Estimation of Atmospheric PSF Parameters for Hyperspectral Imaging

S. Berisha†, J. G. Nagy† and R. J. Plemmons††

†Department of Mathematics and Computer Science, Emory University, Atlanta, GA, USA.
††Departments of Computer Science and Mathematics, Wake Forest University, Winston-Salem, NC, USA.

SUMMARY

We present an iterative approach to solve separable nonlinear least squares problems arising in the estimation of wavelength-dependent point spread function (PSF) parameters for hyperspectral imaging. A variable projection Gauss-Newton method is used to solve the nonlinear least squares problem. An 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 other regularization techniques. Experimental results related to hyperspectral PSF parameter identification and star spectrum reconstruction illustrate the effectiveness of the resulting numerical scheme. Copyright © 2015 John Wiley & Sons, Ltd.

KEY WORDS: Hyperspectral imaging, PSF estimation, regularization

1. INTRODUCTION

Spectral imaging data represent 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 the image of the same scene corresponding to a particular wavelength. Let \( S \) denote the spectral cube. Then \( S_{ijk} \) is the measured reflection at the \( k \)th 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. Figure 1 shows an example of a hyperspectral data cube. For illustration purposes the front image of the datacube is a false color image created by transforming hyperspectral image reflectances into an RGB color image.

We denote the reflection intensities of a pixel \((i, j)\) at \( N_w \) wavelengths by a vector \( s_{(ij)} = [s_{ij}(\lambda_1) \ s_{ij}(\lambda_2) \ \cdots \ s_{ij}(\lambda_{N_w})]^T \), where \( \lambda_k \) represents the \( k \)th wavelength, for \( k = 1, \cdots, N_w \). The vector \( s_{(ij)} \) is often referred to as the spectral signature or spectral vector 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.

*Correspondence to: James G. Nagy, Emory University, Atlanta, GA, USA. nagy@mathcs.emory.edu. Work of J. G. Nagy was supported by grant no. DMS-1115627 from the US National Science Foundation, and grant no. AF9550-12-1-0084 from the US Air Force Office of Scientific Research. Work of R. J. Plemmons was supported by grant no. FA9550-11-1-0194 from the US Air Force Office of Scientific Research and by contract no. HM1582-10-C-0011 from the US National Geospatial-Intelligence Agency.

Copyright © 2015 John Wiley & Sons, Ltd.

Prepared using nlaauth.cls (Version: 2010/05/13 v2.00)
Multispectral imaging (MSI) data are spectral data collected at a few different wavelengths while the spectral data associated with hundreds or even thousands of wavelengths in a continuous spectral range are referred to as hyperspectral imaging (HSI) data. For comprehensive discussions see [1, 29, 34] and the references therein.

Hyperspectral imaging is used in a wide range of applications, such as remote surveillance, astrophysics, medical imaging, geophysics, agriculture, mineral exploration, etc. [23, 29, 35]. One advantage of hyperspectral imaging consists in its usage to perform object identification and classification in image scenes that contain objects that are too small to be detected by spatial imaging methods. However, 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 work has been done in developing hyperspectral image processing methods for HSI data compression, spectral signature identification of constituent materials, and determination of their corresponding fractional abundances [1, 6, 7, 11, 12, 17, 18, 22, 24, 30, 40, 41].

Knowledge of the data acquisition system is required to be able to effectively perform hyperspectral image processing tasks. In particular, the data acquisition system induces 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 contributes to the mixing of the spectrum of different objects present in the imaged scene. The hyperspectral PSFs are generally dependent on the image acquisition system and the atmospheric seeing conditions [36, 39]. 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 [36] the authors assume the model of the PSF is a linear combination of Gaussian functions and thus the blur identification process corresponds to finding the scalar weights for each of the Gaussian functions. In [39] the authors model the PSFs at different wavelengths by Moffat functions [25], 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 in 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 paper we use the same PSF model described in [39], but 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.

This paper is outlined as follows. The mathematical framework is set up in Section 2. Specifically, we describe the data formation model, the PSF model used, and the general problem formulation for multispectral wavelengths. In Section 3 we discuss the approach we use, based on solving a separable nonlinear least squares problem, to jointly estimate the hyperspectral PSF parameters and to reconstruct the star spectrum. The structure of the Jacobian matrix is discussed in detail. Numerical results are given in Section 4, and concluding remarks are given in Section 5.

2. 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 spatially invariant, and an elliptical model that can be used for spatially variant blurs.

2.1. PSF Star Image Formation Model

We follow the noisy data model used in [39], where the image formation process for an isolated star, often called a guide star in general ground-based astronomical imaging, e.g. [31], observed at a particular wavelength can be written as:

$$b_{\lambda} = h_{\lambda}s_{\lambda} + \eta_{\lambda},$$

where $b_{\lambda}$ is a vector representing the vectorized form of an observed, blurred, and noisy image of an isolated star corresponding to wavelength $\lambda$, $h_{\lambda}$ is a vector representing the vectorized form of an exact original image of the isolated star corresponding to wavelength $\lambda$, $s_{\lambda}$ is a scalar representing the unknown intensity of the star spectrum at wavelength $\lambda$, and $\eta_{\lambda}$ represents noise and other errors in the data. In this model we assume that the sky background has been subtracted from the observed isolated star images. Note that since we model the PSF using an image of a star, $h_{\lambda}$ also represents the hyperspectral PSF, which is, in general, unknown. However, it is reasonable to assume that a parametric model of $h_{\lambda}$ is known, where the parameters defining $h_{\lambda}$ are wavelength dependent. That is, by assuming a parametrized formula for the PSF, the image formation model becomes

$$b_{\lambda} = h(\phi_{\lambda})s_{\lambda} + \eta_{\lambda},$$

where $\phi_{\lambda}$ is a vector of unknown parameters corresponding to wavelength $\lambda$. The aim, then, is given measured data $b_{\lambda}$ at known wavelengths $\lambda_1, \lambda_2, \ldots, \lambda_{N_w}$, jointly compute estimates of the unknowns $\phi_{\lambda}$ and $s_{\lambda}$. Our approach will 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 $h(\phi_{\lambda})$. We concentrate on two popular parametric models first described by Moffat [25], the circular and elliptical Moffat functions. The terms “circular” and “elliptical” are based on the geometric structures of the PSFs resulting from these two models.

2.2. Circular Moffat

In [33] 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 $\alpha$ and a shape parameter $\beta$. In this case, $\phi = [\alpha\ \beta]^T$, and the PSF has the form:

$$h(\phi)_{i_x,j_y} = h(\alpha, \beta)_{i_x,j_y} = \left(1 + \frac{i_x^2 + j_y^2}{\alpha^2}\right)^{-\beta}.$$  \hspace{1cm} (1)

The flux of the Moffat function is $\int \int h(\alpha, \beta)_{i_x,j_y} \, di_x \, dj_y = \frac{\pi \alpha^2}{\beta - 1}$. 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 [37] that the analytical approximation in (1) provides the best fit to the PSF predicted from atmospheric turbulence theory when $\beta \approx 4.765$. In hyperspectral imaging the PSF is system dependent and it varies with wavelength. In [39] the authors have shown that the variation of the PSF with respect to $\lambda$ can be modeled by fixing the shape parameter $\beta$ as a constant, $\beta(\lambda) = \beta_0$, and by varying the scale parameter $\alpha$ linearly with respect to the wavelengths, $\alpha(\lambda) = \alpha_0 + \alpha_1 \lambda$.

Using this model the parameter vector becomes $\phi = [\alpha_0\ \alpha_1\ \beta]^T$ and the normalized wavelength varying PSF takes the form:

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

Modeling the hyperspectral PSF at a particular wavelength by a circular Moffat function provides a simple model involving only 3 parameters. Figure 2 shows a variety of hyperspectral PSFs corresponding to different wavelengths. We can observe from this figure that the spread of the circular Moffat PSFs is greater at the shorter wavelengths. The same situation is true for elliptical Moffat PSFs, which is discussed in the next subsection, and the effect follows from their defining equations. Specifically, it is well-known that atmospheric turbulence causes much more blur at short wavelengths than at longer ones, which is corroborated by the standard von Karman and Kolmogorov models; see, for example, Roggemann and Welsh [31]. A mesh plot of the Moffat PSF corresponding to the short wavelength $\lambda = 465\text{nm}$ and defined by the parameters $\phi = [\alpha_0\ \alpha_1\ \beta]^T = [2.51\ -7.2 \cdot 10^{-4}\ \ 3.31]^T$ is shown in Figure 3.

2.3. 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. In particular, if MUSE is operated with a Ground Layer Adaptive Optics (GLAO) device then the atmospheric perturbation is not corrected uniformly in the field of view, resulting in a spatially variant PSF [33]. In this case the PSFs are no longer circularly symmetric, and are more accurately modeled with an elliptical Moffat function,

$$h(\phi)_{i_x,j_y} = h(\alpha, \beta, \gamma, \Theta)_{i_x,j_y} = \left[1 + \frac{1}{\alpha^2} \left(\frac{i_x^2 + j_y^2}{\gamma^2}\right)\right]^{-\beta}, \hspace{1cm} (3)$$

where $\gamma$ is the ellipticity parameter, $\Theta$ is the rotation angle, and

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

The property $h(\alpha, \beta, \gamma, \Theta)_{i_x,j_y} = h(\alpha, \beta, \frac{1}{\gamma}, \Theta - \frac{\pi}{2})_{i_x,j_y}$ imposes boundary conditions on the ellipticity and rotation angle parameters, specifically: $1 \leq \gamma < \infty$ and $0 \leq \Theta < \frac{\pi}{2}$. In [33] the authors have proposed a model for the variation of the elliptical Moffat PSF with respect to $\lambda$ and the polar coordinates $(\rho, \theta)$ in the field of view. We use the same model for the variation of the hyperspectral PSF parameters.
Figure 2. 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). PSF images are of size $64 \times 64$ and are defined by parameters: $\alpha_0 = 2.51, \alpha_1 = -7.2 \cdot 10^{-4}, \beta = 3.31$.

Figure 3. A circular Moffat PSF as given in equation (2) with $\beta = 3.31, \alpha_0 = 2.51, \alpha_1 = -7.2 \cdot 10^{-4}$, and $\lambda = 465$ nm.

The ellipticity parameter, $\gamma$, 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, $\beta$, is kept as a constant while the variations of the scale parameter $\alpha$ are modeled as a linear function of $\rho$ and a quadratic function of $\lambda$, specifically: $\alpha(\lambda, \rho) = \alpha_0 + \alpha_1 \rho + \alpha_2 \lambda + \alpha_3 \lambda^2$.

The orientation parameter, $\Theta$, is modeled as a decreasing function of $\theta$ in the field of view, specifically $\Theta = \frac{\pi}{2} - \theta$. Figure 4 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 5.

Figure 4. Image, and corresponding mesh plot, of an isolated star modeled by the elliptical Moffat PSF with wavelength $\lambda = 465$nm and parameters: $\alpha_0 = 3.75, \alpha_1 = -2.99 \cdot 10^{-3}, \alpha_2 = -4.31 \cdot 10^{-3}, \alpha_3 = 1.98 \cdot 10^{-6}, \beta = 1.74, \gamma_0 = 6.86 \cdot 10^{-4}, \gamma_1 = 2.17 \cdot 10^{-6}$.

Figure 5. The variations of the elliptical Moffat PSF ($\lambda = 465$nm) in the first quadrant of the field of view.

Modeling the hyperspectral PSF at a particular wavelength by an elliptical Moffat function results in a model involving 7 unknown parameters. In particular the parameter vector becomes $\phi = [\alpha_0 \ \alpha_1 \ \alpha_2 \ \alpha_3 \ \beta \ \gamma_0 \ \gamma_1]$. 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. 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:

\[
b_{\lambda_i} = h_{\lambda_i} s_{\lambda_i} + \eta_{\lambda_i}, \quad i = 1, 2, \ldots, N_w,
\]

where \( b_{\lambda_i} \in \mathbb{R}^{N_p} \) is the observed star image at wavelength \( \lambda_i \), \( h_{\lambda_i} \in \mathbb{R}^{N_p} \) is the noise-free star image, or PSF, at wavelength \( \lambda_i \), \( \eta_{\lambda_i} \in \mathbb{R}^{N_p} \) represents noise corresponding to a star image at wavelength \( \lambda_i \), and \( s_{\lambda_i} \) is a scalar representing the reflectance of the star at a particular wavelength \( \lambda_i \). Since there is a one-to-one correspondence between \( \lambda_i \) and the index \( i \), without loss of generality, we will use the notation:

\[
b_i = b_{\lambda_i}, \quad h_i = h_{\lambda_i}, \quad s_i = s_{\lambda_i}, \quad \eta_i = \eta_{\lambda_i}.
\]

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

\[
b = H(\phi)s + \eta,
\]

where

\[
s = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_{N_w} \end{bmatrix},
\]

and in the case of a circular Moffat PSF,

\[
b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{N_w} \end{bmatrix}, \quad H(\phi) = \begin{bmatrix} h_1 & 0 & \cdots & 0 \\ 0 & h_2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & h_{N_w} \end{bmatrix}, \quad \phi = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \beta \end{bmatrix}.
\]

In the case of elliptical Moffat PSF, with \( N_w \) wavelengths and \( N_o \) orientations (i.e., \( N_o \) polar coordinates \( (\rho_\ell, \theta_\ell) \), \( \ell = 1, 2, \ldots, N_o \)) in the field of view, we have

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

In [39] the authors have proposed an estimation scheme consisting of two steps. In the first step the star spectrum, \( s \), and the noise, \( \eta \), 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 \( s \) and \( \eta \) 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:

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

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 tuning 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 [14, 20, 32] is applied to the objective function, $f(\phi, s)$, 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 $s$ and optimizing only over the nonlinear term $\phi$.

### 3.1. Variable Projection

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

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

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

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

whose minimum is attained at

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

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

$$\tilde{f}(\phi) = \| b - H(\phi) H(\phi)^\dagger b \|_2^2 = \| (I - H(\phi) H(\phi)^\dagger) b \|_2^2. \tag{4}$$

For notational convenience, we drop “$(\phi)$” in the following equations. Note that,

$$H^T H = \begin{bmatrix} h_1^T \\ h_2^T \\ \vdots \\ h_N^T \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_N \end{bmatrix} = \begin{bmatrix} \| h_1 \|_2^2 \\ \| h_2 \|_2^2 \\ \vdots \\ \| h_N \|_2^2 \end{bmatrix}.$$  

Thus,

$$HH^\dagger = H \left( H^T H \right)^{-1} H^T = \begin{bmatrix} h_1 h_1^T \| h_1 \|_2^2 \\ h_2 h_2^T \| h_2 \|_2^2 \\ \vdots \\ h_N h_N^T \| h_N \|_2^2 \end{bmatrix},$$

and

$$P = I - HH^\dagger = \begin{bmatrix} I - h_1 h_1^T \| h_1 \|_2^2 \\ I - h_2 h_2^T \| h_2 \|_2^2 \\ \vdots \\ I - h_N h_N^T \| h_N \|_2^2 \end{bmatrix}.$$
where $P$ is the projector onto the orthogonal complement of the range of $H$. With this notation the minimization problem in (4) can be reposed as a nonlinear least squares problem:

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

We use the Gauss-Newton algorithm [21, 26, 28] to solve this problem.

### Algorithm: Gauss-Newton

| given: $b$ |
| choose: initial guess for $\phi$ |
| step length parameter $\tau$ |
| for $i = 1, 2, \ldots$, until stop |
| solve |
| $J(\phi)^T J(\phi) d = -J(\phi)^T r$ |
| where |
| $r = b - H(\phi)s = P(\phi)b$ |
| $J = \nabla (P(\phi)b) = \nabla P(\phi)b$ |
| Set $\phi = \phi + \tau d$, where $\tau$ is chosen using line search. |
| end for |
| Set $s = H(\phi)^\dagger b$. |

The behavior of the Gauss-Newton method depends on the conditioning of the Jacobian matrix, e.g. [26].

### 3.2. Jacobian Matrix

In this section we consider the structure and conditioning of the Jacobian matrix. For separable nonlinear least squares problems, Golub and Pereyra [14] show that it is convenient to calculate individual columns of the Jacobian, using derivatives of $H(\phi)$. Specifically (we use similar notation as O’Leary and Rust [27]), for each variable $\phi_k$, define derivative matrices $D_k$ to have $(i,j)$ entries

$$[D_k]_{i,j} = \left[ \frac{\partial H_j}{\partial \phi_k} \right]_i,$$

where $H_j$ is the $j$th column of $H(\phi)$. With this definition, the Jacobian can be written as

$$J = -(C + F),$$

where the $k$th column of $C$ is

$$c_k = PD_k H^\dagger b = PD_k s,$$

and the $k$th column of $F$ is

$$f_k = (PD_k H^\dagger)^T b = (H^\dagger)^T D_k^T P^T b = (H^\dagger)^T D_k^T r. \quad (7)$$

Because it is the more important case, in the rest of this section we consider only the elliptical Moffat PSF. In this case, the $i$th entry of the PSF corresponding to wavelength $\lambda_j$ and $\ell$th orientation $(\rho_\ell, \theta_\ell)$ is

$$[h_j^{(\ell)}]_i = \left[ 1 + \frac{i^2 + (j^2 + \gamma_0 + \gamma_1 \lambda_j + \alpha_2 \lambda_j + \alpha_3 \beta \lambda_j)^2}{(\alpha_0 + \alpha_1 \rho_\ell + \alpha_2 \lambda_j + \alpha_3 \lambda_j^2)^2} \right]^{-\beta}.$$ 

It is only necessary to compute derivatives with respect to $\alpha_0$, $\beta$, and $\gamma_0$. 

By identifying variables $\phi_k$ as

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

the above derivatives define the entries of matrices $D_1$, $D_5$ and $D_6$. Now observe that the remaining derivatives satisfy the relations

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

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

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

$$\frac{\partial \left[ h_j^{(\ell)} \right]_i}{\partial \gamma_0} = \lambda_j \frac{\partial \left[ h_j^{(\ell)} \right]_i}{\partial \gamma_0},$$

from which it follows that

$$D_2 = RD_1,$$

$$D_3 = D_1 \Lambda,$$

$$D_4 = D_1 \Lambda^2,$$

$$D_7 = D_6 \Lambda,$$

where

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

These relations for $D_k$ can be used, according to equations (5-7), 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 $\Lambda$ is a diagonal matrix, we can write

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

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

$$\begin{bmatrix} c_1 & c_3 & c_4 \end{bmatrix} = PD_1 \text{diag}(s) \begin{bmatrix} 1 & \lambda & \lambda^2 \end{bmatrix},$$
where $\lambda^2 = \begin{bmatrix} \lambda_1^2 & \cdots & \lambda_{N_w}^2 \end{bmatrix}^T$. A similar observation can be made for the same subset of columns of $F$, namely

$$\begin{bmatrix} f_1 & f_3 & f_4 \end{bmatrix} = (H^\dagger)^T \text{diag}(D_{T}^T r) \begin{bmatrix} 1 & \lambda & \lambda^2 \end{bmatrix}.$$  

The Vandermonde structure of these corresponding columns of $J$ indicate the potential of ill-conditioning; our numerical experiments (see Section 4) 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.

We remark that the main difference in computational costs and storage requirements between the two models depends mainly on the size of the Jacobian, and the associated costs of doing computations with it. Let $J(\phi) \in \mathbb{R}^{M \times p}$, where $p$ is the number of parameters used to define the hyperspectral PSF. Specifically, for the circular Moffat model $M = N_w \cdot N_p$, whereas in the elliptical Moffat model $M = N_o \cdot N_w \cdot N_p$. Here, $N_w$ is the number of wavelengths, $N_p$ is the number of pixels, and $N_o$ is the number of orientations. Furthermore, in the circular Moffat model $p = 3$ and in the elliptical Moffat model $p = 7$. A standard implementation to solve the Jacobian system will cost $O(M p^2)$ floating point operations, and thus the elliptical Moffat model is more expensive than the circular model. Iterative approaches and/or parallel implementations might be able to reduce some of the extra overhead needed with the elliptical model, but we emphasize that if the elliptical model is the correct model, then one should use it instead of the circular model, even if it is more expensive.

### 3.3. Subset Selection

Subset selection is a method for selecting the most linearly independent columns of a matrix in such a way that the selected subset of columns is a good representation of the original matrix and it is not rank deficient [5, 13, 15, 16, 38]. It has been shown in [19] 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.

Assuming $p$ is the number of unknowns in the model, let $1 \leq k \leq p$ be a sampling parameter. We use subset selection to choose $k$ linearly independent columns of the Jacobian matrix $J(\phi)$. Subset selection applied to the Jacobian matrix $J(\phi)$ produces a permutation matrix $\Pi$ such that

$$J(\phi)\Pi = [J_1 \quad J_2],$$

where $J_1$ contains $k$ columns. The goal is to bring the $k$ linearly independent columns of $J(\phi)$ to the front. This results in identifying the “best” $k$ parameters of the original parameter set $\phi$ while keeping the rest of the $p - k$ parameters unchanged. Thus we solve the new Jacobian system

$$J_1^T J_1 d = -J_1^T r.$$

This guarantees that the nonlinear least squares problem we are trying to solve has a full rank Jacobian. A rank revealing QR factorization (RRQR) can be used to perform subset selection. The goal of an RRQR algorithm is to find a permutation matrix $\Pi$ such that if

$$J(\phi)\Pi = QR = Q \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{bmatrix},$$

is a QR factorization of $J(\phi)\Pi$ then $R_{11} \in \mathbb{R}^{k \times k}$ has a condition number in the order of $\frac{\sigma_{k+1}}{\sigma_k}$ ($\sigma_i$ are the singular values of $J(\phi)$ for $i = 1, \cdots, p$), $||R_{22}||_2$ is on the order of $\sigma_{k+1}$, and $k$ is the numerical rank of $J(\phi)$. For a detailed discussion on different types of rank-revealing factorizations and comparisons of rank-revealing algorithms see [5, 10]. We perform subset selection using an implementation of the Chan-Foster RRQR algorithm [2, 4, 9] from UTV Tools [8].
We remark that an alternative to using the RRQR factorization is to compute a regularized solution of the Jacobian system. For example, Tikhonov regularization solves
\[
(J^T J + \mu^2 I) d = J^T r,
\]
where the regularization parameter \( \mu \) satisfies \( \sigma_p \leq \mu \leq \sigma_1 \), where \( \sigma_1 \) and \( \sigma_p \) are, respectively, the largest and smallest singular values of \( J \). Another common regularization scheme is the truncated singular value decomposition (TSVD). Specifically, suppose the SVD of \( J \) is given by
\[
J = U \Sigma V^T,
\]
where \( U = [u_1 \cdots u_M] \in \mathbb{R}^{M \times M} \) and \( V = [v_1 \cdots v_p] \in \mathbb{R}^{p \times p} \) are orthogonal matrices, and \( \Sigma = \text{diag}(\sigma_1, \ldots, \sigma_p) \), with \( \sigma_1 \geq \cdots \geq \sigma_p \). Then a TSVD solution is given by
\[
d = \sum_{i=1}^{k} \frac{u_i^T r}{\sigma_i} v_i,
\]
where the truncation (or regularization) index \( k \) satisfies \( k \leq p \).

4. NUMERICAL 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 \( \lambda = 465 \text{nm} \) with different levels of noise is shown in Figure 6. In this and subsequent figures of PSFs, the square roots of the PSFs are shown for better contrast.

![Image of an isolated star observed at wavelength \( \lambda = 465 \text{nm} \). The top row is the noise-free image. The bottom row shows the star image with different levels of noise.](image)

Figure 6. Image of an isolated star observed at wavelength \( \lambda = 465 \text{nm} \). The top row is the noise-free image. The bottom row shows the star image with different levels of noise.

Recall from Section 3, we are minimizing:
\[
f(\phi, s) = \| b - H(\phi)s \|_2^2.
\]
Our goal is to find the parameters defining the hyperspectral PSF (e.g. the top row of Figure 6) and to reconstruct the star spectrum (e.g. Figure 7), given the noisy measurements (e.g. the bottom row of Figure 6). The relative error at the $k^{th}$ iteration is defined as
\[
\frac{\|\phi_k - \phi_{true}\|_2}{\|\phi_{true}\|_2},
\]
for the hyperspectral PSF parameters, and
\[
\frac{\|s_k - s_{true}\|_2}{\|s_{true}\|_2},
\]
for the star spectrum.

The downward spikes in the spectrum of the star shown in Figure 7 are the so-called absorption lines which are produced when the photon energy corresponding to certain wavelengths is absorbed and then re-emitted mostly in different directions from the original ones. The strength of the lines indicate the abundance of particular atoms or ions in the star. The set of absorption lines is unique for each atom or ion and thus these lines are used to analyze the chemical composition and physical conditions of stars and other celestial bodies.

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 [33]. The simulated noisy hyperspectral PSFs are of size $64 \times 64$. We use $\phi_{true} = \begin{bmatrix} \alpha_0 & \alpha_1 & \beta \end{bmatrix}^T = \begin{bmatrix} 2.42 & -0.001 & 2.66 \end{bmatrix}^T$ to generate the data and $\phi_0 = \begin{bmatrix} 4.61 & -0.0009 & 4.3 \end{bmatrix}^T$ as an initial guess.
there is no clear gap between the second and third singular values. We plot the objective function (Figure 9), the relative errors for the PSF parameters (Figure 10), and the approximated star spectrum for the first 100 wavelengths (Figure 11) 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 \( \phi_{\text{GN}} = \begin{bmatrix} 2.4155 & -0.001 & 2.6533 \end{bmatrix}^T \) corresponding to a relative error of 0.0022. The estimated star spectrum is a very close approximation of the true star spectrum (Figure 11). Notice that even though we used 466 wavelengths, we are showing the star spectrum approximation results for only the first 100 wavelengths in order that the visual difference between the true spectrum and the approximated spectrum in the figure is more clear.

For the numerical experiments with the elliptical Moffat model we generate the data using
\[
\phi_{\text{true}} = \begin{bmatrix} 3.75 & -2.99 \cdot 10^{-3} & -4.31 \cdot 10^{-3} & 1.98 \cdot 10^{-6} & 1.74 & 6.86 \cdot 10^{-4} & 2.17 \cdot 10^{-6} \end{bmatrix}^T,
\]
as the exact parameter vector.
PSF ESTIMATION FOR HYPERSPECTRAL IMAGING

Thus we seek to find the parameters $\phi = [\alpha_0 \; \alpha'_1 \; \alpha'_2 \; \alpha'_3 \; \beta \; \gamma'_0 \; \gamma'_1]^T$ which best approximate the data. In this example the vector of true parameters becomes

$$\phi_{true} = [3.75 \; -2.99 \; -4.31 \; 1.98 \; 1.74 \; 6.86 \; 2.17]^T,$$

and we use $\phi_0 = [6.41 \; -1.07 \; -3.15 \; 3.28 \; 2.42 \; 13.18 \; 3.81]^T$ as an initial guess. We use 20 wavelengths with 25 orientations in the first quadrant of the field of view for each wavelength, which corresponds to using 500 PSFs. Since our purpose here is to show the variation of the PSFs in the field of view, we reduce the number of wavelengths to accommodate for the additional orientations in the elliptical Moffat experiments. However, we remark that the number of PSFs could be increased depending on the available computational resources and/or a parallel implementation of our method.

Figure 12 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 ($\text{cond}(J) = 10^5$). Thus the exact Gauss-Newton algorithm has the potential to compute an incorrect step direction vector when solving the system $J^T J d = -J^T r$. To improve the step direction, as previously discussed, we attempt to use subset selection, truncated SVD (TSVD), and Tikhonov regularization.

For the computation using subset election we use $k = 6$ most linearly independent columns of the Jacobian matrix (note that in this example $J$ has 7 columns). Subset selection picks $\alpha_3$ to be the problematic parameter and thus we fix $\alpha_3$ 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. 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 13. One can observe that the Tikhonov regularization method achieves the lowest relative errors. With Tikhonov regularization the approximated parameters are:

$$\phi_{TIK} = [3.8173 \; -2.9962 \; -4.5704 \; 2.2541 \; 1.7446 \; 7.0326 \; 2.1308]^T.$$

If we assume that we know the exact value for $\alpha_3$ and apply subset selection (SST in the right part of Figure 13) we achieve lower relative errors compared to fixing $\alpha_3$ to its initial guess value. We also tried a combined approach by applying subset selection for the first 7 iterations (fixing $\alpha_3$ to its initial guess value) 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 13. 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

\[ \phi_{100}^{SS+TIK} = \begin{bmatrix} 3.7710 & -2.9962 & -4.3749 & 2.0476 & 1.7446 & 7.0123 & 2.1351 \end{bmatrix}^T, \]

which provides a fair comparison with the result obtained using Tikhonov regularization. From Figure 13, we see that if we terminate at iteration 66, then we obtain a solution that minimizes the relative error. The computed solution in this case is

\[ \phi_{66}^{SS+TIK} = \begin{bmatrix} 3.7630 & -2.9860 & -4.3739 & 2.0541 & 1.7376 & 6.8686 & 2.1700 \end{bmatrix}^T. \]

Figure 14 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 15.

We remark that the dimension, \( k \), used for subset selection is problem dependent, and its choice is analogous to choosing regularization parameters; \( k \) should be large enough to accurately represent information in the Jacobian, but not so large as to cause sensitivity in the computations. The left plot in Figure 16 illustrates that for the elliptical Moffat model, \( k = 6 \) provides the best results. Finally, we remark that the noise level in these applications can be large, and rather complicated; see [3] for more details. The right plot in Figure 16 illustrates that our approach is robust with increasing noise levels.

5. CONCLUDING REMARKS

We have described an iterative approach for solving nonlinear least squares problems related to hyperspectral imaging. Circular and elliptical Moffat function models were used to define the spectrally and spatially varying hyperspectral 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.
Figure 13. 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 ($\alpha_3$) and assume that we know the true value of that parameter for the rest of the iterations (SST).

Figure 14. Iteration history of the objective function and the norm of gradient for the combined approach of subset selection and Tikhonov regularization.

which in turn causes stagnation in the convergence history of relative errors for PSF parameters 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 hyperspectral 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. The proposed approach was shown to successfully estimate jointly the PSF parameters and the star spectrum.

We remark that it could be interesting to study the HSI unmixing and deblurring methods, as in [41], while also using the system and wavelength dependent hyperspectral imaging PSFs obtained in our current work. Concerning the effectiveness of our PSF estimation methods applied to the hyperspectral image deblurring problem, we point to the paper by Zhao, et al., [42] in which the authors study the image restoration problem where the point spread function is corrupted by errors. In their model, they study the objective function by minimizing two variables: the restored image and the estimated error of the point spread function. Since our estimated hyperspectral image PSFs
Figure 15. 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.

Figure 16. Comparisons of the parameter relative errors (using subset selection and Tikhonov regularization) for the elliptical Moffat model with different $k$ values (left), and increasing noise levels with $k = 6$ (right).

In the above plots, the horizontal axis is the iteration index, and the vertical axis is relative error.

are based on HSI data from a guide star point source, they can be inaccurate due to noise and blur. The use of the total variation structured total least squares method, such as in [42], can be important in future work on joint HSI deblurring and sparse unmixing since the HSI PSFs can be corrupted by errors.

REFERENCES


