

Tensor Methods for Hyperspectral Data Analysis: A Space Object Material Identification Study

Qiang Zhang,^{1,*} Han Wang,² Robert J. Plemmons^{2,3} and V. Paul Pauca³

¹*Department of Biostatistical Sciences, Wake Forest University Health Sciences,
Medical Center Boulevard, Winston-Salem, NC 27109, USA*

²*Department of Mathematics, Wake Forest University,
Winston-Salem, NC 27109, USA*

³*Department of Computer Science, Wake Forest University,
Winston-Salem, NC 27109, USA*

**Corresponding author: qizhang@wfubmc.edu*

An important and well studied problem in hyperspectral image data applications is to identify materials present in the object or scene being imaged and to quantify their abundance in the mixture. Due to the increasing quantity of data usually encountered in hyperspectral datasets, effective data compression is also an important consideration. In this paper, we develop novel methods based on tensor analysis that focus on all three of these goals: material identification, material abundance estimation, and data compression. Test results are reported in all three perspectives. © 2008 Optical Society of America

Key Words: hyperspectral imaging, tensor analysis, nonnegativity constraints, space object material identification

OCIS codes: 000.3860, 100.3008, 100.6890, 110.4234.

1. Introduction

Hyperspectral remote sensing technology allows one to capture images using a range of spectra from ultraviolet to visible to infrared. Multiple images of a scene or object are created using light from different parts of the spectrum. These hyperspectral images can be used, for example, to detect and identify objects at a distance, to identify surface minerals, objects and buildings from space, and to enable Space Object Identification (SOI) from the ground. In this particular study within the domain of SOI, we concentrate on the material identification only.

Three major objectives in processing hyperspectral image data of an object (target) are data compression, spectral signature identification of constituent materials, and determination of their corresponding fractional abundances. Here we propose a novel approach to processing hyperspectral data using Nonnegative Tensor Factorization (NTF), which reduces a large tensor into three nonnegative factor matrices, the Khatri-Rao product which approximates the original tensor, see e.g. [1–6]. This approach preserves physical characteristics of the data such as nonnegativity and is a natural extension of nonnegative least squares approximate nonnegative matrix factorization, see e.g. [7].

In (approximate) Nonnegative Matrix Factorization (NMF), an $m \times n$ (nonnegative) mixed data matrix X is approximately factored into a product of two nonnegative rank- k matrices, with k small compared to m and n , $X \approx WH$. This factorization has the advantage that W and H can provide a physically realizable representation of the mixed data. Since the early work of Paatero and Tapper [8] and Lee and Seung’s seminal paper on learning the parts of objects [9], NMF algorithms have been developed and applied in numerous areas of engineering, science, and medicine. In particular, NMF is widely used in a variety of applications, including air emission control, image and spectral data processing, text mining, chemometric analysis, neural learning processes, sound recognition, remote sensing, spectral unmixing and object characterization, see, e.g. [3, 10].

Nonnegative Tensor Factorization (NTF) is a natural extension of NMF to higher

dimensional data. In NTF, high-dimensional data, such as hyperspectral or other image cubes, is factored directly and is approximated by a sum of rank-1 nonnegative tensors. See Figure 5 for an illustration of 3-D tensor factorization. The ubiquitous tensor approach, originally suggested by Einstein to explain laws of physics without depending on inertial frames of reference, is now becoming the focus of extensive research, e.g. [3]. Here, we develop and apply NTF algorithms for the analysis of spectral and hyperspectral image data. The method described in this paper combines features from both NMF and NTF methods.

For safety and other considerations in space, non-resolved space object characterization is an important component of Space Situational Awareness (SSA). The key problem in non-resolved space object characterization is to use spectral reflectance data to gain knowledge regarding the physical properties (e.g., function, size, type, status change) of space objects that cannot be spatially resolved with normal panchromatic telescope technology. Such objects may include geosynchronous satellites, rocket bodies, platforms, space debris, or nano-satellites. Spectral reflectance data of a space object can be gathered using ground-based spectrometers, such as the SPICA system, see [11–13], located on the 1.6 meter Gemini telescope and the ASIS system, see [14–16], located on the 3.67 meter telescope at the Maui Space Surveillance Complex (MSSC), and contains essential information regarding the makeup or types of materials comprising the object. Different materials, such as aluminum, mylar, paint, plastics and solar cell, possess fairly unique characteristic wavelength-dependent ab-

sorption features, or spectral *signatures*, that mix together in the spectral reflectance measurement of an object.

Spectral unmixing is a problem that originated within the hyperspectral imaging community and several computational methods to solve it have been proposed over the last few years. A thorough study and comparison of various computational methods for endmember or spectral signature computation, in the related context of hyperspectral unmixing, can be found in the work of Plaza *et al.*, [17]. An information-theoretic approach has been provided by Chang [18].

In an earlier project on spectral data analysis for SOI, some of the authors have employed Non-negative Matrix Factorization (NMF) algorithms for unmixing of spectral reflectance data from a single pixel imaged by the SPICA spectrometer at MCSS to find endmember candidates. In that work, regularized inverse problem methods for determining corresponding fractional abundances were developed [12, 13].

A new spectral imaging sensor, capable of collecting hyperspectral images of space objects, has been installed on the 3.67 meter Advanced Electro-Optical System (AEOS) at the MSSC. The AEOS Spectral Imaging Sensor (ASIS) is used to collect adaptive optics compensated spectral images of astronomical objects and satellites. In a series of papers, Blake *et al.*, [14–16], have developed model-based spectral image deconvolution methods that can simultaneously remove much of the spatial and spectral blurring introduced by the ASIS sensor. See Figure 5 for a simulated hyperspectral image of the Hubble Space Telescope from [14], similar to that collected by

ASIS.

Other methods and algorithms have been developed for some of the three main objectives for processing hyperspectral data. For compressing the hyperspectral data, while maintaining the endmembers' presence in the original data, a set of techniques under the umbrella of dimensionality reduction have been developed, including the traditional Principal Component Analysis (PCA), Independent Component Analysis (ICA) [18], Wavelet [19] and vector quantization [20] methods. For a single material identification purpose, a group of algorithms under the name of Endmember Extraction Algorithms (EEA) has been used to match endmembers with a material library, e.g. Pixel Purity Index (PPI) [21], N-finder (N-FINDR) algorithm [22], Iterative Error Analysis (IEA) [23] and Automated Morphological Endmember Extraction (AMEE) algorithm [17]. However, in remote sensing image analysis the difficulty arises in the fact that information captured by a detector pixel is mixed linearly or nonlinearly by different materials resident in the physical area of the scene associated with the pixel. Here direct application of approaches described above may not work well [21, 24].

One popular approach to linearly unmix spectral signatures is to solve a regular least square problem to fit for the fractional abundances by minimizing the l_2 norm difference between the observed vector and the material signature matrix times the fractional abundance vector. Due to the nature of fractional abundances, two constraints are usually desired, i.e. non-negativity and sum-to-one [25, 26]. A linear matched filter approach [21] computes fast and reveals those prevalent signatures by

suppressing the background. It can also satisfy the sum-to-one constraint, but not necessarily the non-negativity constraint [27]. Keshava [24] provides a nice comparison and classification of various spectral unmixing algorithms.

NMF algorithms have been developed and applied to various applications, see e.g. [3, 10]. In [12], additional constraints are explored to better recover material spectra. Two sets of factors, one as endmembers and the other as fractional abundances, are optimally chosen to balance fit to data and smoothness of endmembers and sparsity of fractional abundances. And, due to reduced size of factors all three purposes, including data compression, can be fulfilled at the same time. When compared with linear unmixing methods, NMF inherently imposes a non-negativity constraint and the sum-to-one constraint can also be posed by adding an extra term in the cost function, see [12]. Du et al. [28] proposed a similar compression method to save only the material signatures and their fractional abundances within each pixel, which is exactly the two factor matrices of NMF, and they also imposed the non-negative and sum-to-one constraints.

For 3-D hyperspectral data, Shashua and Levin [6] have shown better compression and preservation of components can be achieved with NTF than with NMF [5]. The purpose of our paper is to extend work by Pauca et al. [12, 13] on the use of NMF for space object material identification, where a single image pixel of the object collected by a spectrometer was used. Here we consider more recent hyperspectral image data of the type collected by the AEOS ASIS system on Maui. Work related in various

ways to ours can be found, e.g., in papers by Blake et al. [14–16], Hege et al. [29], and Scholl et al. [30].

In Section 2, we define the NTF problem and present a block optimization approach to divide the NTF problem into three NMF sub-problems, the solutions of which are sought through an improved projected gradient method. In Section 3, we test our method using four simulated datasets of the Hubble Space Telescope, considering the presence of both noise and blurring in hyperspectral images, and we present test results, including a comparison with a linear unmixing method with constraints. We conclude with a brief discussion and ideas for future research.

2. Tensor Methods for Spectral Unmixing

Next we introduce notation commonly used within tensor analysis literature, followed by the core NTF problem and its solution.

2.A. Notational Preliminaries

We call a three-way data array $\mathcal{T} = (t_{ijk})$ a tensor, where $i = 1, \dots, D_1$, $j = 1, \dots, D_2$ and $k = 1, \dots, D_3$. For fixed i and j , we call the vector $\mathbf{t}_{ij} = (t_{ij1}, \dots, t_{ijD_3})^T$ a fiber.

In our hyperspectral data application, the third dimension represents wavelength. The symbol \circ denotes the usual outer product of two vectors, $\mathbf{x} \circ \mathbf{y} = \mathbf{x}\mathbf{y}^T$. A three-way outer product of three vectors $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^m$, and $\mathbf{z} \in \mathbb{R}^\ell$, leads to a three-way

tensor of rank one,

$$\mathcal{T} = \mathbf{x} \circ \mathbf{y} \circ \mathbf{z} \in \mathbb{R}^{n \times m \times \ell},$$

where $t_{ijk} = x_i y_j z_k$, for $i = 1, \dots, n$, $j = 1, \dots, m$, and $k = 1, \dots, \ell$. Moreover, a tensor $\mathcal{T} \in \mathbb{R}^{n \times m \times \ell}$ that can be written as a finite sum of rank-one tensors,

$$\mathcal{T} = \sum_{i=1}^k \mathbf{x}^{(i)} \mathbf{y}^{(i)} \mathbf{z}^{(i)},$$

where $\mathbf{x}^{(i)} \in \mathbb{R}^n$, $\mathbf{y}^{(i)} \in \mathbb{R}^m$ and $\mathbf{z}^{(i)} \in \mathbb{R}^\ell$ is said to be in CANDECOMP (CP) canonical factored form.

The symbol \otimes denotes the Kronecker product, which for two vectors \mathbf{x} and \mathbf{y} is given by,

$$\mathbf{x} \otimes \mathbf{y} = (x_1 \mathbf{y}, x_2 \mathbf{y}, \dots, x_n \mathbf{y}).$$

The symbol \odot denotes the Khatri-Rao product, which for two matrices, $\mathbf{X} \in \mathbb{R}^{m \times k}$, $\mathbf{Y} \in \mathbb{R}^{n \times k}$, having the same number of columns is given by,

$$\mathbf{X} \odot \mathbf{Y} = (\mathbf{x}_1 \otimes \mathbf{y}_1, \dots, \mathbf{x}_n \otimes \mathbf{y}_n) \in \mathbb{R}^{mn \times k},$$

where \mathbf{x}_i and \mathbf{y}_i denote columns of \mathbf{X} and \mathbf{Y} respectively.

Tensor unfolding is the process of rearranging tensor entries into a matrix, similar to unfolding a data cube to a flat matrix. For a three-way tensor, $\mathcal{T} \in \mathbb{R}^{D_1 \times D_2 \times D_3}$,

there are three typical ways of unfolding \mathcal{T} along each dimension, i.e.

$$(\mathbf{T}_1)_{(k-1)*D_3+j,i} = t_{ijk}, \quad \mathbf{T}_1 \in \mathbb{R}^{D_2 D_3 \times D_1},$$

$$(\mathbf{T}_2)_{(k-1)*D_3+i,j} = t_{ijk}, \quad \mathbf{T}_2 \in \mathbb{R}^{D_1 D_3 \times D_2},$$

$$(\mathbf{T}_3)_{(j-1)*D_2+i,k} = t_{ijk}, \quad \mathbf{T}_3 \in \mathbb{R}^{D_1 D_2 \times D_3},$$

where $i = 1, \dots, D_1$, $j = 1, \dots, D_2$ and $k = 1, \dots, D_3$. The Frobenius norm of a tensor is defined as the square root of the sum of squares of all its entries, i.e.

$$\|\mathcal{T}\|_F = \sqrt{\sum_i \sum_j \sum_k t_{ijk}^2}.$$

Definition 1. Let $\mathcal{T} \in \mathbb{R}^{D_1 \times D_2 \times D_3}$ be a nonnegative tensor and $\hat{\mathcal{T}} = \sum_{i=1}^k \mathbf{x}^{(i)} \circ \mathbf{y}^{(i)} \circ \mathbf{z}^{(i)}$ a tensor in CP factored form, where $\mathbf{x}^{(i)} \in \mathbb{R}^{D_1}$, $\mathbf{y}^{(i)} \in \mathbb{R}^{D_2}$, $\mathbf{z}^{(i)} \in \mathbb{R}^{D_3}$. Then a rank- k nonnegative approximate tensor factorization problem is defined as:

$$\min_{\tilde{\mathcal{T}}} \|\mathcal{T} - \tilde{\mathcal{T}}\|_F^2, \quad \text{subject to } \tilde{\mathcal{T}} \geq 0. \quad (1)$$

The factor matrices associated with the CP tensor $\tilde{\mathcal{T}}$ can then be expressed as

$$\mathbf{X} = [\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k)}] \in \mathbb{R}^{D_1 \times k}, \quad \mathbf{Y} = [\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(k)}] \in \mathbb{R}^{D_2 \times k}, \quad \mathbf{Z} = [\mathbf{z}^{(1)}, \dots, \mathbf{z}^{(k)}] \in \mathbb{R}^{D_3 \times k},$$

leading to the following representation of (1) in terms of unfolded matrices.

Remark 1. *The norm of the residual defined in Definition 1 can be equivalently written in the Frobenius norm of unfolded matrices, i.e.*

$$\|\mathbf{T} - \tilde{\mathbf{T}}\|_F^2 = \|\mathbf{T}_1 - (\mathbf{Z} \odot \mathbf{Y})\mathbf{X}^T\|_F^2 = \|\mathbf{T}_2 - (\mathbf{Z} \odot \mathbf{X})\mathbf{Y}^T\|_F^2 = \|\mathbf{T}_3 - (\mathbf{X} \odot \mathbf{Y})\mathbf{Z}^T\|_F^2.$$

2.B. Alternating Least Squares Method

A common approach in solving the optimization problem in Definition (1), the Alternating Least Squares (ALS) method [2, 4, 31, 32], is a special case of the block coordinate descent method, also known as the Block Gauss-Seidel (BGS) method [33]. At each iteration step, the BGS method alternately optimizes only a subset of the variables, while keeping the rest fixed, and turns the original non-convex problem into a sequence of convex least squares sub-problems. In NTF, this means holding two matrix factors fixed while fitting for the other one. Thus the original NTF problem is transformed into three semi-NMF sub-problems in each iteration. Here we use the term “semi” to represent the optimization only for one of the two factor matrices, while assuming the other is given.

Definition 2. *Given $\mathbf{A} \in \mathbb{R}^{m \times n} \geq 0$ and $\mathbf{W} \in \mathbb{R}^{m \times k} \geq 0$, a semi-NMF problem is defined as,*

$$\min_{\mathbf{H}} \Phi(\mathbf{H}) = \|\mathbf{A} - \mathbf{W}\mathbf{H}\|_F^2, \text{ subject to } \mathbf{H} \geq 0. \quad (2)$$

The solution of (2) can also be specified as a function Ψ of \mathbf{A} and \mathbf{W} ,

$$\Psi(\mathbf{A}, \mathbf{W}) = \min_{\mathbf{H}} \|\mathbf{A} - \mathbf{W}\mathbf{H}\|_F^2, \text{ subject to } \mathbf{H} \geq 0. \quad (3)$$

A semi-NMF problem for \mathbf{W} given \mathbf{A} and \mathbf{H} can be similarly defined, leading to a 2-block BGS method. NTF can be correspondingly transformed into a 3-block BGS method. Grippo and Sciandrone [33] proved for BGS methods that if the objective function is componentwise strictly quasi-convex for $b - 2$ blocks, where b is the total number of blocks, and the sequence generated by the BGS methods has limit points, then every limit point is a critical point also. For an unfolded tensor $\tilde{\mathcal{T}}$ with factor matrices \mathbf{X} , \mathbf{Y} , and \mathbf{Z} , the functions $\Phi(\mathbf{X})$, $\Phi(\mathbf{Y})$, and $\Phi(\mathbf{Z})$ are convex, and the remaining issue is to ensure their sequences, e.g. $\{\mathbf{X}^p\}$, have at least one limit point, which often comes from the boundedness of the feasible region. The nonnegative constraint provides a lower bound, i.e. a zero matrix, and thus we will need an upper bound also, which can be easily added.

Now we can redefine the semi-NMF problem in Definition 2 by adding an upper bound and gain confidence about the convergence of our ALS method to solve the NTF algorithm previously described. Much of the success of applying tensor analysis is attributed to the Kruskal essential uniqueness, see [34], of the NTF problem in Definition 1. In practice, we may not need the upper bound, and so far we have not observed results moving up to infinity, possibly because the line search used by

the Projected Gradient Descent method might be essentially bounded above. Some further analysis in this perspective might be helpful.

Definition 3. We define a bounded semi-NMF problem as,

$$\min_{\mathbf{H}} \Phi(\mathbf{H}) = \|\mathbf{A} - \mathbf{W}\mathbf{H}\|_F^2, \text{ subject to } 0 \leq \mathbf{H} \leq \mathbf{U}, \quad (4)$$

where $\mathbf{A} \in \mathbb{R}^{m \times n} \geq 0$ and $\mathbf{W} \in \mathbb{R}^{m \times k} \geq 0$ are given. The solution of (4) can also be defined as a function Ψ_b of \mathbf{A} and \mathbf{W} ,

$$\Psi_b(\mathbf{A}, \mathbf{W}) = \min_{\mathbf{H}} \|\mathbf{A} - \mathbf{W}\mathbf{H}\|_F^2. \quad (5)$$

We summarize the discussion above in Algorithm 1, providing a method for solving the approximate nonnegative tensor factorization problem of Definition 1. Next, we

Algorithm 1: Alternating Least Squares Algorithm for NTF

Input: $\mathcal{T} \in \mathbb{R}^{D_1 \times D_2 \times D_3}$, k
Output: $\mathbf{X} \in \mathbb{R}^{D_1 \times k}$, $\mathbf{Y} \in \mathbb{R}^{D_2 \times k}$, $\mathbf{Z} \in \mathbb{R}^{D_3 \times k}$
Initialize randomly $\mathbf{X}^{(0)}$, $\mathbf{Y}^{(0)}$;
 $n \leftarrow 0$;
repeat
 $\mathbf{Z}^{(n+1)} \leftarrow \Psi_b(\mathcal{T}_3, \mathbf{X}^{(n)} \odot \mathbf{Y}^{(n)})^T$;
 $\mathbf{Y}^{(n+1)} \leftarrow \Psi_b(\mathcal{T}_2, \mathbf{Z}^{(n+1)} \odot \mathbf{X}^{(n)})^T$;
 $\mathbf{X}^{(n+1)} \leftarrow \Psi_b(\mathcal{T}_1, \mathbf{Z}^{(n+1)} \odot \mathbf{Y}^{(n+1)})^T$;
until converged ;

describe our numerical approach for efficiently solving the three semi-NMF problems needed for computing $\mathbf{Z}^{(n)}$, $\mathbf{Y}^{(n)}$, and $\mathbf{X}^{(n)}$.

2.C. Improved Projected Gradient Descent Method

Various approaches for solving NMF have been proposed over the last few years, several of which are based in part on early work by Lawson and Hanson [35] on Nonnegative Least Squares (NNLS) computations, see e.g. [7, 10, 36–39] and references therein. The seminal work of Lee and Seung [9] provided an alternating approach for computing the nonnegative factors \mathbf{W} and \mathbf{H} of a matrix \mathbf{A} , Bro and Jong [36] improved the active/passive set method of Lawson and Hanson [35], Recently, Chu and Lin [37] improved Lee and Seung method significantly with respect to the convergence time. Here we use a modified non-alternating version of a Projected Gradient Descent (PGD) method developed by Lin [38] to solve the three semi-NMF problems of the form $\mathbf{H} \leftarrow \Psi_b(\mathbf{A}, \mathbf{W})$ in Algorithm 1.

In the PGD method, determining an appropriate step size at each iteration can be difficult. Lin [38] pointed out that Armijo’s Rule is an effective criteria to determine step sizes. The rule can be expressed as

$$\Phi(\mathbf{H}^{(p+1)}) - \Phi(\mathbf{H}^{(p)}) \leq \sigma \langle \text{vec}(\nabla \Phi(\mathbf{H}^{(p)})), \text{vec}(\mathbf{H}^{(p+1)} - \mathbf{H}^{(p)}) \rangle, \quad (6)$$

where $\mathbf{H}^{(p+1)} = P_+[\mathbf{H}^{(p)} - \alpha_p \nabla \Phi(\mathbf{H}^{(p)})]$, α_p is the step size at the p^{th} iteration and P_+ is the projection function that replaces all negative entries with zeros and all entries greater than the specified upper bound, U , with U . Armijo’s Rule ensures sufficient decrease of Φ for each iteration. And by trying $\alpha_p = 1, \beta, \beta^2, \dots$ where $\beta \in (0, 1)$, it is

proved by Bertsekas [40] that one can find a positive α_p satisfying (6).

In general, the step size search is the most time consuming part in the iterations. To address this problem, Lin suggested to transform (6) by using the quadratic and convex properties of Φ , and then show that the new form leads to a much lower computational cost.

$$(1-\sigma)\langle \text{vec}(\nabla\Phi(\mathbf{H})), \text{vec}(\tilde{\mathbf{H}}-\mathbf{H}) \rangle + \frac{1}{2}\langle \text{vec}(\tilde{\mathbf{H}}-\mathbf{H}), \text{vec}((\mathbf{W}^T\mathbf{W})(\tilde{\mathbf{H}}-\mathbf{H})) \rangle \leq 0. \quad (7)$$

Table 1 shows a comparison of the computational costs associated with (6) and (7), with respect to the number of multiplications for parts of these computations.

From the table, we conclude that, if $\mathbf{W}^T\mathbf{W}$ and $\nabla\Phi(\mathbf{H})$ are given and if $k = n$, (7) approximately reduces the complexity of (6) by a factor of 2. When $k \ll n$, which is common in real situations, the improvement in performance will be considerable. Further significant reduction in computational cost can be obtained using α_{p-1} as the initial guess for α_p , as suggested by Lin [38]. We summarize our improved Projected Gradient Method in Algorithm 2.

Table 2 illustrates the reduction in computational cost obtained by the improvements made to PGD, with matrices of increasing size. For each matrix size, we set $k = 50$ and $maxiter = 30$. Test 0 shows the performance associated with the original PGD method. Test 1 shows the performance of PGD with (6) replace by (7). Test 2 shows the performance of PGD with α_{p-1} as initial guess for α_p . Test 3 shows

Algorithm 2: An Improved Projected Gradient Method

Input: $\mathbf{A} \in \mathbb{R}^{m \times n} \geq 0$, $\mathbf{W} \in \mathbb{R}^{m \times k} \geq 0$
Output: $\mathbf{H} \in \mathbb{R}^{k \times m} \geq 0$
Set $\beta = 0.1$, $\sigma = 0.01$, $\alpha_0 = 1$, and $p = 1$;
Initialize randomly $\mathbf{H}^1 \in \mathbb{R}^{k \times m} \geq 0$;
while $p < \text{maxiter}$ **do**
 $\alpha_p \leftarrow \alpha_{p-1}$;
 $\tilde{\mathbf{H}} \leftarrow P_+[\mathbf{H}^p - \alpha_p \nabla \Phi(\mathbf{H}^p)]$;
 if α_p satisfies (7) **then**
 while α_p satisfies (7) **do**
 $\alpha_p \leftarrow \alpha_p / \beta$;
 $\tilde{\mathbf{H}} \leftarrow P_+[\mathbf{H}^p - \alpha_p \nabla \Phi(\mathbf{H}^p)]$;
 end
 else
 while α_p does not satisfy (7) **do**
 $\alpha_p \leftarrow \alpha_p * \beta$;
 $\tilde{\mathbf{H}} \leftarrow P_+[\mathbf{H}^p - \alpha_p \nabla \Phi(\mathbf{H}^p)]$;
 end
 end
 $\mathbf{H}^{p+1} \leftarrow P_+[\mathbf{H}^p - \alpha_p \nabla \Phi(\mathbf{H}^p)]$;
 $p \leftarrow p + 1$;
end

the performance of PGD with both improvements. Clearly, when $n \gg m$, the computational cost is greatly reduced when both improvements to PGD are used. This observation is critical for nonnegative tensor factorization, because we usually expect $D_i \gg k$, $i = 1, 2, 3$. Note that the first improvement appears to be much more effective than the second improvement suggested by Lin [38]. All tests were run on a Dell Precision 690 platform with an Intel Xeon 3.0GHz processor and 2GB available memory.

2.D. Adaptive Resampling Method

Sharp changes in magnitude along the wavelength dimension, such as narrow peaks and valleys, are often exploited in fields such as chemistry to identify ions that respond with certain absorption characteristics to specific wavelength of light [41]. These sharp changes, however, often reduce the quality of reconstructed images using NTF and other techniques [13]. For space object identification it is further observed that features of interest are often relative smooth and large, covering fairly broad spectral bands. We propose an adaptive resampling method that increases the amount of data along spectral regions with sharp changes in magnitude, leading to a reduction of frequency change of the discrete data along the z -direction.

We define spectral change using the concept of total variations. Assume the hyperspectral tensor we observe is a discrete sampling of a continuous, bounded and differentiable multivariate function, $t(x, y, \lambda)$. Note that we replace z by λ , since in our application, the third dimension is wavelength. We define the *variation* at a given wavelength $\lambda \in [\lambda_1, \lambda_{D_3}]$ as:

$$v(\lambda) = \iint_{\Omega} \left| \frac{\partial t(x, y, \lambda)}{\partial \lambda} \right| dx dy, \quad (8)$$

where Ω is a finite region in xy space determined by the detector. A discretization of

(8) using forward differences to approximate the partial derivative can be written as

$$v(\lambda_k) = v_k \approx \sum_{i=1}^{D_1} \sum_{j=1}^{D_2} \sum_{k=1}^{D_3-1} |t_{ij,k+1} - t_{ijk}|, \quad (9)$$

where $t_{ijk} = t(x_i, y_j, \lambda_k)$. Thus, v_k measures the intensity changes between neighboring wavelengths λ_k over all fibers of a tensor \mathcal{T} .

Figure 3(a) shows the variation $v(\lambda)$ that results from applying (9) to the simulated dataset used for experimentation in Section 3. The large variation at the first few wavelengths is due to sharp changes near $.4\mu m$ found in the spectral signature of composing solar panel material used in the mixture (see Figure 4). Wang [42] uses a cumulative function of $v(\lambda)$, defined as

$$c(\lambda) = \int_{\lambda_1}^{\lambda} v(\hat{\lambda}) d\hat{\lambda}, \quad (10)$$

and looks for a set of λ values that evenly divide $c(\lambda)$ from 0 to its maximum. Figure 3(b) shows a plot of $c(\lambda)$ for the set of λ values at which to resample the data. Observe that there are more sampling points around steeper slopes of $c(\lambda)$.

We propose herein a more efficient and information preserving approach that identifies and resamples $t(x, y, \lambda)$ along regions that incur large changes in magnitude, considerably reducing computation time and memory requirement, while keeping much of the original information untouched. Let \bar{v} and σ_v denote the mean and standard

deviation of v , respectively, and identify the set of discrete wavelength values, \mathbf{S} , for which $v(\lambda) > \bar{v} + 2\sigma_v$. This is the set of points for which the data must be resampled. The number of equally spaced sampling points to be introduced for each $\lambda_i \in \mathbf{S}$ is given by $v(\lambda_i)/\bar{v}$, and $t(x, y, \lambda)$ is linearly interpolated at these sampling points. Thus we are able to smooth out sharp changes by adaptively inserting interpolated data into the original observed dataset, leaving regions for which $v(\lambda) \leq \bar{v} + 2\sigma_v$ untouched. This approach significantly improves the reconstruction quality of NTF as shown in Section 3.

3. Numerical Results

In this section, we explore the efficacy of the tensor factorization approached presented in Section 2 for the reconstruction, compression, and material identification of hyperspectral data.

3.A. Simulation Data

We constructed an initial dataset of simulated spectral data using a 3-D model of the Hubble Space Telescope and a library of 262 material spectral signatures [43]. These are lab-measured the absolute reflectance by comparing the measured reflectance of each material to a known reflectance of a white reference. The 3-D model was discretized into a 128×128 array of pixels for which a specific mixture of 8 materials was assigned based on orientation of the Hubble telescope. Figure 4 shows the spectral

signatures of these materials. The signatures cover a band of spectrum from $.4\mu m$ to $2.5\mu m$ for 100 evenly distributed sampling points, leading to a data cube or tensor \mathcal{T}_0 of size $128 \times 128 \times 100$. Three other datasets, \mathcal{T}_1 , \mathcal{T}_2 , and \mathcal{T}_3 were then constructed from the \mathcal{T}_0 by considering: i) spatial blur (Gaussian point spread function with standard deviation of 2 pixels), ii) noise (independent Gaussian, and signal-dependent Poisson noise associated with the light detection process), and iii) a mixture of blur and noise. In this first study of hyperspectral analysis using NTF, we try to simulate the practical situations of observing hyperspectral data in a simple way, i.e. to simulate the atmospheric compensation using a Gaussian PSF and the various noises present in the atmosphere and in the imaging system with a noise model. These simplifications can easily be replaced by for example, more complicated turbulence models, for any future studies.

We adopt a widely-used noise model [44] for modifying the tensor \mathcal{T} to $\tilde{\mathcal{T}}$ specified by:

$$\tilde{t}_{ijk} = t_{ijk} + \eta_{ijk}^{(1)}\sqrt{t_{ijk}} + \eta_{ijk}^{(2)}, \quad (11)$$

where $\eta^{(1)} \in N(0, \sigma_1)$ and $\eta^{(2)} \in N(0, \sigma_2)$. Here we set $\sigma_1 = .05$ and $\sigma_2 = .005$.

Figure 5(a) shows images at $1.4\mu m$, for each of the four datasets. As an illustration of original material signatures assigned to each image pixel, we render an image of Hubble Space Telescope, Figure 5(b), using the color map defined in Table 3.

3.B. Compression and Reconstruction

We report the reconstruction and compression results obtained from applying Algorithm 1 together with Algorithm 2 to each of the four of datasets, \mathcal{T}_0 , \mathcal{T}_1 , \mathcal{T}_2 and \mathcal{T}_3 , for a variety of values of k . By reconstruction we mean the approximation of a tensor \mathcal{T} with a nonnegative tensor $\tilde{\mathcal{T}}$ of low rank, as described in Section 2. We report on results obtained for $k = 50$ as this value provides an appropriate balance of reconstruction quality and compression versus good recovery of original material spectral signatures. For $k = 50$ we obtained a compression ratio of **92**. Decreasing k results in a higher compression ratio but poorer reconstruction quality relative. This is not an uncommon phenomena [3, 10]. Compared with a compression ratio of 76 achieved by Du et al. [28], our result is higher but in the same order of magnitude. The possible reason is that the method from Du et al, is similar to NMF. Their method unfolds the original data cube into a large matrix and factor it into two matrices, while NTF factors the original data cube directly into three smaller matrices, and thus a higher compression ratio should be expected from NTF.

Figure 6 shows a comparison of original versus reconstructed images at wavelengths, $0.4\mu m$ and $2.0\mu m$. Figure 7 shows how the relative residual norm error decreases and converges through 25 iterations, the number of which is determined by the stopping criteria. It takes less than 1 minute to fully process each dataset \mathcal{T}_i on an Intel Xeon 3.0GHz processor. It is of interest to see that relative norm error curves of the clean

dataset and the noisy dataset almost coincide after only 3 iterations, which shows that our method is robust to noise. Note that for all relative residual norm errors, we use the original tensor \mathcal{T}_0 as the only reference to be compared with reconstructed tensors from all four datasets. The best reconstruction quality is achieved in the band of spectrum values from $1.0\mu m$ to $2.0\mu m$, as seen in Figure 9.

3.C. Adaptive Resampling and Reconstruction Quality

As previously mentioned, appropriate resampling of the spectral data can lead to better reconstruction results of NTF type algorithms. We applied the adaptive resampling approach described in Section 2 to each of the four simulation datasets, \mathcal{T}_0 (original), \mathcal{T}_1 (blurred), \mathcal{T}_2 (noisy), and \mathcal{T}_3 (blurred and noisy), increasing the number of frames along the wavelength or z-direction from 100 to 121, 121, 103, and 104, respectively. The small increase in size for the last two datasets is due to the relative increase of the total variation due to noise, which reduces the number of sampling spectral points. In general, resampling results in less than 20% increase in the tensor size for all datasets.

Figure 8 shows a comparison of the reconstruction of the original tensor, with and without resampling. Apparently, without resampling, the reconstruction quality at the $0.4\mu m$ wavelength is rather poor, due to a sharp peak observed in the spectral trace of the 'Solar Cell' material (Figure 4). After resampling, a significant improvement in reconstruction is observed.

A plot of relative residual norm error at each wavelength between the original and reconstructed tensors is given in Figure 9, where we see a large decrease at $.4\mu m$ and $.5\mu m$ for the clean dataset and the blurred dataset. For the noisy dataset and the blurred and noisy dataset we observe no improvement due to fewer sampling points added.

3.D. Endmembers Matched

For each computed tensor $\tilde{\mathcal{T}}$, we matched each column of factor \mathbf{Z} with material spectral signatures from the library, by comparing the cosine of the angle between their gradients. We chose the material for which the gradient has the smallest cosine of the angle. This choice comes after experimenting with different measures including Kullback-Liebler Divergence, Du’s measure [45], Euclidean distance, and the cosine angle measure.

In Figure 10 we compared the recovered endmember spectral signatures with the original ones using the noisy dataset and we see good matches for all materials except ‘Black Rubber Edge’, whose spectral signature possesses little feature to be detected and its prevalence is also small. Similar results are seen in the other three datasets that are not shown here. Table 4 gives a count of matched Z-factors in each dataset. Out of a total of 100 Z-factors (columns of \mathbf{Z}), we are able to match around and over 90% for all four datasets, but in the next subsection when we estimate the fractional abundances, we do see that the blurring and noise will bring down the identifica-

tion performance. Among the eight original materials, five major materials, ‘Hubble Aluminum’, ‘Hubble Green Glue’, ‘Solar Cell’, ‘Copper Stripping’, and ‘Hubble Honeycomb Top’ are matched well. Quite a few materials in the material library have signatures similar to that of ‘Hubble Honeycomb Side’. As a result we do not have a good match for this material.

3.E. Identification of Materials and their Fractional Abundances

If we unfold the original tensor \mathcal{T} along the z direction as a matrix \mathbf{T}_3 , the approximation of the tensor becomes

$$\mathbf{T}_3 \approx (\mathbf{X} \odot \mathbf{Y})\mathbf{Z}.$$

Here $\mathbf{X} \odot \mathbf{Y}$ is the mixing matrix, while \mathbf{Z} is the component matrix. Each fiber of the data cube, i.e. $\mathbf{t}_{ij} \in \mathbb{R}^{D_3}$, is approximated by

$$\mathbf{t}_{ij} \approx \sum_{l=1}^k (x_{il}y_{jl})\mathbf{z}_l,$$

which is a linear combination of Z-factors, i.e. \mathbf{z}_l . After matching Z-factors, we may have a group of Z-factors matching to the same material. To account for the material’s abundance at pixel (i, j) , we sum up the mixing coefficients of the Z-factors in the same group. Then we are left with only a few mixing coefficients at each pixel, and picking the largest one will give us the most abundant material at pixel (i, j) .

Knowing the most abundant material at each pixel, we use the same color map as given by Table 3 to render a recovered image of HST in Figure 11 to be compared with Figure 5(b). This shows that we are able to identify major materials and their spatial presence for all four datasets. However, the quality of our identification is affected by noise and blurring, especially the presence of both within our observed data.

Table 5 compares the recovered material prevalence with the true values which we have assigned to image pixels. Note that due to the usual definition of the fractional abundance as the percentage of material within each pixel, we use the term material prevalence to indicate amount of materials present within the scene. We have successfully recovered the prevalence of five major materials. However, we do see ‘Solar Cell’ abundance went up from the true 37% to around 50%, due to the difficulty in detecting ‘Black Rubber Edge’. These pixels are mostly regarded as ‘Solar Cell’. This reduces their prevalence but increases that of ‘Solar Cell’. We see the same happening to “Copper Stripping” which is located between pixels of “Solar Cell” and thus blurring would reduce its presence and attribute it to “Solar Cell”.

3.F. Comparison with a Linear Unmixing Method

By definition, NTF is a linear unmixing method with non-negativity constraints but not necessarily the sum-to-one constraint, because it further factors the coefficient matrix to achieve a better compression ratio. To compare our results with a linear unmixing method with two constraints, we first define the method using our semi-

NMF notation.

Definition 4. We define a bounded semi-NMF problem with a sum-to-one constraint

as,

$$\min_{\mathbf{h}} \Phi(\mathbf{h}) = \|\mathbf{t}_{ij} - \mathbf{S}\mathbf{h}\|_2^2 + \alpha \mathbf{h}^T \mathbf{e}, \text{ subject to } 0 \leq \mathbf{h} \leq \mathbf{u}, \quad (12)$$

where $\mathbf{e} = (1, 1, \dots, 1)$, α is the weight given to the sum-to-one term, $\mathbf{t}_{ij} \in \mathbb{R}^{D_3 \times 1} \geq 0$ is a fiber out of our data cube and $\mathbf{W} \in \mathbb{R}^{D_3 \times l} \geq 0$ is the known material signature matrix.

To simulate the supervised and unsupervised cases, we solve Problem (12) by setting the matrix \mathbf{W} with signatures either from known 8 materials or from the entire library of 262 materials. In the former case, we assume a complete prior knowledge of materials present, while in the latter case, we assume no prior knowledge on materials present. For each fiber \mathbf{t}_{ij} , Problem (12) is solved by the improved PGD method and the maximum coefficient in \mathbf{h} is regarded as the matched material and color-coded in Figures 11 and 12.

Comparing Figure 11 and 12, we can see the unsupervised NTF method matches slightly worse than the supervised linear unmixing results, but much better than the unsupervised linear unmixing results, which can only identify ‘solar cell’ and ‘copper stripping’.. Even in the supervised case, the linear unmixing method can not differentiate ‘Hubble Honeycomb Side’ from ‘Bolt’, shown as the white ellipse and two white triangles in Figure 12(a) respectively, because their signatures look somewhat

similar. The large difference between our result and the unsupervised one, might be due to the endmember matching method we have chosen, which uses the cosine of the angle between the gradients of two signatures, rather than the signatures themselves, because in the material library we see many similar signatures whose gradients would differ significantly. One last note is that after experimenting with various weights, α , we found the best results were achieved by setting α to zero, which might be due to the relatively weak mixing we simulated using the Gaussian blur with a small standard deviation. In the future research, we plan to expand the range of blurring and also simulate the atmospheric compensation through a turbulence model [46].

4. Concluding Remarks and Further Research

We have developed methods based on tensor analysis for solving three major problems in processing hyperspectral image data of an object or scene: data compression, spectral signature identification of constituent materials, and determination of their corresponding fractional abundances. Here we have proposed an approach to processing hyperspectral data which reduces a large 3-D tensor into three factor matrices, the Khatri-Rao product which approximates the original tensor. This approach preserves physical characteristics of the data such as nonnegativity. Test results have been reported for space object material identification and indicate the effectiveness of the tensor methods developed in this paper for this application.

Below is a list of tasks that we plan to consider in further work.

- Deblurring. Observe that in Figure 11 and Table 5(b) the spatial blurring affects the material identification. Running deblurring algorithms on the blurred dataset before our algorithm will likely help much in this regard. Also, we have not yet considered spectral blurring, e.g. [16]. We plan to incorporate spectral along with spatial blurring in our data simulations and to enhance the resolution (deblurring and denoising) of the spectral images before our tensor analysis of the hypercube data. Fast PDE-based total variation minimization schemes will be used, applying the approach in [47] to hyperspectral data cubes in joint work with Y. Huang and M. Ng.
- Atmospheric compensation. Imaging through turbulence is an indispensable problem that optical SOI studies must face. In the next three years through a grant funded by AFOSR, we plan to degrade our simulated dataset by a turbulence model [46], and then apply our methods described here plus the total variation deblurring method to further evaluate their success in situations closer to reality.
- Selection of k . The number, k , of factors in \mathbf{X} , \mathbf{Y} and \mathbf{Z} is selected to achieve better reconstruction quality and a better compression ratio, while covering all possible endmembers in the \mathbf{Z} -factors. In our computations we have considered various values of k , and reported on our tests for $k = 50$, leading to a compression factor of 92. We have observed in the tests that only 8 material signatures

can sufficiently describe variations in the z direction, while $k = 8$ would not be large enough to cover all the details in x and y directions. If we had chosen $k = 8$, the reconstruction quality would have been rather poor. However, choosing $k = 50$, one order higher than 8, we observed many highly correlated Z-factors, which can be collapsed into one Z-factor to further improve the compression ratio. Certain criteria need to be identified and developed to find an optimum range of values for k , which can even vary in different factors, i.e. we could have k_x, k_y and k_z values.

- The sum-to-one constraint. Due to the definition of NTF, imposing an sum-to-one constraint is more difficult, because the fractional abundances are estimated by the multiplication of two factors in two spatial dimensions. However, due to its significance in the abundance estimation, we intend to explore further in this direction.
- Further experiments. We plan to obtain unclassified AEOS Spectral Imaging Sensor (ASIS) hyperspectral data, or data based on the model of ASIS developed by Blake et al. [14], for further tests on using the tensor analysis methods developed in this paper for space object material identification.
- A new variation of NTF. We will consider a more physically motivated and less aggressive variant of the NTF defined below, in which one does not insist on a fully factorized tensor form for each term in the sum. Rather, one merely

factorizes the 1-D spectral dependence from the 2-D spatial dependence in each term. This preserves the 2-D spatial correlations of the data cube while separating out the spectral or material components. We consider a special rank- k nonnegative approximate factorization given by:

$$\min_{\mathbf{I}^{(i)}, \mathbf{z}^{(i)}} \left\| \mathcal{T} - \sum_{i=1}^k \mathbf{I}^{(i)} \circ \mathbf{z}^{(i)} \right\|, \quad \text{subject to } \mathbf{I}^{(i)} \geq \mathbf{0}, \mathbf{z}^{(i)} \geq \mathbf{0}, \quad (13)$$

where $\mathcal{T} \in \mathbb{R}^{D_1 \times D_2 \times D_3}$, $\mathbf{I}^{(i)} \in \mathbb{R}^{D_1 \times D_2}$, $\mathbf{z}^{(i)} \in \mathbb{R}^{D_3}$. Thus we anticipate a decomposition of the data into a sum of elementary images, each corresponding to a specific material and expected to be spatially sparse. For example, the matrix $\mathbf{I}^{(i)}$ could correspond to the solar panels which are localized to certain surface regions and are thus sparse over the full 2-D array, and the vector $\mathbf{z}^{(i)}$ could correspond to a solar panel material spectral signature. This has the advantage of avoiding a factoring of shapes that are not rectangular and thus non-factorizable in the Cartesian basis, e.g., circular boundaries, while one disadvantage is the lower compression ratio. However, we expect the approach to accommodate general physical situations more accurately than the conventional NTF. We are in the process of developing algorithms to compute the factorization given in (13), as well as determining appropriate constraints.

5. Acknowledgments

The authors wish to thank Kira Abercromby from the NASA Johnson Space Center for providing a spectral scan library of materials commonly found on satellites and general space objects that we used for the tests in this paper. We also wish to thank Air Force Maj. Travis Blake for providing helpful information about modeling the ASIS hyperspectral imaging system at the Maui Space Surveillance Complex, and for sharing with us his papers on enhancing the resolution of spectral images.

The research described in this paper was supported in part by grants from the Air Force Office of Scientific Research (AFOSR) with award numbers F49620-02-1-0107 and FA9550-08-1-0151, and by the Army Research Office (ARO), with award numbers DAAD19-00-1-0540 and W911NF-05-1-0402 . Their kind support is sincerely appreciated.

References

1. B. Bader, M. Berry, and M. Browne, "Discussion tracking in Enron email using PARAFAC," In *Survey of Text Mining II Clustering, Classification, and Retrieval*, M. Berry and M. Castellanos Editors, Springer, 147-163 (2008).
2. A. Cichocki, R. Zdunek, and S. Amari, "Hierarchical ALS Algorithms for Nonnegative Matrix and 3D Tensor Factorization", In: *Independent Component Analysis, ICA07*, London, UK, September 9-12, 2007, Lecture Notes in Computer Science, Vol. 4666, Springer, 169-176 (2007).

3. A. Cichocki, R. Zdunek, and S. Amari: “Nonnegative Matrix and Tensor Factorization”, *IEEE Signal Processing Magazine*, 142-145 (2008).
4. M. P. Friedlander and K. Hatz, “Computing nonnegative tensor factorizations,” Technical Report, University of British Columbia (2006).
5. A. Shashua and T. Hazan, “Non-negative tensor factorization with applications to statistics and computer vision,” *Proceedings of the 22nd International Conference on Machine Learning*, 792-799 (Bonn, 2005).
6. A. Shashua and A. Levin, “Linear image coding for regression and classification using the tensor-rank principle,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 42-49 (2001).
7. D. Chen and R. Plemmons, “Nonnegativity Constraints in Numerical Analysis.” Paper presented at the Symposium on the Birth of Numerical Analysis, Leuven Belgium, October 2007. To appear in the Conference Proceedings (2008).
8. P. Paatero and U. Tapper, *Positive matrix factorization a nonnegative factor model with optimal utilization of error-estimates of data value*, *Environmetrics*, **5**, 111-126 (1994).
9. D. Lee and H. Seung, “Learning the Parts of Objects by Non-Negative Matrix Factorization,” *Nature*, **401**, 788-791 (1999).
10. M. Berry, M Browne, A. Langville, P Pauca, and R. Plemmons, “A survey of algorithms and applications for approximate nonnegative matrix factorization,”

- Computational Statistics and Data Analysis, **52**, 155-173 (2007).
11. K. Jorgensen Abercromby, J. Africano, K. Hamada, E. Stansbery, P. Sydney and P. Kervin, "Physical properties of orbital debris from spectroscopic observations", *Advances in Space Research*, **34**, 1021-1025 (2004).
 12. P. Pauca, J. Piper, and R. Plemmons, "Nonnegative matrix factorization for spectral data analysis," *Lin. Alg. Applic.*, **416**, 29-47 (2006).
 13. P. Pauca, J. Piper R. Plemmons, and M. Giffin, "Object characterization from spectral data using nonnegative factorization and information theory," *Proceedings of AMOS Technical Conference* (Maui, 2004).
 14. T. Blake, S. Cain, M. Goda, and K. Jerkatis, "Model of the AEOS Spectral Imaging Sensor (ASIS) for Spectral Image Deconvolution," *Proceedings of AMOS Technical Conference* (Maui, 2005).
 15. T. Blake, S. Cain, M. Goda, and K. Jerkatis, "Enhancing the resolution of Spectral Images," *Proc. SPIE* **6233**, 623309 (2006).
 16. T. Blake, S. Cain, M. Goda, and K. Jerkatis, "Reconstruction of Spectral Images from the AEOS Spectral Imaging Sensor," *Proceedings of AMOS Technical Conference* (Maui, 2006).
 17. A. Plaza, P. Martinez, R. Perez and J. Plaza, "Spatial/spectral endmember extraction by multidimensional morphological operations," *IEEE Trans. on Geoscience and Remote Sensing*, **40**, 2025-2041 (2002).

18. J. Wang and C.-I. Chang, "Applications of independent component analysis (ICA) in endmember extraction and abundance quantification for hyperspectral imagery," *IEEE Trans. on Geoscience and Remote Sensing*, **44**, 2601-2616 (2006).
19. S. Kaewpijit, J. Le Moigne, T. El-Ghazawi, "Automatic Reduction of Hyperspectral Imagery Using Wavelet Spectral Analysis," *IEEE Transactions on Geoscience and Remote Sensing*, **41**, 863-871 (2003).
20. S. Qian, B. Hollinger, D. Williams, and D. Manak, "Fast three-dimensional data compression of hyperspectral imagery using vector quantization with spectral-feature-based binary coding," *Optical Engineering*, **35**, 3242-3249 (1996).
21. J.W. Boardman, F.A. Kruse, and R.O. Green, "Mapping target signatures via partial unmixing of AVIRIS data", *Summaries of the Fifth JPL Airborne Earth Science Workshop*, JPL Publication **1**, 23-26 (1995).
22. M. Winter, "N-finder: an algorithm for fast autonomous spectral endmember determination in hyperspectral data," *Image Spectrometry V, Proc. SPIE* **3753**, 266-277 (1999).
23. R. Neville, K. Staenz, T. Szeredi, J. Lefebvre and P. Hauff, "Automatic endmember extraction from hyperspectral data for mineral exploration," *4th International Airborne Remote Sensing Conf. and Exhibition/21st Canadian Symposium on Remote Sensing*, 21-24 (Canada 1999).
24. N. Keshava, "A Taxonomy of Spectral Unmixing Algorithms and Performance

- Comparisons,” Report HTAP-9, Lincoln Laboratory, MIT (2002).
25. D. Heinz and C.-I Chang, “Fully constrained least squares linear mixture analysis method for material quantification in hyperspectral imagery,” *IEEE Trans. on Geoscience and Remote Sensing*, **39**, 529-545 (2001).
 26. N. Goodwin, N.C. Coops, and C. Stone “Assessing plantation canopy condition from airborne imagery using spectral mixture analysis and fractional abundances”, *International Journal of Applied Earth Observation and Geoinformation*, **7**, 11-28 (2005).
 27. H. Kwon and N.M. Nasrabadi, “Hyperspectral target detection using kernel spectral matched filter”, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, (27), 127-127 (2004).
 28. Q. Du, C.-I Chang, D.C. Heinz, M.L.G. Althouse, I.W. Ginsberg, “A linear mixture analysis-based compression for hyperspectral image analysis”, *IEEE International Geoscience and Remote Sensing Symposium*, **2**, 585-587 (2000).
 29. K. Hege, D. O’Connell, W. Johnson, S. Basty and E. Dereniak, “Hyperspectral imaging for astronomy and space surveillance,” in *Imaging Spectrometry IX*, S. S. Chen and P. E. Lewis, eds., *Proc. SPIE* **5159**, 380-391 (2003).
 30. J. Scholl, K. Hege, M. Lloyd-Hart, D. O’Connell, W. Johnson, and E. Dereniak, “Evaluations of classification and spectral unmixing algorithms using ground based satellite imaging,” *Proc. SPIE* **6233**, 1-12 (2006).

31. N. K. M. Faber, R. Bro, and P. K. Hopke. "Recent developments in CANDECOMP/PARAFAC algorithms: a critical review," *Chemometr. Intell. Lab.*, **65**, 119-137 (2003).
32. R. A. Harshman. "Foundations of the PARAFAC procedure: models and conditions for an explanatory multi-modal factor analysis," *UCLA Working Papers in Phonetics*, **16**, 1-84 (1970).
33. L. Grippo and M. Sciandrone, "On the convergence of the block nonlinear Gauss-Seidel method under convex constraints," *Operations Research Letters*, **26**, 127-136 (2000).
34. J. Kruskal, "Three-way Arrays: Rank and Uniqueness of Trilinear Decompositions, with Applications to Arithmetic Complexity and Statistics," *Linear Alg. and Applic.*, **18**, 95-138 (1977).
35. C. L. Lawson and R. J. Hanson. *Solving Least Squares Problems*, PrenticeHall, (1974).
36. R. Bro and S. D. Jong. "A Fast Non-negativity-constrained Least Squares Algorithm," *Journal of Chemometrics*, **11**, 393-401 (1997).
37. M. Chu and M. M. Lin, "Low dimensional polytope approximation and its application to nonnegative matrix factorization," *SIAM Journal of Computing*, **30**, 1131-1155 (2008).
38. C. Lin, "Projected gradient methods for non-negative matrix factorization," *Neu-*

- ral Computation, **19**, 2756-2779 (2007).
39. R. Zdunek and A. Cichocki, "Non-Negative Matrix Factorization with Quasi-Newton Optimization," *Eighth International Conference on Artificial Intelligence and Soft Computing*, 870-879 (2006).
 40. D. Bertsekas, "On the Goldstein-Levitin-Polyak gradient projection method," *IEEE Transactions on Automatic Control*, **21**, 174-184 (1976).
 41. D. R. Lide (ed.), *CRC Handbook of Chemistry and Physics*, 83rd ed., Boca Raton, FL: CRC Press (2002).
 42. H. Wang, "Nonnegative tensor factorization for hyperspectral data analysis", Master Thesis, Department of Mathematics, Wake Forest University, (2007).
 43. K. Jorgensen Abercromby, NASA Johnson Space Center (personal communication 2006).
 44. A.K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, (1989).
 45. Y. Du, C-I. Chang, H. Ren, C-C. Chang, J.O. Jensen and F.M. D'Amico, "New hyperspectral discrimination measure for spectral characterization", *Optical Engineering*, **43**, 1777-1786 (2004).
 46. M.C. Roggemann and B. Welsh, *Imaging Through Turbulence*, CRC Press, (1996).
 47. Y. Huang, M. Ng and Y. Wen, "A Fast Total Variation Minimization Method for Image Restoration," Preprint (2008).

Table 1. Complexity analysis of Armijo Rules defined in (6) and (7)

Formula	Number of multiplications	Complexity	Pre-computed
$\mathbf{H} \otimes \mathbf{H}$	mk	$O(mk)$	
$\Phi(\mathbf{H})$	$mnk + 2mn$	$O(mnk)$	
$\mathbf{W}^T \mathbf{W}$	nk^2	$O(nk^2)$	
$(\mathbf{W}^T \mathbf{W})\mathbf{H}$	mk^2	$O(mk^2)$	$\mathbf{W}^T \mathbf{W}$
$\nabla \Phi(\mathbf{H})$	$mk^2 + mnk$	$O(mnk)$	$\mathbf{W}^T \mathbf{W}$
(6)	$m(2nk + 4n + 2k)$	$O(mnk)$	$\mathbf{W}^T \mathbf{W}, \nabla \Phi(\mathbf{H})$
(7)	$m(k^2 + 4k)$	$O(mk^2)$	$\mathbf{W}^T \mathbf{W}, \nabla \Phi(\mathbf{H})$

Table 2. Benchmarks for two improvements to the Projected Gradient Descent method. No improvement in Test 0, first improvement only in Test 1, second improvement only in Test 2 and both improvements in Test 3 (in seconds).

Matrix Size ($n \times m$)	Test 0	Test 1	Test 2	Test 3
500×500	7.28	0.38	1.39	0.078
50×5000	13.15	3.13	2.60	1.22
5000×50	7.53	0.083	1.31	0.008
1000×1000	28.51	0.97	5.18	0.20
100×10000	40.64	7.34	11.68	2.74
10000×100	29.97	0.18	5.36	0.037

Table 3. Materials, colors and fractional abundances used for Hubble satellite simulation.

Material	Color	Fractional Abundance (%)
Hubble Aluminum	light gray	19
Hubble Green Glue	dark gray	12
Hubble Honeycomb Top	white	4
Hubble Honeycomb Side	blue	3
Solar Cell	gold	37
Bolts	red	3
Black Rubber Edge	dark gray	8
Copper Stripping	cyan	13

Table 4. Numbers of matched Z-factors in the four datasets.

Material	Clean	Blurred	Noisy	Blurred & Noisy
Hubble Aluminum	5	6	7	5
Hubble Green Glue	11	6	11	8
Black Rubber Edge	0	2	1	0
Bolts	1	2	1	1
Hubble Honeycomb Side	1	0	1	0
Hubble Honeycomb Top	3	2	2	3
Solar Cell	11	13	8	13
Copper Stripping	13	16	19	15
Total matched	45	47	50	45

Table 5. True and recovered material prevalences in percent.

Material	True	Clean	Blurred	Noisy	Blurred & Noisy
Hubble Aluminum	19	19	12	19	19
Hubble Green Glue	12	14	16	13	14
Hubble Honeycomb Top	4	5	5	5	5
Hubble Honeycomb Side	3	4	0	3	1
Solar Cell	37	48	55	50	52
Bolts	3	0	0	1	2
Black Rubber Edge	8	0	0	0	0
Copper Stripping	13	11	12	10	7

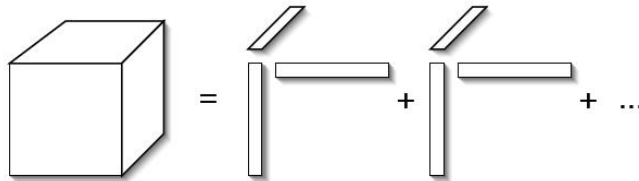


Fig. 1. An illustration of 3-D tensor approximate factorization using a sum of rank one tensors.

Table 6. List of Figures

Figure 1	An illustration of 3-D tensor approximate factorization.
Figure 2	A blurred and noisy simulated hyperspectral image of the Hubble Space Telescope
Figure 3	Total variation and cumulative variation for the simulated data
Figure 4	Spectral signatures of materials assigned to a HST model
Figure 5	(a) Simulated image at wavelength $.4\mu m$. (b) Materials assigned to pixels in HST
Figure 6	Comparison of frames from the original and the reconstructed data
Figure 7	Relative residual norm error descent through the iterations
Figure 8	Reconstructed images from NTF at wavelength $0.4212\mu m$, with and without resampling
Figure 9	Comparison of reconstruction error with and without resampling in preprocessing
Figure 10	Matched endmembers using the noisy dataset
Figure 11	Reconstructed color image of HST for materials identification
Figure 12	Results from a linear unmixing method with and without assuming prior knowledge

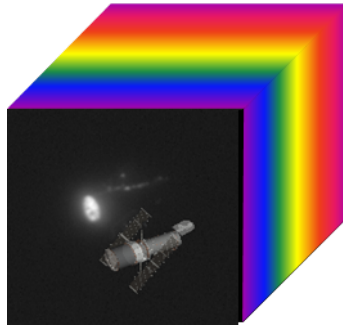
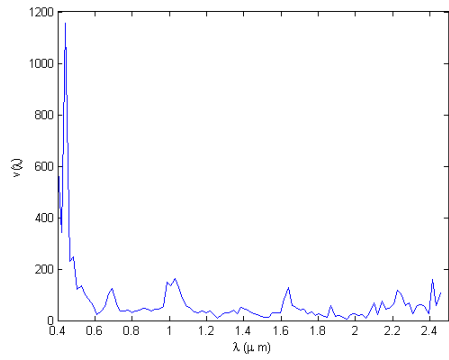
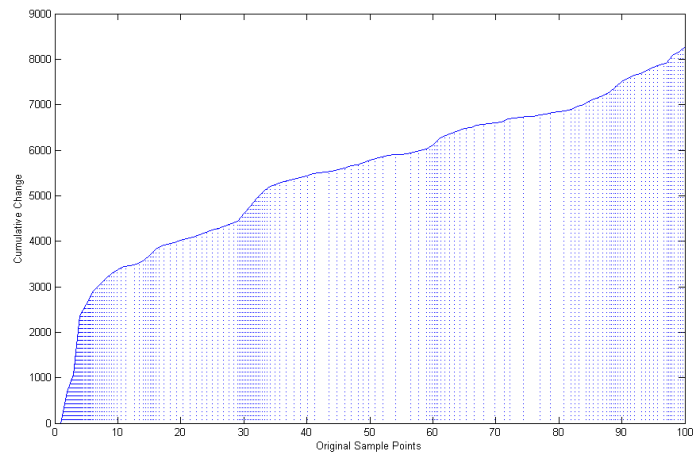


Fig. 2. A blurred and noisy simulated hyperspectral image above the original simulated image of the Hubble Space Telescope [14], representative of the data collected by the Maui ASIS system.



(a)



(b)

Fig. 3. (a) Total variation between observations at neighboring wavelengths for the simulated data from Section 3. (b) Cumulative variation across all wavelengths.

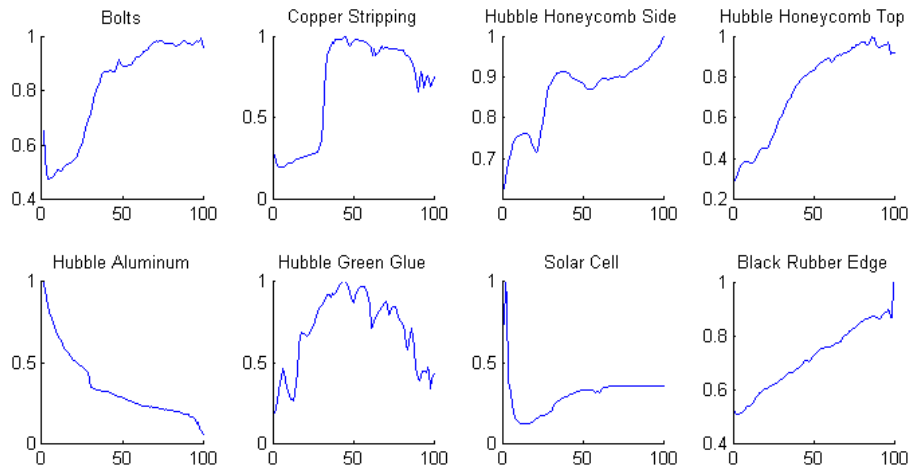


Fig. 4. Spectral signatures of eight materials assigned to a Hubble Space Telescope model.

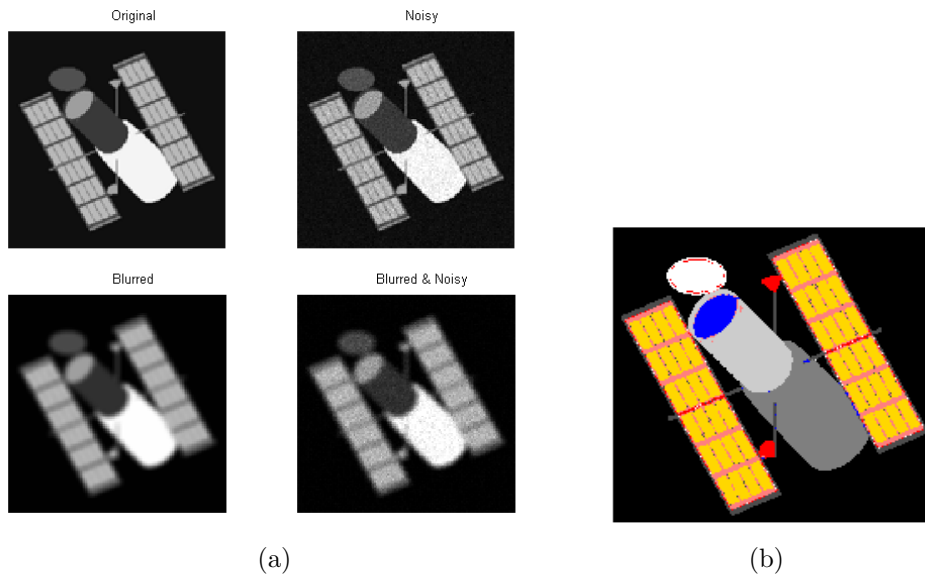
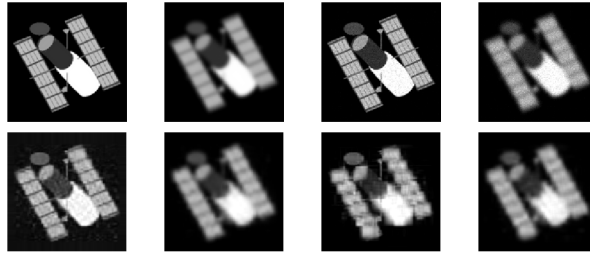
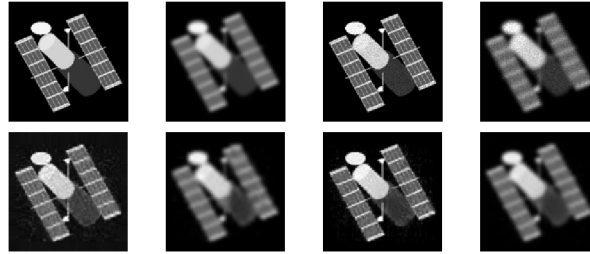


Fig. 5. (a) Simulated data sets with images from the data cube at wavelength $4\mu m$. (b) Materials assigned to pixels in Hubble Satellite Telescope image. All images have a size of 128×128 .



(a) Images at $.4\mu m$.



(b) Images at $2.0\mu m$.

Fig. 6. Comparison of frames from the original data $\mathcal{T}_0, \mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3$ (top row in (a) and (b)) with frames from the reconstructed data $\tilde{\mathcal{T}}_0, \tilde{\mathcal{T}}_1, \tilde{\mathcal{T}}_2, \tilde{\mathcal{T}}_3$ (bottom row in (a) and (b)). From left to right, each column corresponds to original, blurred, noisy, and blurred and noisy data.

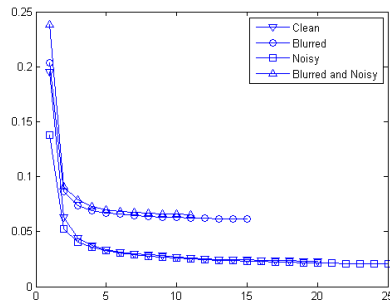


Fig. 7. Relative residual norm error descent through the iterations. The stopping criteria is when the relative norm difference of two successive approximations falls below a lower limit, e.g. 10^{-5} .

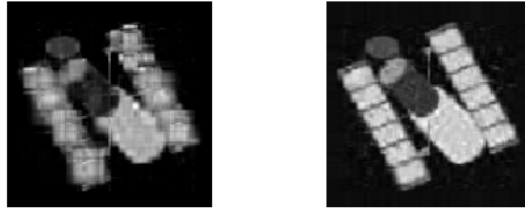


Fig. 8. Reconstructed images from NTF at wavelength $0.4212\mu m$. The one on the right was resampled before processing. Both images have a size of 128×128 .

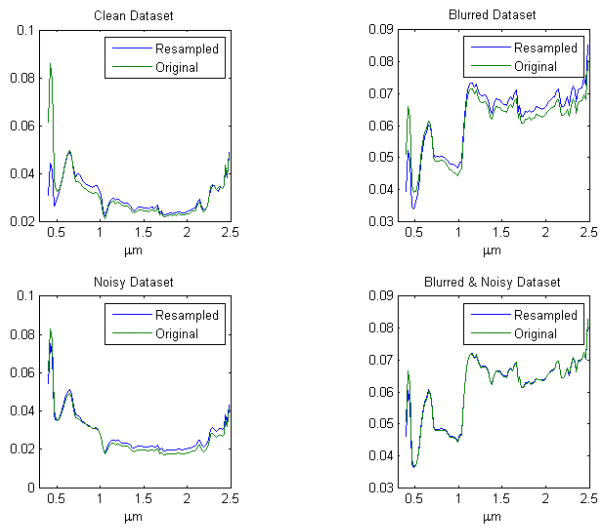


Fig. 9. Comparison of reconstruction error between with and without resampling in preprocessing for all four datasets.

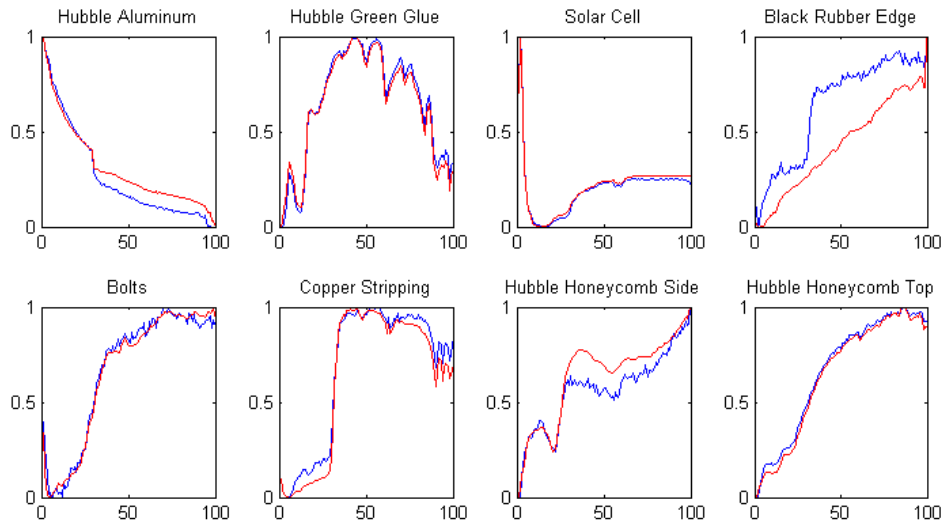


Fig. 10. Matched endmembers using the noisy dataset, $k=50$. Red is the original material spectrum and blue is the best matched Z-factor.

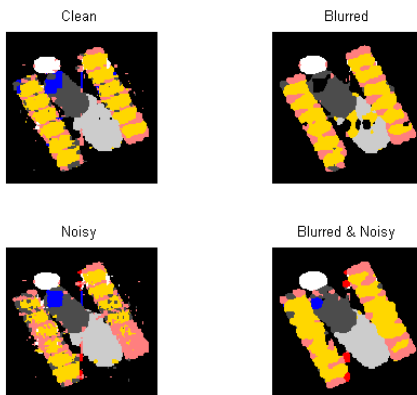
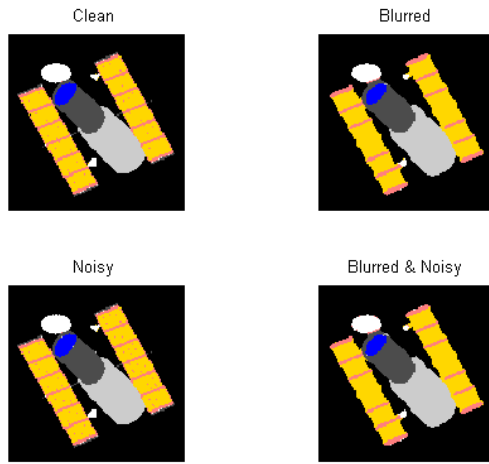
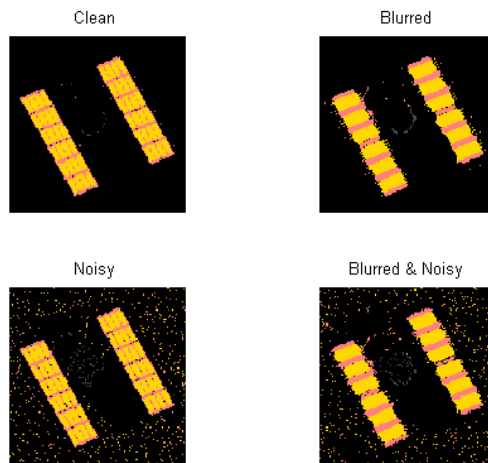


Fig. 11. Reconstructed color image of HST using the color map in Table 3 for original materials.



(a)



(b)

Fig. 12. (a) Linearly unmixed and matched materials using only 8 known materials. (b) Linearly unmixed and matched materials using the whole library of 262 materials.