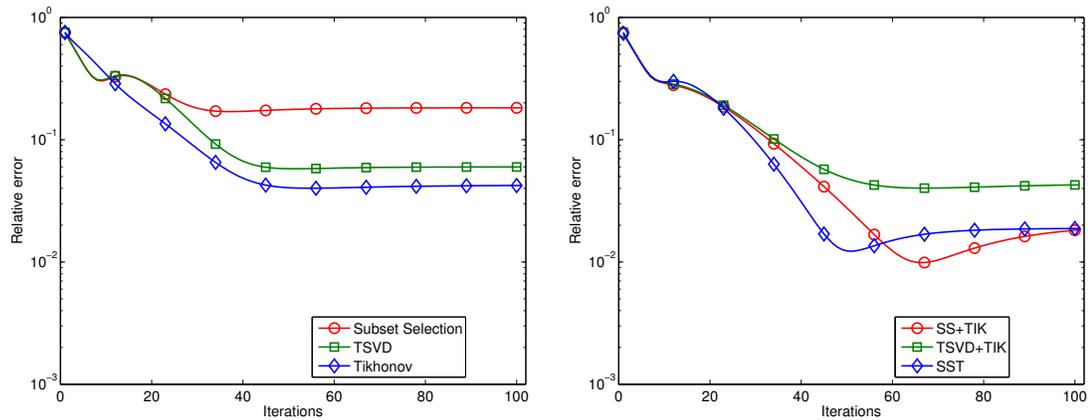
Figure 5.8: PSF parameter relative errors corresponding to the elliptical Moffat model. Left: parameter relative errors for subset selection, TSVD, and Tikhonov regularization. Right: parameter relative errors for the combined approaches of applying subset selection and Tikhonov regularization (SS+TIK), TSVD and Tikhonov regularization (TSVD+TIK), and for the approach of using subset selection to identify the problematic parameter (α_3) and assume that we know the true value of that parameter for the rest of the iterations (SST).
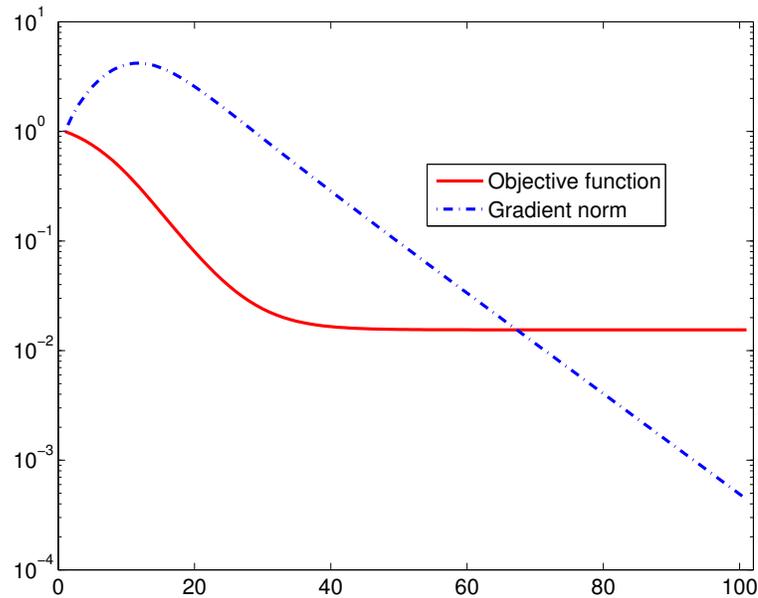
Figure 5.9: Iteration history of the objective function and the norm of gradient for the combined approach of subset selection and Tikhonov regularization.
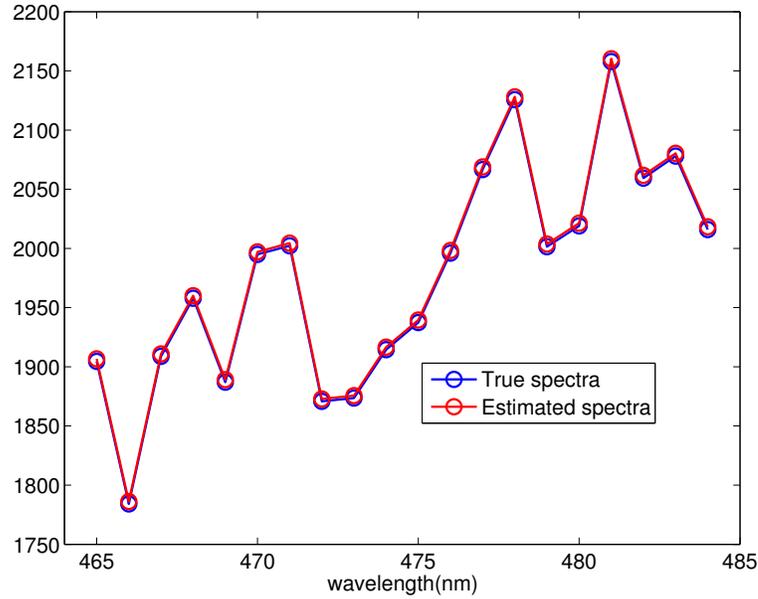
Figure 5.10: True and computed spectra for 20 wavelengths (from $\lambda = 465$nm to $\lambda = 484$nm with a step size of 1nm between wavelengths), corresponding to the elliptical Moffat model.

For the Tikhonov regularization approach we use the sixth singular value of the Jacobian matrix as the regularization parameter. We plot the relative errors

$$\frac{\|\phi_k - \phi_{true}\|_2}{\|\phi_{true}\|_2},$$

for the three approaches in the left part of Figure 5.8. One can observe that the Tikhonov regularization method achieves the lowest relative errors. With Tikhonov regularization the approximated parameters are:

$$\phi^{TIK} = \begin{bmatrix} 3.8173 & -2.9962 & -4.5704 & 2.2541 & 1.7446 & 7.0326 & 2.1308 \end{bmatrix}^\top.$$

If we assume that we know the exact value for α_3 and apply subset selection (SST in the right part of Figure 5.8) we achieve lower relative errors compared to fixing α_3 to its initial guess value. We also tried a combined approach by applying subset

selection for the first 7 iterations and then for the rest of the iterations we solve the Jacobian system using Tikhonov regularization. The basic idea is to apply subset selection for the first few iterations so that we can provide a better initial guess for Tikhonov regularization. For comparison purposes we applied the same combined approach using truncated SVD and Tikhonov regularization.

The parameter relative errors are shown in the right part of Figure 5.8. We observe that using subset selection and Tikhonov regularization (SS+TIK) accelerates the convergence and also achieves lower relative errors compared to the other approaches. After running subset selection with Tikhonov regularization for 100 iterations the approximated parameter vector is

$$\phi^{SS+TIK} = \begin{bmatrix} 3.771 & -2.9962 & -4.3749 & 2.0476 & 1.7446 & 7.0123 & 2.1351 \end{bmatrix}^\top$$

Figure 5.9 shows the convergence history for the objective function and the norm of the gradient corresponding to the SS+TIK approach. One can observe that gradient norm decreases as the iterations proceed. The true star spectrum and the approximated star spectrum using the SS+TIK approach are plotted in Figure 5.10.

5.1.3 Hyperspectral Unmixing and Deblurring Results

In this section we test the proposed numerical scheme for hyperspectral deblurring and unmixing using single and multiple PSFs. We consider a simulated hyperspectral image of the Hubble Space Telescope, which is also used for testing in [147]. Similar data was also used in [92] and [146]. The quality of the estimated fractional abundances of endmembers is evaluated by using the relative error defined by:

$$\frac{||X_{true} - X||_2}{||X_{true}||_2},$$

where X_{true} is the true fractional abundances of endmembers and X is the computed fractional abundances of endmembers by the proposed method.

The signatures cover a band of spectra from 400nm to 2500 nm. We use 99 evenly distributed sampling points, leading to a hyperspectral datacube of size $128 \times 128 \times 99$. The synthetic map of the satellite image is shown in Figure 5.1. The spectral signatures of the materials used in our experiment are shown in Figure 5.11. The hyperspectral datacube is blurred by multiple circular Moffat point spread functions (see Figure 5.12), i.e. each column of $\boldsymbol{Y}$ is blurred by a different circular Moffat function corresponding to a particular wavelength. Note that as the wavelength, λ, increases there is less blurring present in those columns compared to the blurring in the columns observed at shorter wavelengths, as expected e.g. [79]. Gaussian white noise in the level of 30dB is also added to the data cube.

In the experiments for the numerical scheme with multiple PSFs we use all the PSFs for the reconstruction of the fractional abundances. In the single PSF scheme we use an average of all the PSFs for reconstruction. That is, the columns of $\boldsymbol{Y}$ are blurred with different PSFs in both cases and we use one PSF representing the average of all PSFs for reconstruction in the single PSF numerical scheme. The plot of relative errors is shown in Figure 5.13. One can observe that the use of multiple PSFs provides lower relative reconstruction errors compared to the relative errors obtained by the single PSF case. It is a known fact that the blurring varies with different wavelengths. Figure 5.13 shows that by taking this fact into account we can achieve much lower relative errors in the reconstruction of the fractional abundances.

The following values are used for the parameters in the alternating minimization scheme: $\beta = 10^{-2}, \mu_1 = 10^{-1}, \mu_2 = 5 \cdot 10^{-4}$. The convergence of the alternating direction method is theoretically guaranteed as long as the penalty parameter β is positive, see e.g. [147]. We note that the conjugate gradient method required $1,000$ iterations for solving each $\boldsymbol{X}$-subproblem (for multiple PSFs) at the same accuracy level as the
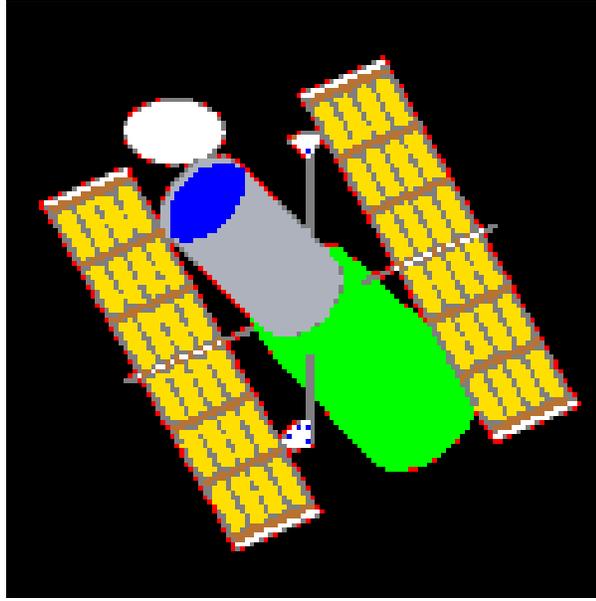
Material	Color	Endmembers (%)	Frac. Abundances (%)
Material 1	light gray	Em. 1 (100)	11
Material 2	green	Em. 2 (70), Em. 9 (30)	18
Material 3	red	Em. 3 (100)	4
Material 4	dark gray	Em. 4 (60), Em. 10 (40)	19
Material 5	brown	Em. 5 (100)	7
Material 6	gold	Em. 6 (40), Em. 11 (30), Em. 12 (30)	32
Material 7	blue	Em. 7 (100)	3
Material 8	white	Em. 8 (100)	6

Table 5.1: Top: Synthetic map representation of the hyperspectral satellite image. Bottom: Materials, corresponding colors, fractional abundances of constituent end-members, and fractional abundances of the materials used for the Hubble satellite simulation.
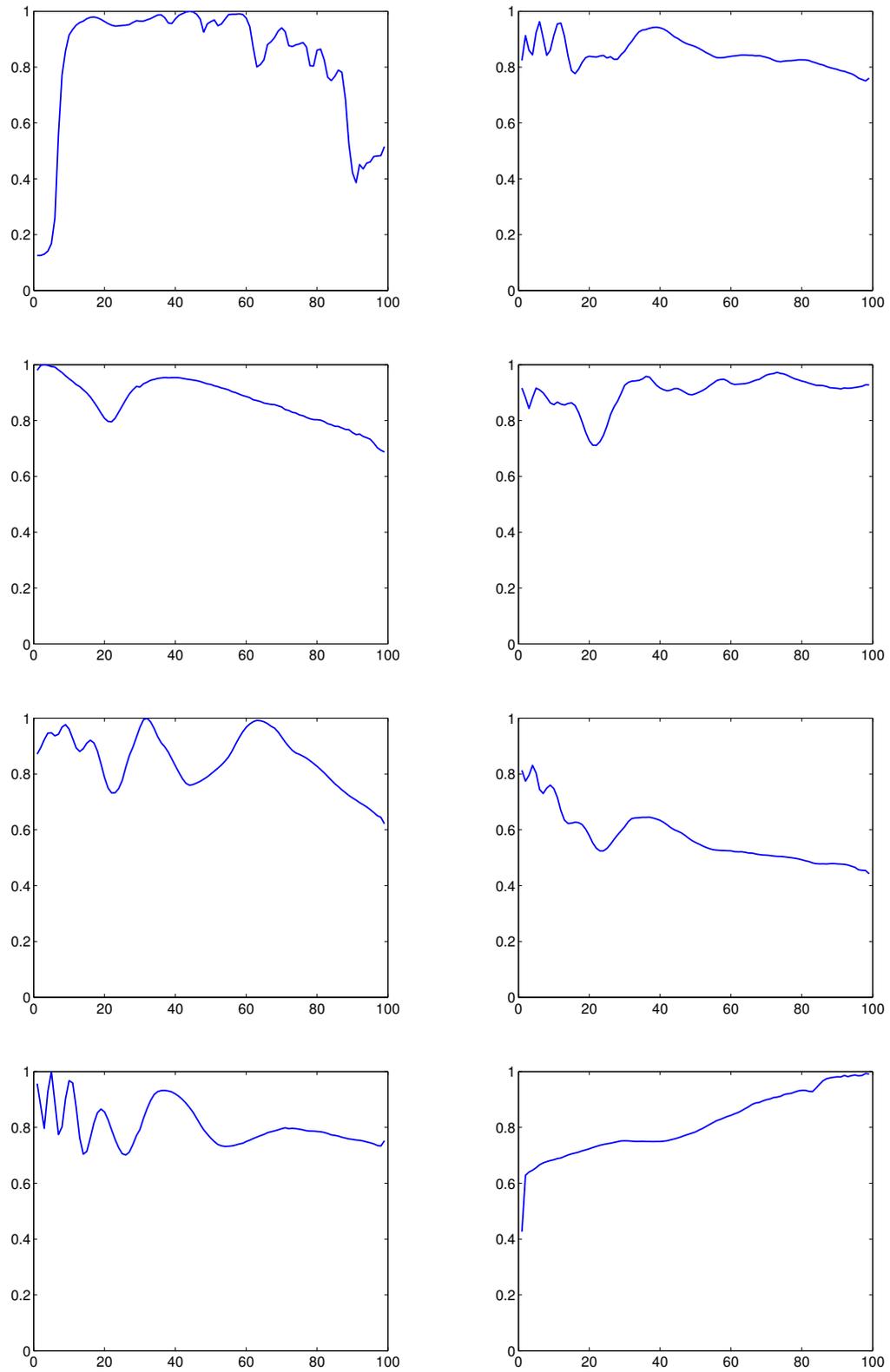
Figure 5.11: Spectral signatures of eight materials assigned to the simulated Hubble Telescope model.
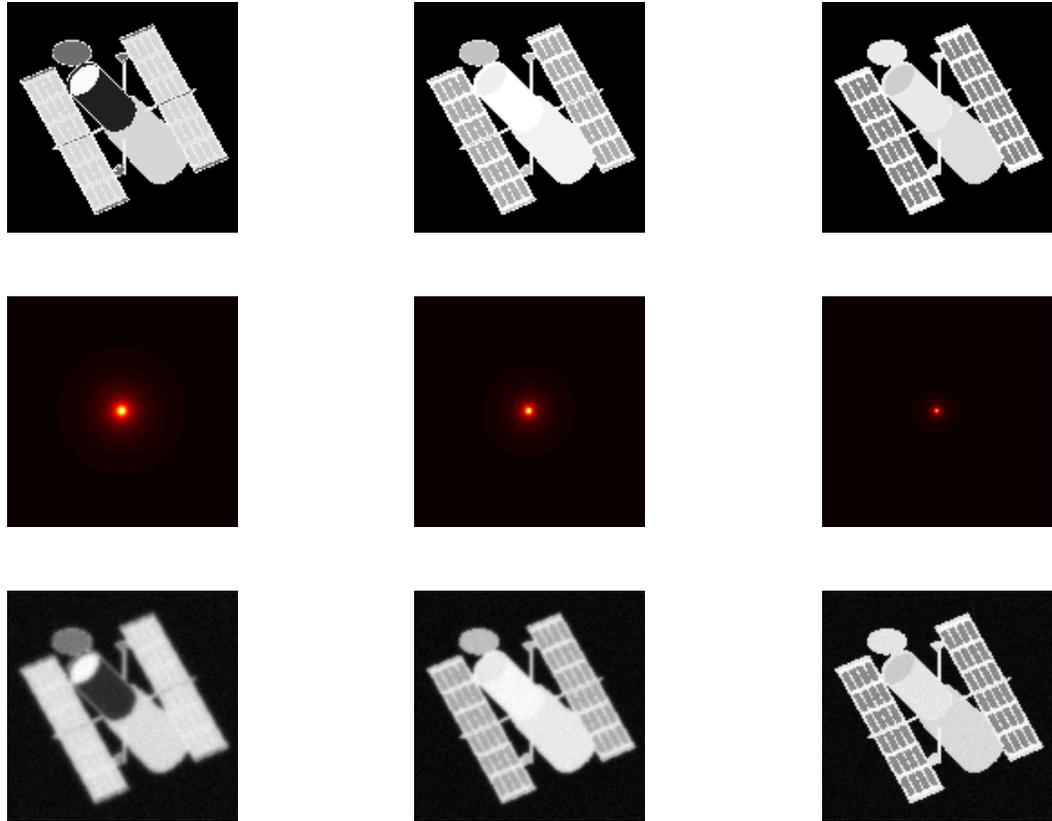
Figure 5.12: First row: the true columns of Y observed at different wavelengths (from left to right: 400nm, 1107.1nm, 1814.3nm). Second row: the circular Moffat PSFs used to blur the columns of Y at different wavelengths. Third row: the corresponding blurred and noisy columns of Y blurred with different Moffat PSFs corresponding to different wavelengths.

single PSF method. By using the preconditioner (presented in Chapter 4) we were able to reduce the number of iterations required for convergence to 20. In Figure 5.14 we show the relative residual norms for the first 95 iterations of preconditioned conjugate gradient (PCG) and conjugate gradient (CG) without preconditioning. The relative residual norms for PCG continue to decrease until they reach the default tolerance level of 10^{-6} whereas for CG the relative residual norms stagnate and oscillate in the interval between 10^{-2} and 10^{-3}. Figures 5.15 and 5.16 show the reconstructed columns of X for both the single PSF and multiple PSF methods.
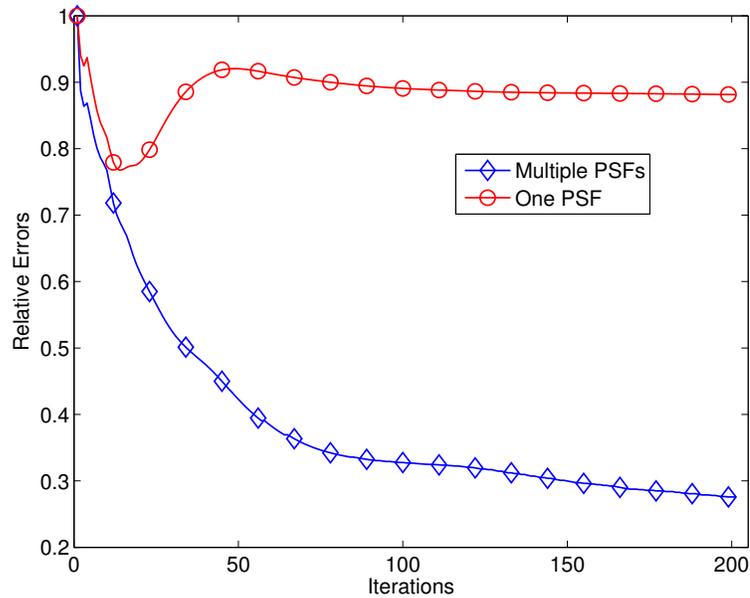
Figure 5.13: Relative errors for the computed fractional abundances using a single PSF and multiple PSFs.
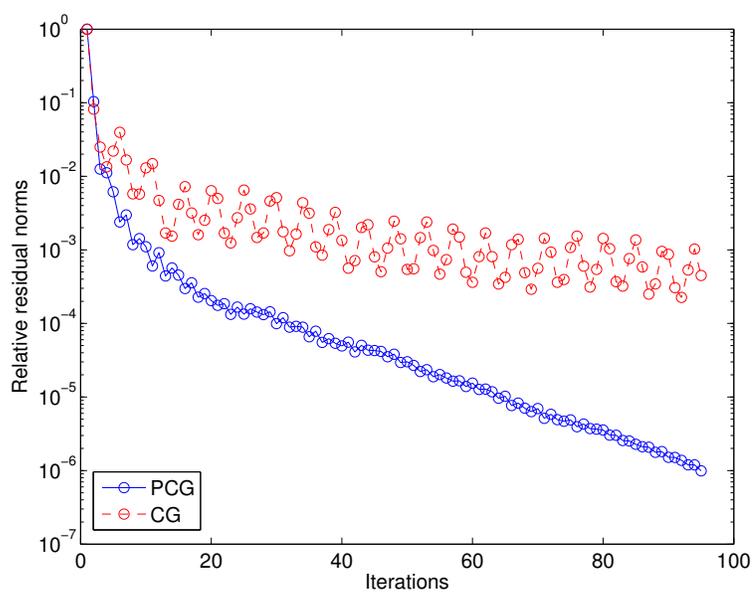
Figure 5.14: Relative residual norms for the first 95 iterations of PCG and CG.
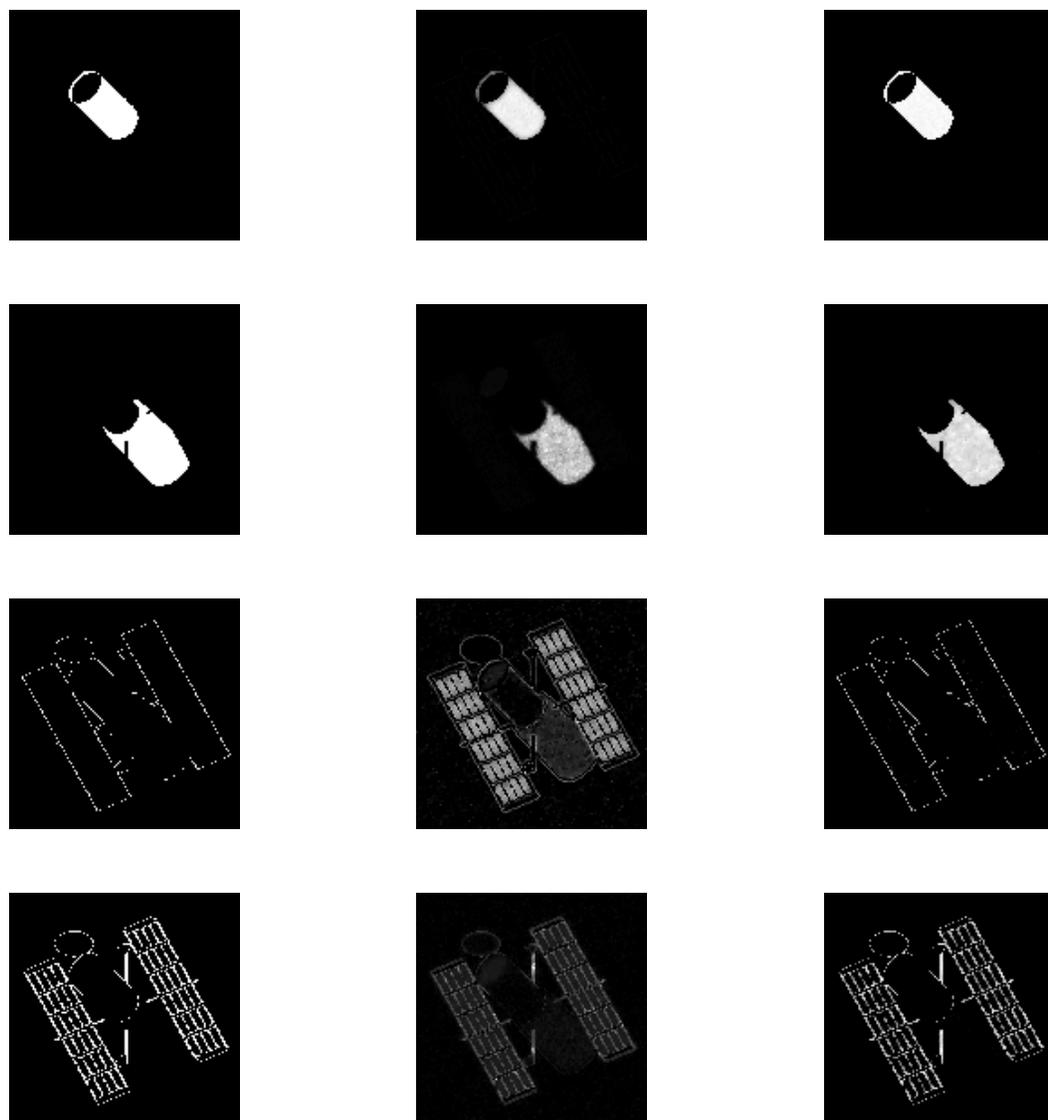
Figure 5.15: Fractional abundances for materials 1 to 4 (the first 4 materials in Table 5.1). First column: true fractional abundances; second column: the estimated fractional abundances using the single PSF approach; third column: the estimated fractional abundances using the multiple PSF approach.
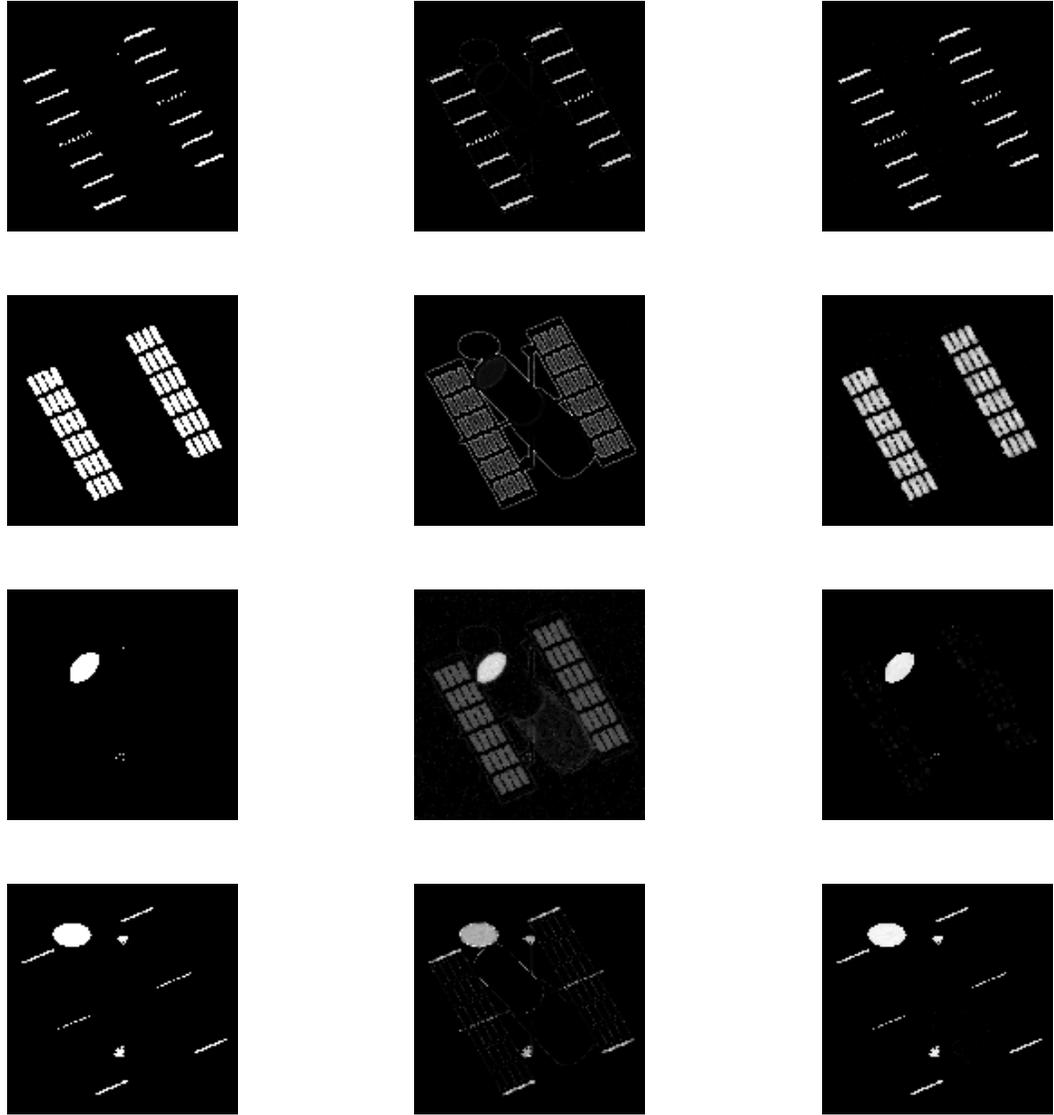
Figure 5.16: Fractional abundances for materials 5 to 8 (the last 4 materials in Table 5.1). First column: true fractional abundances; second column: the estimated fractional abundances using the single PSF approach; third column: the estimated fractional abundances using the multiple PSF approach.
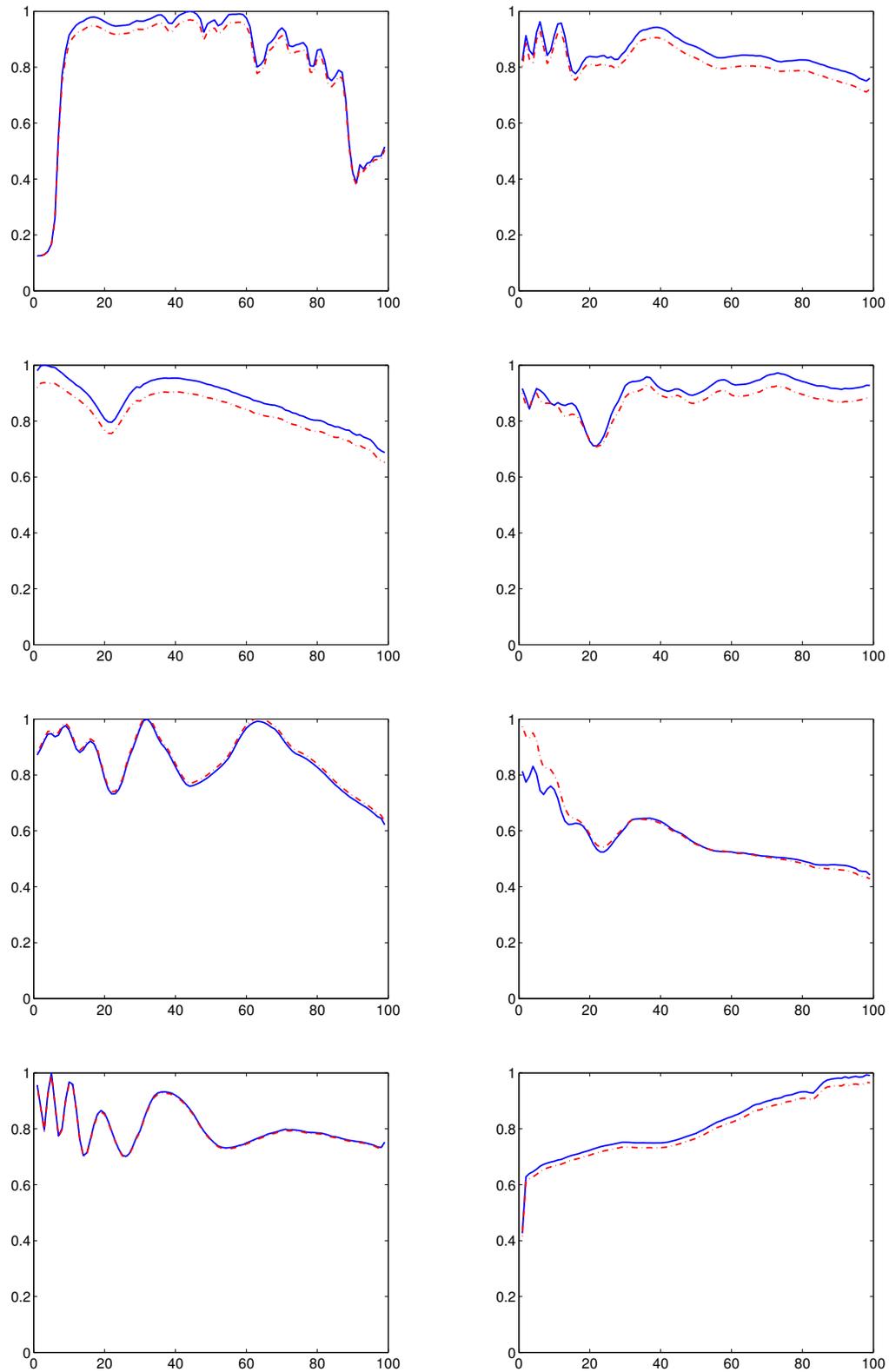
Figure 5.17: The true material spectral signatures (blue and −) and the computed spectral signatures (red and −.) after applying the multiple PSF numerical scheme on the blurred and noisy data.
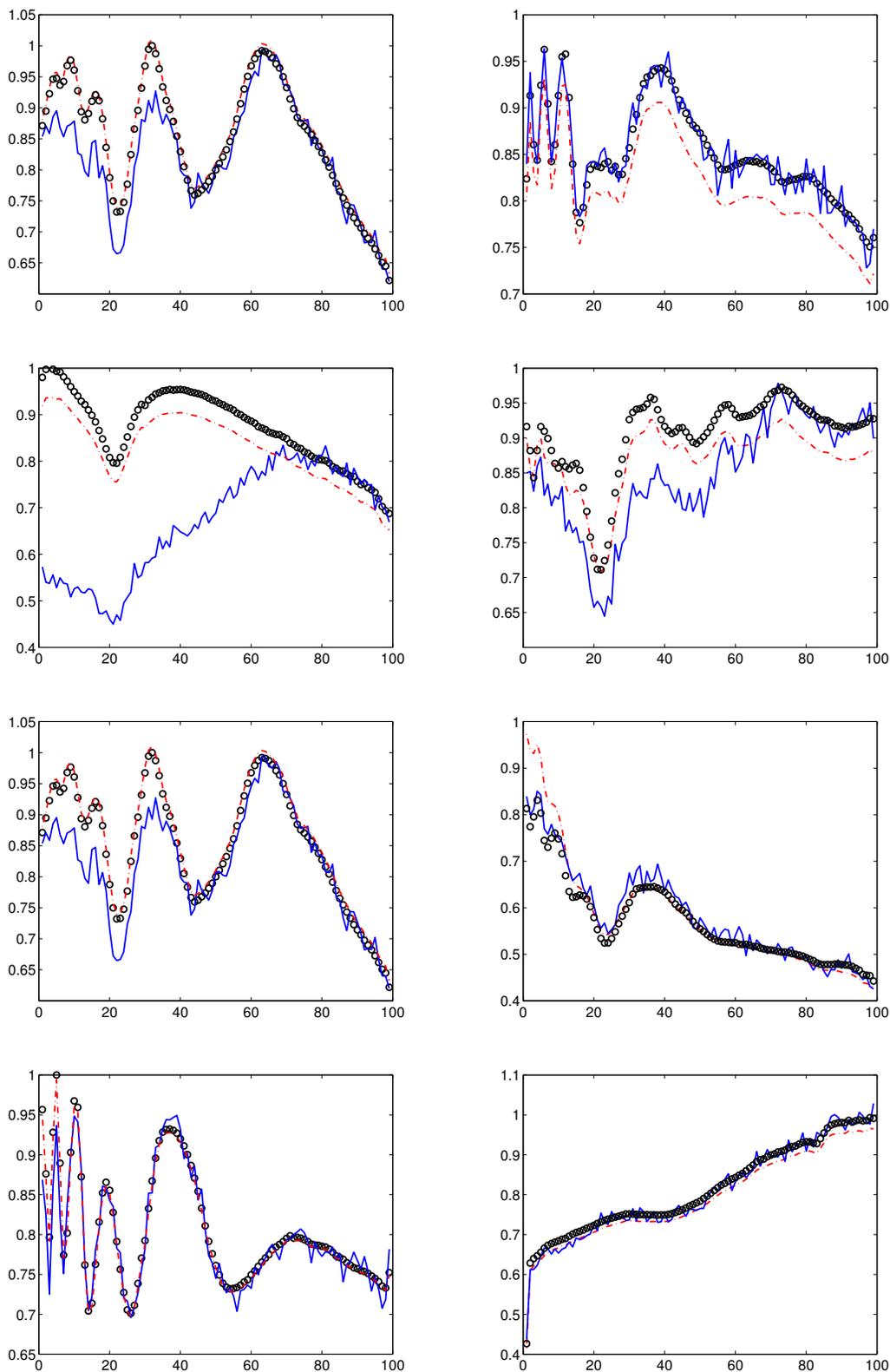
Figure 5.18: The true material spectral signatures (black and circle), the computed spectral signatures (red and −.) after applying the multiple PSF numerical scheme on the blurred and noisy data, and the observed blurred and noisy material spectral signatures (blue and −). Note that the y-axis has not been scaled in order to show more clearly the differences between spectral signatures in the three cases.

Chapter 6

Conclusions and Future Work

We have described linear and separable nonlinear inverse problems using an image restoration model problem. We discussed properties of ill-posed inverse problems and explained why regularization is necessary to obtain a meaningful solution. We have considered different regularization techniques for linear inverse problems. Preconditioning of inverse problems was discussed and we explained some considerations that should be taken into account in the case of preconditioning of ill-posed problems.

Hyperspectral imaging is now a mature technology which is finding applications in many fields. We have explained some important concepts in hyperspectral imaging ranging from the electromagnetic spectrum to the hyperspectral unmixing models that are used for hyperspectral data analysis and exploitation.

Next we described an iterative approach for solving nonlinear least squares problems related to hyperpsectral imaging. Circular and elliptical Moffat function models were used to define the spectrally and spatially varying hyperpsectral PSF. We observed that the Jacobian matrix associated with the nonlinear least squares formulation of the problem tends to be very ill-conditioned. This causes numerical instability in the solution of the Jacobian system inside the Gauss-Newton iteration, which in turn causes stagnation in the convergence history of relative errors for PSF param-

eters and star spectrum. We have addressed this problem and have shown that by combining a Gauss-Newton approach for minimizing a reduced cost functional with subset selection and regularization, for improving the conditioning of the Jacobian matrix, one can solve the large-scale nonlinear inverse problem for hyperpspectral PSF parameter identification and star spectrum reconstruction with higher accuracy. We have provided examples for simulated HSI data of an isolated star corresponding to different noise levels and different number of wavelengths.

Our work shows that a variable projection Gauss-Newton scheme can be used to estimate a wavelength varying PSF, from the hyperspectral datacube of an isolated star, with satisfactory results on simulated data. This approach allows one to jointly estimate the star spectrum and the hyperspectral PSF parameters without the need to tune the number of wavelengths used for spectral binning. We show that using subset selection to choose the most linearly independent columns of the Jacobian matrix can make the computations more robust in the case of elliptical Moffat PSF. Furthermore, our work indicates that combining subset selection with Tikhonov regularization avoids the need to know exactly the true value of one or more parameters.

The next important step would be to estimate the hyperspectral PSF parameters from real MUSE data of isolated bright stars, which are known to be faintly sampled spatially. In the future it is also important to account for noise from the imperfectly subtracted sky background. The hyperpsectral PSF model we used in our work neglects the MUSE instrument PSF and assumes that the data reduction software does not introduce any bias. In order to obtain better results with realistic data, it is important to take into consideration the possible degradations due to the instrument PSF and the data reduction software.

In the second part of our work on hyperspectral imaging, we combined the proposed hyperpsectral PSF estimation approach with hyperpsectral deblurring and sparse unmixing schemes. Specifically, we were interested in studying hyperspectral

unmixing and deblurring using the overall system dependent hyperspectral imaging PSF (estimated at multiple wavelengths) for space object identification applications. In particular, we considered deblurring and spectrally analyzing ground-based astronomical images of space objects. A numerical approach was provided for deblurring and sparse unmixing of ground-based hyperspectral images of objects taken through the atmosphere at multiple wavelengths with narrow spectral channels. Here, the problem is quite challenging since the point spread function (PSF) depends on the imaging system as well as the seeing conditions and is generally wavelength dependent. We provided a numerical method for joint deblurring and sparse unmixing in order to spectrally analyze the image objects. Our work shows that the number of iterations can be significantly reduced by using a simple conjugate gradient preconditioner to solve the $\boldsymbol{X}$-subproblem of the ADMM method for the multiple PSF case. The method was illustrated with numerical experiments on a commonly used test example, a simulated hyperspectral image of the Hubble Space Telescope satellite.

In this work we have assumed that the hyperpsectral PSF parameters have already been estimated prior to performing hyperspectral deblurring and unmixing. In future work we plan to estimate the hyperspectral PSF parameters and perform hyperspectral unmixing directly from hyperspectral data. We will also estimate the instrument PSF, which relates generally to axial optical aberrations, and combine this process with our proposed deblurring and sparse unmixing method for the multiple PSF case. We plan to use tensor decomposition to exploit the structure of the problem in constructing the preconditioner. We will test our methods with astronomical data collected by the AEOS Spectral Imaging Sensor (ASIS).

Bibliography

[1] AGGARWAL, S. Principles of remote sensing. *Satellite Remote Sensing and GIS Applications in Agricultural Meteorology* (2004).

[2] ALTMANN, Y., DOBIGEON, N., TOURNERET, J., AND BERMUDEZ, J. A robust test for nonlinear mixture detection in hyperspectral images. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)* (2013), pp. 2149–2153.

[3] ANDREWS, H. C., AND HUNT, B. R. *Digital Image Restoration.* Prentice-Hall, Englewood Cliffs, NJ, 1977.

[4] BACHMAYR, M., AND BURGER, M. Iterative total variation schemes for nonlinear inverse problems. *Inverse Problems 25*, 10 (2009).

[5] BARDSLEY, J., AND VOGEL, C. R. A nonnegativly constrained convex programming method for image reconstruction. *SIAM J. Sci. Comput. 25*, 4 (2004), 1326–1343.

[6] BARDSLEY, J. M. An efficient computational method for total variation-penalized Poisson likelihood estimation. *Inverse Problems and Imaging 2*, 2 (2008), 167–185.

[7] BARDSLEY, J. M. Stopping rules for a nonnegatively constrained iterative method for ill-posed Poisson imaging problems. *BIT 48*, 4 (2008), 651–664.

[8] BARDSLEY, J. M. An efficient MCMC method for uncertainty quantification in inverse problems. Tech. Rep. UM Math Tech Report #28, Mathematics Department, University of Montana, 2010.

[9] BARDSLEY, J. M., CALVETTI, D., AND SOMERSALO, E. Hierarchical regularization for edge-preserving reconstruction of PET images. *Inverse Problems 26*, 3 (2010), doi: 10.1088/0266–5611/26/3/035010.

[10] BARDSLEY, J. M., AND NAGY, J. G. Covariance-preconditioned iterative methods for nonnegatively constrained astronomical imaging. *SIAM J. Matrix Anal. Appl. 27* (2006), 1184–1197.

[11] BASEDOW, R., SILVERGLATE, P., RAPPOPORT, W., ROCKWELL, R., ROSENBERG, D., SHU, K., WHITTLESEY, R., AND ZALEWSKI, E. The hydice instrument design and its application to planetary instruments. In *Advanced Technologies for Planetary Instruments* (1993), vol. 1, p. 1.

[12] BENZI, M. Preconditioning techniques for large linear systems: A survey. *Journal of Computational Physics 182* (2002), 418–477.

[13] BERISHA, S., AND NAGY, J. Iterative image restoration. *R. Chellappa and S. Theodoridis, eds., Academic Press Library in Signal Processing 4* (2013), 193–243.

[14] BERTERO, M., AND BOCCACCI, P. *Introduction to Inverse Problems in Imaging.* Institute of Physics Publishing, London, 1998.

[15] BESAG, J. On the statistical analysis of dirty pictures (with discussion). *J. Royal Stat. Soc., Ser. B 48*, 5–6 (1986), 259–302.

[16] BESAG, J. Towards Bayesian image analysis. *J. Appl. Stat. 16* (1989), 395–406.

[17] BIENIARZ, J., CERRA, D., AVBELJ, J., REINARTZ, P., AND MÜLLER, R. Hyperspectral image resolution enhancement based on spectral unmixing and information fusion. Tech. rep., ISPRS XXXVIII-4, 2011.

[18] BIOUCAS-DIAS, J. M. A variable splitting augmented lagrangian approach to linear spectral unmixing. In *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 2009. WHISPERS'09. First Workshop on* (2009), IEEE, pp. 1–4.

[19] BIOUCAS-DIAS, J. M., AND NASCIMENTO, J. M. Hyperspectral subspace identification. *Geoscience and Remote Sensing, IEEE Transactions on 46*, 8 (2008), 2435–2445.

[20] BIOUCAS-DIAS, J. M., AND PLAZA, A. Hyperspectral unmixing: Geometrical, statistical, and sparse regression-based approaches. In *Remote Sensing* (2010), International Society for Optics and Photonics, pp. 78300A–78300A.

[21] BIOUCAS-DIAS, J. M., PLAZA, A., DOBIGEON, N., PARENTE, M., DU, Q., GADER, P., AND CHANUSSOT, J. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of 5*, 2 (2012), 354–379.

[22] BJÖRCK, Å. A bidiagonalization algorithm for solving large and sparse ill-posed systems of linear equations. *BIT 28*, 3 (1988), 659–670.

[23] BJÖRCK, Å., GRIMME, E., AND VAN DOOREN, P. An implicit shift bidiagonalization algorithm for ill-posed systems of linear equations. *BIT 34*, 4 (1994), 510–534.

[24] CALVETTI, C., HANSEN, P. C., AND REICHEL, L. L-curve curvature bounds via lanczos bidiagonalization. *Electron. Trans. Numer. Anal. 14* (2002), 20–35.

[25] CALVETTI, D., AND SOMERSALO, E. Image inpainting and bootstrap priors. *Image and Vision Computing 24* (2006), 782–793.

[26] CALVETTI, D., AND SOMERSALO, E. Gaussian hypermodel and recovery of blocky objects. *Inverse Problems 23* (2007), 733–754.

[27] CALVETTI, D., AND SOMERSALO, E. *Introduction to Bayesian Scientific Computing: Ten Lectures on Subjective Computing*, vol. 2. Springer Science+Business Media, LLC, New York, 2007.

[28] CALVETTI, D., AND SOMERSALO, E. Hypermodels in the Bayesian imaging framework. *Inverse Problems 24*, 3 (2008), doi: 10.1088/0266–5611/24/3/034013.

[29] CANDÉS, E. J., ROMBERG, J. K., AND TAO, T. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math. LIX* (2006), 1207–1223.

[30] CHAN, R. H., NAGY, J. G., AND PLEMMONS, R. J. FFT-based preconditioners for Toeplitz-block least squares. *SIAM J. Numer. Anal. 30* (1993), 1740–1768.

[31] CHAN, T. F. Rank revealing QR factorizations. *Linear Algebra and Its Applications 88* (1987), 67–82.

[32] CHAN, T. F., AND SHEN, J. *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*. SIAM, Philadelphia, PA, 2005.

[33] CHANG, C.-I. *Hyperspectral imaging: techniques for spectral detection and classification*, vol. 1. Springer, 2003.

[34] CHEN, J., RICHARD, C., AND HONEINE, P. A novel kernel-based nonlinear unmixing scheme of hyperspectral images. In *Signals, Systems and Computers*

(ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on (2011), IEEE, pp. 1898–1902.

[35] CHEN, J., RICHARD, C., AND HONEINE, P. Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model. *IEEE Transactions on Signal Processing 6* (2013), 480–492.

[36] CHEN, J., RICHARD, C., LANTÉRI, H., THEYS, C., AND HONEINE, P. A gradient based method for fully constrained least-squares unmixing of hyperspectral images. In *Statistical Signal Processing Workshop (SSP), 2011 IEEE* (2011), IEEE, pp. 301–304.

[37] CHUNG, J. *Numerical Approaches for Large-Scale Ill-Posed Inverse Problems.* PhD thesis, Emory University, 2009. `http://www.cs.umd.edu/~jmchung/Home/CV_files/Julianne_Chung_Dissertation.pdf`.

[38] CHUNG, J., KNEPPER, S., AND NAGY, J. G. Large-scale inverse problems in imaging. In *Handbook of Mathematical Methods in Imaging*, O. Scherzer, Ed. Springer, New York, 2010.

[39] CHUNG, J., NAGY, J. G., AND O'LEARY, D. P. A weighted GCV method for Lanczos hybrid regularization. *Elec. Trans. Numer. Anal. 28* (2008), 149–167.

[40] CLARK, R. N., AND ROUSH, T. L. Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *Journal of Geophysical Research: Solid Earth (1978–2012) 89*, B7 (1984), 6329–6340.

[41] DENNIS, J. E., AND SCHNABEL, R. B. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* SIAM, Philadelphia, 1996.

[42] Dobigeon, N., Tourneret, J.-Y., Richard, C., Bermudez, J., McLaughlin, S., and Hero, A. O. Nonlinear unmixing of hyperspectral images: models and algorithms. *arXiv preprint arXiv:1304.1875* (2013).

[43] Dópido, I., Villa, A., Plaza, A., and Gamba, P. A quantitative and comparative assessment of unmixing-based feature extraction techniques for hyperspectral image classification. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of 5*, 2 (2012), 421–435.

[44] Eastman, J. R. Guide to gis and image processing volume. *Clark Labs* (2001).

[45] Eismann, M. T. *Hyperspectral Remote Sensing*. SPIE Press, 2012.

[46] Engl, H. W., Hanke, M., and Neubauer, A. *Regularization of Inverse Problems*. Kluwer Academic Publishers, Dordrecht, 2000.

[47] Engl, H. W., and Kügler, P. Nonlinear inverse problems: Theoretical aspects and some industrial applications. In *Multidisciplinary Methods for Analysis Optimization and Control of Complex Systems*, V. Capasso and J. Périaux, Eds. Springer, Berlin, 2005, pp. 3–48.

[48] Engl, H. W., Kunisch, K., and Neubauer, A. Convergence rates for Tikhonov regularisation of nonlinear ill-posed problems. *Inverse Problems 5*, 4 (1999), 523–540.

[49] Epstein, C. L. *Introduction to the Mathematics of Medical Imaging, Second Edition*. SIAM, Philadelphia, PA, 2007.

[50] Eriksson, J., and Wedin, P. Truncated Gauss-Newton algorithms for ill-conditioned nonlinear least squares problems. *Optimization Methods and Software 19*, 6 (2004), 721–737.

[51] FIGUEIREDO, M. A., NOWAK, R. D., AND WRIGHT, S. J. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *Selected Topics in Signal Processing, IEEE Journal of 1*, 4 (2007), 586–597.

[52] FOSTER, L. V. Rank and null space calculations using matrix decomposition without column interchanges. *Linear Algebra and its Applications 74* (1986), 47–71.

[53] GILLIS, N., AND PLEMMONS, R. J. Dimensionality reduction, classification, and spectral mixture analysis using non-negative underapproximation. *Optical Engineering 50*, 2 (2011).

[54] GILLIS, N., AND PLEMMONS, R. J. Sparse nonnegative matrix underapproximation and its application to hyperspectral image analysis. *Linear Algebra and its Applications 438*, 10 (2013), 3991–4007.

[55] GOLUB, G., HEATH, M., AND WHABA, G. Generalized cross-validation as a method for choosing a good ridge paremeter. *Technometrics 21* (1979), 215–223.

[56] GOLUB, G., AND PEREYRA, V. The differentiation of pseudo-inverses and nonlinear least squares whose variables separate. *SIAM J. Numer. Anal. 10* (1973), 413–432.

[57] GOLUB, G., AND PEREYRA, V. Separable nonlinear least squares: the variable projection method and its applications. *Inverse Problems 19* (2003), R1–R26.

[58] GOLUB, G. H., HEATH, M., AND WAHBA, G. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics 21*, 2 (1979), 215–223.

[59] GOLUB, G. H., AND VAN LOAN, C. F. *Matrix computations*, vol. 3. JHU Press, 2012.

[60] GREENBAUM, A. *Iterative Methods for Solving Linear Systems.* SIAM, Philadelphia, 1997.

[61] GROETSCH, C. W. *The Theory of Tikhonov Regularization for Fredholm Integral Equations of the First Kind.* Pitman, Boston, 1984.

[62] GUO, Z., WITTMAN, T., AND OSHER, S. L1 unmixing and its application to hyperspectral image enhancement. In *SPIE Defense, Security, and Sensing* (2009), International Society for Optics and Photonics, pp. 73341M–73341M.

[63] HADAMARD, J. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin 13* (1902), 49–52.

[64] HANKE, M. On lanczos based methods for the regularization of discrete ill-posed problems. *BIT 41*, 5 (2001), 1008–1018.

[65] HANKE, M., AND NAGY, J. G. Restoration of atmospherically blurred images by symmetric indefinite conjugate gradient techniques. *Inverse Problems 12* (1996), 157–173.

[66] HANKE, M., NAGY, J. G., AND PLEMMONS, R. J. Preconditioned iterative regularization for ill-posed problems. In *Numerical Linear Algebra*, L. Reichel, A. Ruttan, and R. S. Varga, Eds. de Gruyter, Berlin, 1993, pp. 141–163.

[67] HANSEN, P. C. *Rank-deficient and discrete ill-posed problems.* SIAM, Philadelphia, PA, 1997.

[68] HANSEN, P. C. *Discrete Inverse Problems: Insight and Algorithms.* SIAM, Philadelphia, PA, 2010.

[69] HANSEN, P. C., NAGY, J. G., AND O'LEARY, D. P. *Deblurring Images: Matrices, Spectra and Filtering.* SIAM, Philadelphia, PA, 2006.

[70] HAPKE, B. Bidirectional reflectance spectroscopy: 1. theory. *Journal of Geophysical Research: Solid Earth (1978–2012) 86*, B4 (1981), 3039–3054.

[71] HENDRIX, E. M. T., GARCÍA, I., PLAZA, J., MARTÍN, G., AND PLAZA, A. A new minimum-volume enclosing algorithm for endmember identification and abundance estimation in hyperspectral data. *Geoscience and Remote Sensing, IEEE Transactions on 50*, 7 (2012), 2744–2757.

[72] HOFMANN, B. Regularization of nonlinear problems and the degree of ill-posedness. In *Inverse Problems: Principles and Applications in Geophysics, Technology, and Medicine*, G. Anger, R. Gorenflo, H. Jochmann, H. Moritz, and W. Webers, Eds. Akademie Verlag, Berlin, 1993.

[73] HSU, S. M., AND BURKE, H.-H. K. Multisensor fusion with hyperspectral imaging data: detection and classification. *Lincoln laboratory journal 14*, 1 (2003), 145–159.

[74] IORDACHE, D. A sparse regression approach to hyperspectral unmixing. *Universidade Técnica de Lisboa, PhD thesis* (2011).

[75] IORDACHE, M.-D., BIOUCAS-DIAS, J. M., AND PLAZA, A. Total variation spatial regularization for sparse hyperspectral unmixing. *Geoscience and Remote Sensing, IEEE Transactions on 50*, 11 (2012), 4484–4502.

[76] IORDACHE, M.-D., BIOUCAS-DIAS, J. M., AND PLAZA, A. Collaborative sparse regression for hyperspectral unmixing. *Geoscience and Remote Sensing, IEEE Transactions on 52*, 1 (2014), 341–354.

[77] IPSEN, I. C. F., KELLEY, C. T., AND POPE, S. R. Rank-deficient nonlinear least squares problems and subset selection. *SIAM Journal on Numerical Analysis 49*, 3 (2011), 1244–1266.

[78] JAIN, A. K. *Fundamentals of Digital Image Processing*. Prentice-Hall, Englewood Cliffs, NJ, 1989.

[79] JEFFERIES, S. M., AND HART, M. Deconvolution from wavefront sensing using the frozen flow hypothesis. *Optics express 19*, 3 (2011), 1975–1984.

[80] KAUFMAN, L. A variable projection method for solving separable nonlinear least squares problems. *BIT 15* (1975), 49–57.

[81] KELLEY, C. T. *Iterative methods for optimization*, vol. 18. Society for Industrial and Applied Mathematics, 1987.

[82] KESHAVA, N. A survey of spectral unmixing algorithms. *Lincoln Laboratory Journal 14*, 1 (2003), 55–78.

[83] KILMER, M. E., HANSEN, P. C., AND ESPAÑOL, M. I. A projection-based approach to general-form Tikhonov regularization. *SIAM J. Sci. Comput. 29*, 1 (2007), 315–330.

[84] KILMER, M. E., AND O'LEARY, D. P. Choosing regularization parameters in iterative methods for ill-posed problems. *SIAM J. Matrix Anal. Appl. 22* (2001), 1204–1221.

[85] KUMAR, U., RAJA, S. K., MUKHOPADHYAY, C., AND RAMACHANDRA, T. A multi-layer perceptron based non-linear mixture model to estimate class abundance from mixed pixels. In *Students' Technology Symposium (TechSym), 2011 IEEE* (2011), IEEE, pp. 148–153.

[86] LAGENDIJK, R. L., AND BIEMOND, J. *Iterative Identification and Restoration of Images.* Kluwer Academic Publishers, Boston/Dordrecht/London, 1991.

[87] LAMBERT, A. J., AND NICHOLS, G. Wavelength diversity in restoration from atmospheric turbulence effected surveillance imagery. In *Frontiers in Optics 2009, Laser Science XXV, Fall 2009 OSA Optics & Photonics Technical Digest, OSA Technical Digest (CD).*

[88] LANDWEBER, L. An iteration formula for Fredholm integral equations of the first kind. *Amer. J. of Math. 73*, 3 (1951), 615–624.

[89] LARSEN, R. M. Lanczos bidiagonalization with partial reorthogonalization. *DAIMI PB 27*, 537 (1998).

[90] LI, C., SUN, T., KELLY, K. F., AND ZHANG, Y. A compressive sensing and unmixing scheme for hyperspectral data processing. *Image Processing, IEEE Transactions on 21*, 3 (2012), 1200–1210.

[91] LI, F., NG, M. K., PLEMMONS, R., PRASAD, S., AND ZHANG, Q. Hyperspectral image segmentation, deblurring, and spectral analysis for material identification. In *SPIE Defense, Security, and Sensing* (2010), International Society for Optics and Photonics, pp. 770103–770103.

[92] LI, F., NG, M. K., AND PLEMMONS, R. J. Coupled segmentation and denoising/deblurring models for hyperspectral material identification. *Numerical Linear Algebra with Applications 19*, 1 (2012), 153–173.

[93] MA, W., BIOUCAS-DIAS, J., GADER, P., CHAN, T., GILLIS, N., PLAZA, A., AMBIKAPATHI, A., AND CHI, C. A signal processing perspective on hyperspectral unmixing. *IEEE Signal Processing Mag 31*, 1 (2013), 67–81.

[94] Manolakis, D., Marden, D., and Shaw, G. A. Hyperspectral image processing for automatic target detection applications. *Lincoln Laboratory Journal 14*, 1 (2003), 79–116.

[95] Martin, G., Bioucas-Dias, J. M., and Plaza, A. A new technique for hyperspectral compressive sensing using spectral unmixing. In *SPIE Optical Engineering+ Applications* (2012), International Society for Optics and Photonics, pp. 85140N–85140N.

[96] MATLAB. *Student version 7.10.0 (R2010a)*. The MathWorks Inc., Natick, Massachusetts, 2010.

[97] Matson, C. L., Borelli, K., Jefferies, S., Beckner Jr, C. C., Hege, E. K., Lloyd-Hart, M., et al. Fast and optimal multiframe blind deconvolution algorithm for high-resolution ground-based imaging of space objects. *Applied Optics 48*, 1 (2009), A75–A92.

[98] Mianji, F. A., and Zhang, Y. Svm-based unmixing-to-classification conversion for hyperspectral abundance quantification. *Geoscience and Remote Sensing, IEEE Transactions on 49*, 11 (2011), 4318–4327.

[99] Miller, K. Least squares methods for ill-posed problems with a prescribed bound. *SIAM J. Math Anal 1*, 1 (1970), 52–74.

[100] Mishchenko, M. I., Travis, L. D., and Lacis, A. A. *Scattering, absorption, and emission of light by small particles*. Cambridge university press, 2002.

[101] Moses, W. J. *Atmospheric Correction of Hyperspectral Data–Execution of and Comparison Between FLAASH and TAFKAA_6S*. PhD thesis, Cornell University, 2005.

[102] NAGY, J. G., AND O'LEARY, D. P. Fast iterative image restoration with a spatially varying PSF. In *Advanced Signal Processing Algorithms, Architectures, and Implementations VII* (1997), F. T. Luk, Ed., vol. 3162, SPIE, pp. 388–399.

[103] NAGY, J. G., AND O'LEARY, D. P. Restoring images degraded by spatially-variant blur. *SIAM J. Sci. Comput. 19* (1998), 1063–1082.

[104] NAGY, J. G., PALMER, K. M., AND PERRONE, L. Iterative methods for image deblurring: A Matlab object oriented approach. *Numerical Algorithms 36* (2004), 73–93. See also: http://www.mathcs.emory.edu/~nagy/RestoreTools.

[105] NAGY, J. G., PAUCA, V. P., PLEMMONS, R. J., AND TORGERSEN, T. C. Space-varying restoration of optical images. *J. Optical Soc. Amer. A 14* (1997), 3162–3174.

[106] NAGY, J. G., PLEMMONS, R. J., AND TORGERSEN, T. C. Iterative image restoration using approximate inverse preconditioning. *IEEE Trans. on Image Processing 15* (1996), 1151–1162.

[107] NASCIMENTO, J. M. P., AND BIOUCAS-DIAS, J. M. Hyperspectral unmixing algorithm via dependent component analysis. In *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International* (2007), IEEE, pp. 4033–4036.

[108] NATTERER, F. *The Mathematics of Computerized Tomography.* Classics in Applied Mathematics. SIAM, 2001.

[109] NOCEDAL, J., AND WRIGHT, S. *Numerical Optimization.* Springer, New York, 1999.

[110] O'LEARY, D. P., AND SIMMONS, J. A. A bidiagonalization-regularization procedure for large scale discretizations of ill-posed problems. *SIAM J. Sci. Stat. Comp. 2* (1981), 474–489.

[111] ORTEGA, J. M., AND RHEINBOLDT, W. C. *Iterative solution of nonlinear equations in several variables*, vol. 30. Society for Industrial and Applied Mathematics, 1987.

[112] OSBORNE, M. Separable least squares, variable projection, and the gauss-newton algorithm. *Electronic Transactions on Numerical Analysis 28*, 2 (2007), 1–15.

[113] OLEARY, D. P., AND RUST, B. W. Variable projection for nonlinear least squares problems. *Computational Optimization and Applications 54*, 3 (2013), 579–593.

[114] PAIGE, C. C., AND SAUNDERS, M. A. LSQR: An algorithm for sparse linear equations and sparse least squares. *ACM Trans. Math. Software 8*, 1 (1982), 43–71.

[115] PAPP, É., AND CUDAHY, T. Hyperspectral remote sensing. *Geophysical and Remote Sensing Methods for Regolith Exploration 144* (2002), 13–21.

[116] PARGAL, S. Hyperspectral subspace identification and endmember extraction by integration of spatial-spectral information.

[117] PHILLIPS, D. L. A technique for the numerical solution of certain integral equations of the first kind. *Journal of the Association for Computing Machinery 9*, 1 (1962), 84–97.

[118] PLAZA, A., AND CHANG, C.-I. An improved n-findr algorithm in implementation. In *Defense and Security* (2005), International Society for Optics and Photonics, pp. 298–306.

[119] PLAZA, A., MARTINEZ, P., PÉREZ, R., AND PLAZA, J. Spatial/spectral endmember extraction by multidimensional morphological operations. *Geoscience and Remote Sensing, IEEE Transactions on 40*, 9 (2002), 2025–2041.

[120] RAJWADE, A., KITTLE, D., TSAI, T.-H., BRADY, D., AND CARIN, L. Coded hyperspectral imaging and blind compressive sensing, 2012.

[121] RICHARDS, J. A., AND JIA, X. *Remote sensing digital image analysis*, vol. 3. Springer, 1999.

[122] ROGGEMANN, M. C., AND WELSH, B. M. *Imaging through turbulence*. CRC press, 1996.

[123] RUDIN, L., OSHER, S., AND FATEMI, E. Nonlinear total variation based noise removal algorithms. *Physica D 60* (1992), 259–268.

[124] RUHE, A., AND WEDIN, P. Algorithms for separable nonlinear least squares problems. *SIAM Review 22*, 3 (1980), 318–337.

[125] SAAD, Y. *Iterative Methods for Sparse Linear Systems*. PWS Publishing Company, Boston, 1996.

[126] SCHMIDT, F., SCHMIDT, A., TRÉGUIER, E., GUIHENEUF, M., MOUSSAOUI, S., AND DOBIGEON, N. Implementation strategies for hyperspectral unmixing using bayesian source separation. *Geoscience and Remote Sensing, IEEE Transactions on 48*, 11 (2010), 4003–4013.

[127] SERRE, D., VILLENEUVE, E., CARFANTAN, H., JOLISSAINT, L., MAZET, V., BOURGUIGNON, S., AND JARNO, A. Modeling the spatial psf at the vlt focal

plane for muse wfm data analysis purpose. In *SPIE Astronomical Telescopes and Instrumentation: Observational Frontiers of Astronomy for the New Decade* (2010), International Society for Optics and Photonics, pp. 773649–773649.

[128] SHAW, G. A., AND BURKE, H.-H. K. Spectral imaging for remote sensing. *Lincoln Laboratory Journal 14*, 1 (2003), 3–28.

[129] SHIPPERT, P. Introduction to hyperspectral image analysis. *Online Journal of Space Communication 3* (2003).

[130] SINGER, R. B., AND MCCORD, T. B. Mars-large scale mixing of bright and dark surface materials and implications for analysis of spectral reflectance. In *Lunar and Planetary Science Conference Proceedings* (1979), vol. 10, pp. 1835–1848.

[131] SMITH, R. B. Introduction to hyperspectral imaging. *Microimages. Retrieved on June 30* (2006), 2008.

[132] SOULEZ, F., THIEBAUT, E., AND DENIS, L. Restoration of hyperspectral astronomical data with spectrally varying blur. *European Astronomical Society Publications Series 59* (2013), 403–416.

[133] ŠPICLIN, Ž., PERNUŠ, F., AND LIKAR, B. Correction of axial optical aberrations in hyperspectral imaging systems. In *Proc. of SPIE Vol* (2011), vol. 7891, pp. 78910S–1.

[134] THEMELIS, K. E., RONTOGIANNIS, A. A., AND KOUTROUMBAS, K. Sparse semi-supervised hyperspectral unmixing using a novel iterative bayesian inference algorithm.

[135] TIKHONOV, A. N. Regularization of incorrectly posed problems. *Sov Math Dockl 4* (1963), 1624–1627.

[136] TIKHONOV, A. N. Solution of incorrectly formulated problems and the regularization method. *Soviet Math. Doklady 4* (1963), 501–504.

[137] TIKHONOV, A. N., AND ARSENIN, A. N. *Solutions of ill-posed problems.* Winston & Sons, Washington, 1977.

[138] TIKHONOV, A. N., LEONOV, A. S., AND YAGOLA, A. G. *Nonlinear Ill-Posed Problems, Volumes I and II.* Chapman and Hall, London, 1998.

[139] TRUJILLO, I., AGUERRI, J. A. L., CEPA, J., AND GUTIÉRREZ, C. M. The effects of seeing on sérsic profiles–ii. the moffat psf. *Monthly Notices of the Royal Astronomical Society 328*, 3 (2001), 977–985.

[140] TSAIG, Y., AND DONOHO, D. L. Extensions of compressed sensing. *Signal Processing 86*, 3 (2006), 549–571.

[141] VILLENEUVE, E., CARFANTAN, H., AND SERRE, D. Psf estimation of hyperspectral data acquisition system for ground-based astrophysical observations. In *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2011 3rd Workshop on* (2011), IEEE, pp. 1–4.

[142] VOGEL, C. R. *Computational Methods for Inverse Problems.* SIAM, Philadelphia, PA, 2002.

[143] ZARE, A., AND GADER, P. Pce: Piecewise convex endmember detection. *Geoscience and Remote Sensing, IEEE Transactions on 48*, 6 (2010), 2620–2632.

[144] ZHANG, L., WU, B., HUANG, B., AND LI, P. Nonlinear estimation of subpixel proportion via kernel least square regression. *International Journal of Remote Sensing 28*, 18 (2007), 4157–4172.

[145] ZHANG, Q., PLEMMONS, R., KITTLE, D., BRADY, D., AND PRASAD, S. Joint segmentation and reconstruction of hyperspectral data with compressed measurements. *Applied Optics 50*, 22 (2011), 4417–4435.

[146] ZHANG, Q., WANG, H., PLEMMONS, R., AND PAUCA, V. Tensor methods for hyperspectral data analysis: a space object material identification study. *Journal of the Optical Society of America A 25*, 12 (2008), 3001–3012.

[147] ZHAO, X.-L., WANG, F., HUANG, T.-Z., NG, M. K., AND PLEMMONS, R. J. Deblurring and sparse unmixing for hyperspectral images. *Geoscience and Remote Sensing, IEEE Transactions on 51*, 7 (2013), 4045–4058.