

Nonnegative Matrix Factorization for Spectral Data Analysis

V. Paul Pauca* J. Piper† Robert J. Plemmons‡

Abstract

Data analysis is pervasive throughout business, engineering and science. Very often the data to be analyzed is nonnegative, and it is often preferable to take this constraint into account in the analysis process. Here we are concerned with the application of analyzing data obtained using astronomical spectrometers, which provide spectral data which is inherently nonnegative. The identification and classification of space objects that cannot be imaged in the normal way with telescopes is an important but difficult problem for tracking thousands of objects, including satellites, rocket bodies, debris, and asteroids, in orbit around the earth. In this paper we develop an effective nonnegative matrix factorization algorithm with novel smoothness constraints for unmixing spectral reflectance data for space object identification and classification purposes. Promising numerical results are presented using laboratory and simulated datasets.

Keywords: Nonnegative matrix factorization, spectral data, blind source separation, data mining, space object identification and classification.

1 Introduction

We are concerned with methods based on nonnegative matrix factorization (NMF) for the analysis of spectral reflectance data associated with space objects. As will be described later,

*Department of Computer Science, Wake Forest University, Winston-Salem, NC 27109. His research was supported in part by the Air Force Office of Scientific Research under grant FA49620-03-1-0215, and by the Army Research Office under grant DAAD19-00-1-0540. Email: paucavp@wfu.edu

†Department of Computer Science, Wake Forest University, Winston-Salem, NC 27109. His research was supported in part by the Air Force Office of Scientific Research under grant F49620-02-1-0107. Email: pipejw02@wfu.edu

‡Departments of Computer Science and Mathematics, Wake Forest University, Winston-Salem, NC 27109. His research was supported in part by the Air Force Office of Scientific Research under grant F49620-02-1-0107 and by the Army Research Office under grant DAAD19-00-1-0540. Email: plemmons@wfu.edu

the nonnegative spectral values are distinguished by different wavelengths of visible light, and each observation of an object is stored as a column in a data matrix, which we denote by Y , in which the rows correspond to different wavelengths.

In general, the NMF problem is the following: given a nonnegative data matrix Y , find reduced rank nonnegative matrices W and H so that

$$Y \approx WH.$$

Here, W is often thought of as the endmember or source matrix, and H the fractional abundance or mixing matrix associated with the data in Y . A more formal definition of the problem is given later. This approximate factorization process is an active area of research in several disciplines (a Google search on this topic recently provided over 200 references to papers involving nonnegative matrix factorization and applications, written in the past ten years), and the subject is certainly a fertile area of research for linear algebraists. It has been shown to be an especially effective tool in several areas of data mining, e.g., [6, 7, 9, 12, 15, 21, 22, 23, 26], and in this paper we consider the application of NMF to the mining and analysis of spectral data.

There are several concerns in modern data analysis. First, information gathering devices, such as spectrometers, have only finite bandwidth. One thus cannot avoid the fact that the data collected often are not exact. For example, signals received by antenna arrays often are contaminated by noise and other degradations. Before useful deductive science can be applied, it is often important to first reconstruct or represent the data so that the inexactness is reduced while certain feasibility conditions are satisfied. Secondly, in many situations the data observed from complex phenomena represent the integrated result of several interrelated variables acting together. When these variables are less precisely defined, the actual information contained in the original data might be overlapping and ambiguous.

A reduced system model could provide a fidelity near the level of the original system. One common ground in the various approaches for noise removal, model reduction, feasibility reconstruction, and so on, is to replace the original data by a lower dimensional representation obtained via subspace approximation. The notion of low rank approximations therefore arises in a wide range of important applications. Factor analysis and principal component analysis are two of the many classical methods used to accomplish the goal of reducing the number of variables and detecting structures among the variables.

However, as indicated above, often the data to be analyzed is nonnegative, and the low rank data are further required to be comprised of nonnegative values only in order to avoid contradicting physical realities. Classical tools cannot guarantee to maintain the nonnegativity. The approach of low-rank **nonnegative matrix factorization** (NMF) thus becomes particularly appealing. The NMF problem can be stated in *generic form* as follows:

(The NMF Problem) *Given a nonnegative matrix $Y \in R^{m \times n}$ and a positive integer $k < \min\{m, n\}$, find nonnegative matrices $W \in R^{m \times k}$ and $H \in R^{k \times n}$ to minimize the functional*

$$f(W, H) := \frac{1}{2} \|Y - WH\|_F^2. \quad (1)$$

The product WH is called a nonnegative matrix factorization of Y , although Y is not necessarily equal to the product WH . Clearly the product WH is of rank at most k . An appropriate decision on the value of k is critical in practice, but the choice of k is very often problem dependent. The objective function (1) can be modified in several ways to reflect the application need. For example, penalty terms can be added to $f(W, H)$ in order to enforce sparsity or to enhance smoothness in the solution matrices W and/or H [11, 22]. Also, because $WH = (WD)(D^{-1}H)$ for any invertible matrix $D \in R^{k \times k}$, sometimes it is desirable to “normalize” the columns of W . The question of uniqueness of the nonnegative

factors W and H thus arises, which is easily seen by considering the case where the matrices D and D^{-1} are nonnegative, see e.g. Donaho and Stodden [8], for clearly the optimization problem (1) can have local minima. In many applications lack of uniqueness can be mitigated using additional constraints. There is also often a need to further constrain the problem by determining as few factors as possible relative to the application, in order to compress the data, and hence the need for a low rank NMF of the data matrix Y arises.

Our particular application developed in this paper is the identification and classification of space objects by applying novel NMF-based approaches. The determination of space objects whose orbits are significantly distant (e.g. geosynchronous satellites) or whose dimensions are small (e.g. nanosatellites) is a challenging problem in astronomical imaging. Because of their distant orbits or small dimensions these objects are difficult to be resolved spatially using ground-based telescope technology, hence they are denoted as nonimaging objects. Recently, alternate approaches have been proposed in an attempt to circumvent this problem. Among these, the collection and analysis of wavelength-resolved data is a promising and viable approach currently under investigation. Here a spectrometer is employed to collect spectral reflectance traces of a space object over a specific period of time.

A fundamental difficulty with this approach is the determination of characteristics of the material that make up an object from its collection of spectral traces. More specifically, it is desired to determine i) the type of constituent material and ii) the proportional amount of these materials that make up the object. The first problem involves the detection of material spectral signatures, or *endmembers* from the spectral data. The second problem involves the computation of corresponding proportional amounts, or *fractional abundances*.

Here, we propose an unsupervised method for detecting endmembers as well as for computing fractional abundances from the spectral reflectance traces of a non-imaging space object. Our methodology extends previous research [4, 18, 22, 21] in at least two ways. It deals with the material spectral unmixing problem numerically by solving an optimization

problem with nonnegative and smoothness constraints in the data. It then uses the Kullback-Leibler divergence, an information-theoretic measure, to select computed endmembers that best match known materials in the database. Fractional abundances are then computed using the selected endmembers using constrained least squares methods.

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 computation, in the related context of hyperspectral unmixing can be found in the work of Plaza *et al.*, [24]. An information-theoretic approach has been provided by Chang [5]. Recently, nonnegative matrix factorization (NMF), proposed by Lee and Seung [15], has been used with relative success in this problem. We extend NMF by considering a class of smoothness constraints on the optimization problem for more accurate endmember computation.

The Kullback-Leibler divergence measure is selected over other methods for matching computed endmembers against material traces stored in a database. It has several advantages over simpler methods, such as the cosine of the angle between material traces. The Kullback-Leibler divergence is an information, or cross-entropy measure that captures the spectral correlation between traces and facilitates more accurate matching and selection of computed endmembers.

This paper is organized as follows. First we introduce the standard [13] linear mixing model for spectral data in Section 2, and relate it to our application. Our novel techniques for endmember extraction, selection, and quantification are described in Section 3. Results from some extensive numerical simulations studies are provided in Section 4. These computations illustrate the effectiveness of these techniques for our application.

2 A Linear Mixing Model and Spectral Data Analysis

A spectral trace is often modeled as a linear combination of a set of endmembers, weighted by corresponding material quantities or fractional abundances. Let \mathbf{s}_i ($1 \leq i \leq k$) be endmembers corresponding to different material spectral signatures. Each endmember \mathbf{s}_i is a vector containing nonnegative spectral measurements along m different spectral bands. Then, an observed spectral trace \mathbf{y} can be expressed as $\mathbf{y} = S\mathbf{x} + \mathbf{n}$, where $S = \begin{bmatrix} \mathbf{s}_1 & \dots & \mathbf{s}_k \end{bmatrix}$ is an $m \times k$ matrix whose columns are the endmembers, \mathbf{x} is a vector of fractional abundances, and \mathbf{n} is an additive observation noise vector. Moreover, $\mathbf{x} \geq 0$ and $\sum \mathbf{x}_i = 1$.

When a sequence of n time-varying spectral traces $\{\mathbf{y}_i\}$ is considered, matrix notation is appropriate to express all traces as

$$Y = SX + N, \quad (2)$$

where

$$Y = \begin{bmatrix} \mathbf{y}_1 & \dots & \mathbf{y}_n \end{bmatrix}, \quad X = \begin{bmatrix} \mathbf{x}_1 & \dots & \mathbf{x}_n \end{bmatrix}, \quad N = \begin{bmatrix} \mathbf{n}_1 & \dots & \mathbf{n}_n \end{bmatrix}. \quad (3)$$

The linear mixing assumption is typical in many hyperspectral imaging applications [13] and appropriate within the context of nonimaging object identification. Previous work [4, 18, 22] has demonstrated its effectiveness in the recovery of fractional abundances, when the endmember matrix S is known *a priori*.

However, spectral data analysis for nonimaging object identification involves the more difficult problem of computing S as well as X given only the matrix of spectral traces Y . Recently several computational methods have been proposed [4, 18, 21, 22] that provide reasonable approximations to S and X . Other techniques have been proposed in related areas such as blind image deconvolution that include nonnegativity constraints [20].

3 Nonnegative Matrix Factorization Algorithms

Nonnegative matrix factorization (NMF) has been shown to be a very useful technique in approximating high dimensional data where the data are comprised of nonnegative components. In a seminal paper published in *Nature* [15], Lee and Seung proposed the idea of using NMF techniques to find a set of basis functions to represent image data where the basis functions enable the identification and classification of intrinsic “parts” that make up the object being imaged by multiple observations. Low-rank nonnegative matrix factorizations not only enable the user to work with reduced dimensional models, they also often facilitate more efficient statistical classification, clustering and organization of data, and lead to faster searches and queries for patterns or trends, e.g., Pauca, Shahnaz, Berry and Plemmons [23].

3.1 General Nonnegative Matrix Factorization

One nonnegative matrix factorization algorithm developed by Lee and Seung [15] is based on multiplicative update rules of W and H . We refer to this multiplicative scheme simply by **NMF**. A formal statement of the method is given in Algorithm 1.

Algorithm 1 (NMF) Calculate W and H such that $Y \approx WH$

Require: $m \times n$ matrix $Y \geq 0$, and $k > 0, k \ll \min(m, n)$

$W_{ij} \leftarrow$ nonnegative value ($1 \leq i \leq m, 1 \leq j \leq k$)

$H_{ij} \leftarrow$ nonnegative value ($1 \leq i \leq k, 1 \leq j \leq n$)

Scale columns of W to sum to one.

while converge or stop **do**

$$H_{cj} \leftarrow H_{cj} \frac{(W^T Y)_{cj}}{(W^T W H)_{cj} + \epsilon}, \quad (1 \leq c \leq k, 1 \leq j \leq n)$$

$$W_{ic} \leftarrow W_{ic} \frac{(Y H^T)_{ic}}{(W H H^T)_{ic} + \epsilon}, \quad (1 \leq i \leq m, 1 \leq c \leq k)$$

Scale columns of W to sum to one.

end while

Clearly the approximations W and H remain nonnegative during the updates. It is generally best to update W and H “simultaneously”, instead of updating each matrix fully

before the other. In this case, after updating a row of H , we update the corresponding column of W . Matlab performs well with these computations. In the implementation, a small positive quantity, say the square root of the machine precision, should be added to the denominators in the approximations of W and H at each iteration step. We use a parameter $\epsilon = 10^{-9}$ in our Matlab code for this purpose.

It is often important to normalize the columns of Y in a pre-processing step, and in the algorithm to normalize the columns of the basis matrix W at each iteration. In this case we are optimizing on a unit hypersphere, as the column vectors of W are effectively mapped to the surface of a hypersphere by the repeated normalization.

The computational complexity of **Algorithm 1 (NMF)** can be shown to be $O(knm)$ operations per iteration. However, the initial factorization needs only be computed in its entirety once. If data are added to the database then it can either be added directly to the basis matrix W along with a minor modification of H , or else if k is fixed, then further iterations can be applied starting with the current W and H as initial approximations.

Lee and Seung [16] proved that under the NMF update rules the distance $\|Y - WH\|_F^2$ is monotonically non-increasing. In addition it is invariant if and only if W and H are at a stationary point of the objective function (1). From the viewpoint of nonlinear optimization, the algorithm can be classified as a diagonally-scaled gradient descent method [9].

Several other methods have been proposed for nonnegative matrix factorization. (See [17] for a recent survey). In particular, Lee and Seung [15] have also provided an additive algorithm. Both the multiplicative and additive algorithms are related to expectation-maximization approaches used in image processing computations such as image restoration, e.g., [25]. Hoyer [11] has suggested a novel nonnegative sparse coding scheme based on ideas from the study of neural networks, and the scheme has been applied to the decomposition of databases into independent feature subspaces by Hyvärinen and Hoyer [12]. Other methods based in part on orthogonal factorizations have been considered by Liu and Yi [17]. Studies

and comparisons of various algorithms for nonnegative factorizations have been given by D. Guillamet et al [9, 10] and by Liu and Yi [17]. Each approach has advantages and disadvantages, but in general the algorithms examined in these papers have been found to be effective on certain applications, and so the method of choice is often application dependent.

3.2 Incorporating Additional Constraints

Here, we propose a new constrained NMF optimization problem,

$$\min_{W,H} \{ \|Y - WH\|_F^2 + \alpha J_1(W) + \beta J_2(H) \}, \quad (4)$$

for $W \geq 0$ and $H \geq 0$, where $\|\cdot\|_F$ is the Frobenius norm, the columns of $W = \begin{bmatrix} \mathbf{w}_1 & \dots & \mathbf{w}_k \end{bmatrix}$, for some specified value of k , are the possible endmembers, and H is a by-product matrix of mixing coefficients. The functions $J_1(W)$ and $J_2(H)$ are penalty terms used to enforce certain constraints on the solution of Equation (4), and α and β are their corresponding Lagrangian multipliers, or regularization parameters. Different penalty terms may be used depending upon the desired effects on the computed solution. For the current application, we set

$$J_1(W) = \|W\|_F^2, \quad (5)$$

in order to enforce smoothness in W . A similar constraint may be placed on H via $J_2(H)$ to enforce, for example, statistical sparseness [11], which is useful in certain text clustering applications [23]. The effectiveness of these additional yet simple constraints is demonstrated in the next section, where it is shown that endmembers of higher quality can be obtained than with NMF alone.

We derive an algorithm to solve Equation (4) with smoothness constraint (5). For completeness we apply the same constraint on H , namely letting $J_2(H) = \|H\|_F^2$. Let $F(W, H)$

be a scalar-valued function defined by

$$F(W, H) = \frac{1}{2}\|Y - WH\|_F^2 + \frac{\alpha}{2}\|W\|_F^2 + \frac{\beta}{2}\|H\|_F^2. \quad (6)$$

NMF is based on an alternating gradient descent mechanism. That is, for some starting matrices $W^{(0)}$ and $H^{(0)}$, the sequences $\{W^{(t)}\}$ and $\{H^{(t)}\}$, $t = 1, 2, \dots$, are computed by means of the formulas,

$$W_{ij}^{(t)} = W_{ij}^{(t-1)} - \theta_{ij} \frac{\partial F(W, H)}{\partial W_{ij}}, \quad 1 \leq i \leq m, \quad 1 \leq j \leq k \quad (7)$$

$$H_{ij}^{(t)} = H_{ij}^{(t-1)} - \phi_{ij} \frac{\partial F(W, H)}{\partial H_{ij}}, \quad 1 \leq i \leq k, \quad 1 \leq j \leq n. \quad (8)$$

Taking the derivatives of $F(W, H)$ with respect to W_{ij} and H_{ij} and after some algebraic manipulations, the gradients about W_{ij} and H_{ij} can be expressed in matrix-vector form as,

$$\frac{\partial F(W, H)}{\partial W_{ij}} = -YH^T + WHH^T + \alpha W, \quad (9)$$

$$\frac{\partial F(W, H)}{\partial H_{ij}} = -W^TY + W^TWH + \beta H. \quad (10)$$

The scalar quantities θ_{ij} and ϕ_{ij} determine the lengths of the steps to take along the negative of the gradients about W_{ij} and H_{ij} , respectively, and must be chosen carefully. Hoyer [11] selects $\theta_{ij} = c_1$ and $\phi_{ij} = c_2$ for some constants c_1 and c_2 . Here we follow Lee and Seung [15] and choose these quantities based on current estimates of W and H ,

$$\theta_{ij} = \frac{W_{ij}^{(t-1)}}{(W^{(t-1)}HH^T)_{ij}}, \quad (11)$$

$$\phi_{ij} = \frac{H_{ij}^{(t-1)}}{(W^TWH^{(t-1)})_{ij}}. \quad (12)$$

Substituting (9), (10), (11), and (12) into (7) and (8), we obtain the following update

rules,

$$W_{ij}^{(t)} = W_{ij}^{(t-1)} \cdot \frac{(YH^T)_{ij} - \alpha W_{ij}^{(t-1)}}{(W^{(t-1)}HH^T)_{ij}} \quad (13)$$

$$H_{ij}^{(t)} = H_{ij}^{(t-1)} \cdot \frac{(W^TY)_{ij} - \beta H_{ij}^{(t-1)}}{(W^TWH^{(t-1)})_{ij}}. \quad (14)$$

It is not difficult to prove that for our choices of $J_1(W)$ and $J_2(H)$, the function $F(W, H)$ is non-increasing under the updates rules (13) and (14). Next, we summarize our new constrained version of NMF.

Algorithm 2 (CNMF) Calculate W and H such that $Y \approx WH$

Require: $m \times n$ matrix $Y \geq 0$, and $k > 0, k \ll \min(m, n)$

$W_{ij} \leftarrow$ nonnegative value ($1 \leq i \leq m, 1 \leq j \leq k$)

$H_{ij} \leftarrow$ nonnegative value ($1 \leq i \leq k, 1 \leq j \leq n$)

Scale columns of W to sum to one.

while converge or stop **do**

$$H_{cj}^{(t)} \leftarrow H_{cj}^{(t-1)} \frac{((W^{(t-1)})^TY)_{cj} - \beta H_{cj}^{(t-1)}}{((W^{(t-1)})^TW^{(t-1)}H^{(t-1)})_{cj} + \epsilon} \quad (1 \leq i \leq m, 1 \leq c \leq k)$$

$$W_{ic}^{(t)} \leftarrow W_{ic}^{(t-1)} \frac{(Y(H^{(t)})^T)_{ic} - \alpha W_{ic}^{(t-1)}}{(W^{(t-1)}H^{(t)}(H^{(t)})^T)_{ic} + \epsilon} \quad (1 \leq c \leq k, 1 \leq j \leq n)$$

Scale columns of W to sum to one.

end while

The regularization parameters $\alpha \in \mathcal{R}$ and $\beta \in \mathcal{R}$ are used to balance the trade-off between accuracy of the approximation and smoothness of the computed solution and their selection is typically problem dependent. Of course the matrices W and H are generally not unique. Conditions resulting in uniqueness in the special case of equality, $Y = WH$, have been recently studied by Donoho and Stodden [8], using cone theoretic techniques (See also Chapter 1 in Berman and Plemmons [2]). The case where Y is symmetric is studied by Catral, Han, Neumann and Plemmons [3].

4 Endmember Selection using Information-Theoretic Techniques

In this section we describe our use of Kullback-Leibler information theoretic measures to select computed endmembers by calculating their proximity to laboratory-obtained spectral signatures of space object material. Related measures have been recently considered in the context of hyperspectral imaging, in particular by Chang [5]. Let a vector $\mathbf{w} \geq 0$ (a column of W) denote a random variable associated with an appropriate probability space. The probability density vector \mathbf{p} associated with \mathbf{w} is given by

$$\mathbf{p} = \frac{1}{\sum w_i} \mathbf{w}. \quad (15)$$

The vector \mathbf{p} can be used to describe the variability of a spectral trace. Moreover, the entropy of each \mathbf{w} can be defined, using base-two logarithms, as

$$h(\mathbf{w}) = - \sum p_i \log p_i, \quad (16)$$

which can then be used to describe the uncertainty [5] in \mathbf{w} . Now, let \mathbf{d} be a laboratory-obtained material spectral signature with probability density vector $\mathbf{q} = (1/\sum d_i)\mathbf{d}$. Then using information theory the relative entropy of \mathbf{d} with respect to \mathbf{w} can be defined by

$$D(\mathbf{w}||\mathbf{d}) = \sum p_i(\log(p_i) - \log(q_i)) = \sum p_i \log\left(\frac{p_i}{q_i}\right), \quad (17)$$

where the quantities $-\log p_i$ and $-\log q_i$ are the self-information measures of \mathbf{w} and \mathbf{d} , respectively, along the i th spectral band. Here, $D(\mathbf{w}||\mathbf{d})$, is known as the Kullback-Leibler divergence (KLD) measure, but it is not symmetric with respect to \mathbf{w} and \mathbf{d} . The following

symmetrized version [5] may instead be considered,

$$D_s(\mathbf{w}, \mathbf{d}) = D(\mathbf{w}||\mathbf{d}) + D(\mathbf{d}||\mathbf{w}). \quad (18)$$

In our approach, the *symmetric Kullback-Leibler Divergence* (SKLD) measure (18) is used to compute a confusion matrix $C(W, L)$ between each column \mathbf{w}_i in matrix W and each trace \mathbf{d}_j in a laboratory-obtained dataset L of material spectral signatures. For each \mathbf{d}_j , $\arg \min_{\mathbf{w}_i} C(W, \mathbf{d}_j)$ indicates the endmember candidate in W to which \mathbf{d}_j is closest to in terms of (18). Consequently each material \mathbf{d}_j in L has associated with it some endmember candidate \mathbf{w}_i . A tolerance parameter τ is used to reject endmembers \mathbf{w}_i for which the SKLD is larger than τ . The selected columns of W are then used to form a matrix B . This matrix is subsequently used in the third step of our approach, which is to quantify fractional abundances.

5 Quantification of Fractional Abundances

For quantifying fractional abundances, given $B \approx S$, we solve

$$J(X) = \min_X \|Y - BX\|_F^2 \quad (19)$$

subject to nonnegativity constraint $X \geq 0$. A variety of solution techniques can be used. We select PMRNSD, a preconditioned fixed point iteration type approach proposed by Nagy and Strakos [20] and implemented by Nagy, Palmer, and Perrone [19]. In this approach, the constrained minimization problem is first transformed into an unconstrained problem by parameterizing X as $X = e^Z$.

We consider, without loss of generality, the equivalent case of minimizing (19) one column

at the time,

$$J(\mathbf{x}) = \min_{\mathbf{x}} \|\mathbf{y} - B\mathbf{x}\|_2^2, \quad \text{for } \mathbf{x} = e^{\mathbf{z}}. \quad (20)$$

A local non negatively constrained minimizer of (20) can be computed via the fixed point iteration,

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \tau_k \mathbf{v}_k, \quad \mathbf{v}_k = -\text{diag}(\nabla J(\mathbf{x}_k)) \mathbf{x}_k, \quad (21)$$

where $\text{diag}(\nabla J(\mathbf{x}_k))$ is a diagonal matrix whose entries are given by $\nabla J(\mathbf{x}_k) = B^T(\mathbf{y} - B\mathbf{x})$.

In the PMRNSD algorithm the parameter τ_k is given by $\tau_k = \min\{\tau_{uc}, \tau_{bd}\}$, where τ_{uc} corresponds to the unconstrained minimizer of J in the direction \mathbf{v}_k and τ_{bd} is the minimum $\tau > 0$ such that $\mathbf{x}_k + \tau \mathbf{v}_k \geq 0$. Circulant type preconditioners are employed to speed up the convergence of the fixed point iteration method [19].

6 Numerical Results

6.1 Construction of Simulated Data

Four satellites were simulated with varying compositions of aluminum, mylar, solar cell, and white paint. The surface of each satellite was composed of 50% of one of the materials and nearly equal amounts of the remaining three materials. These fractional abundances were then allowed to vary sinusoidally with different randomly generated amplitudes, frequencies, and phase shifts, so as to model space object rotation with respect to a fixed detector. Laboratory-obtained spectra corresponding to aluminum, mylar, solar cell, and white paint, shown in Figure 1, were used to generate $p = 100$ spectral traces of length $m = 155$ for each satellite using the linear mixing model equation (2), with 1% additive Gaussian noise.

Representative simulated spectral traces for each satellite are shown in Figure 2. These simulations are consistent with simulated and real data employed in related work [18].

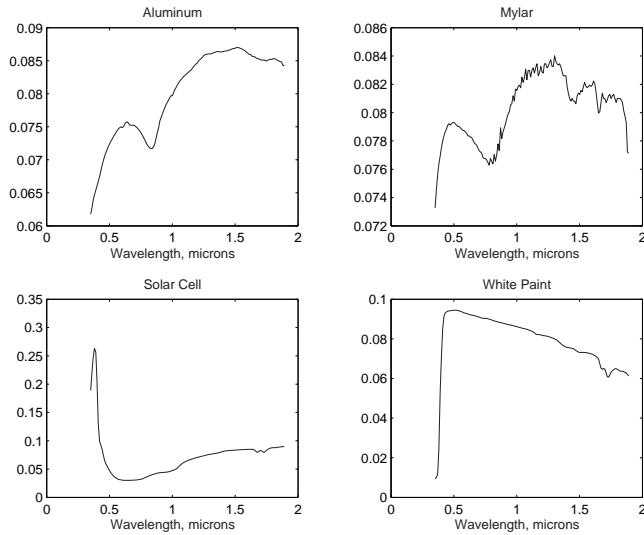


Figure 1: Laboratory spectral traces for the four materials used in data construction: aluminum, mylar, solar cell, and white paint. These spectral traces are in fact averages of several traces of each type of material (see Luu *et al.* [18] for details).

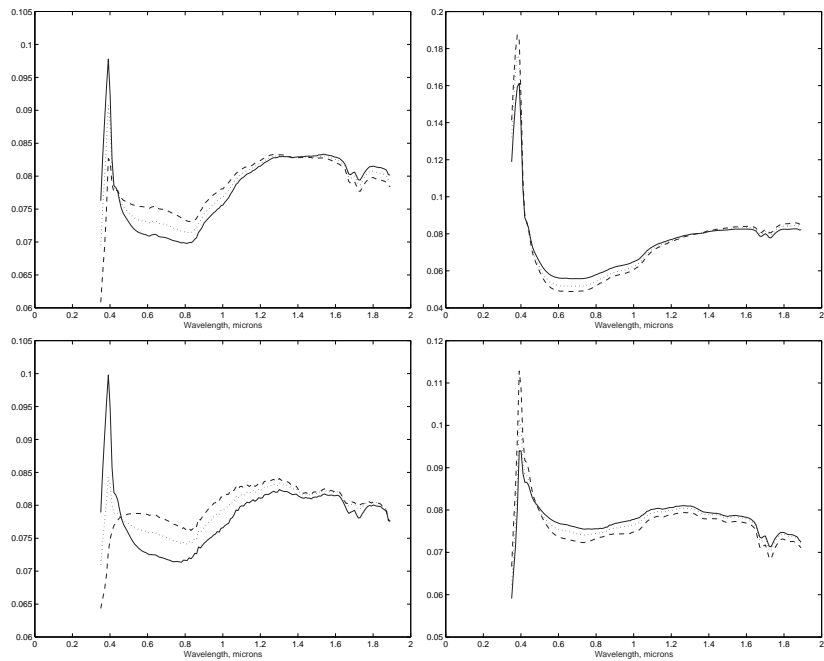


Figure 2: Representative simulated spectral traces for each satellite consisting of mostly: (top-left) 50% aluminum, (top-right) 50% solar cell, (bottom-left) 50% mylar, and (bottom-right) 50% white paint.

6.2 Comparison of Endmember Computation using NMF and CNMF

We applied the sequential approach for computing and selecting endmembers and quantifying fractional abundances, using first **Algorithm NMF** and then **Algorithm CNMF**. In other words, we run each algorithm (NMF and CNMF) 20 times with $k = 6$ on each of the four satellite datasets shown in Figure 2. As a result for each algorithm we obtained a total of 120 endmember candidates for each satellite dataset. In addition we considered a fifth dataset consisting of the aggregation of traces for all four satellites. For CNMF we set the regularization parameters $\alpha = 1$ and $\beta = 0$ to enforce smoothness only in W .

The SKLD was then computed between each endmember and each material spectrum. For each dataset, the endmember with the smallest SKLD was then selected for each of the four materials. The resulting SKLD scores for NMF are shown in Figure 3. The resulting

	Aluminum	Mylar	Solar Cell	White Paint
50% Aluminum	0.0740	0.0609	0.6844	0.1661
50% Mylar	0.0615	0.0606	0.5033	0.1827
50% Solar Cell	0.1681	0.1358	0.0266	0.1916
50% White Paint	0.0882	0.0571	0.8481	0.2845
All Four	0.1137	0.0853	0.0161	0.0346

Figure 3: Confusion matrix for best matched NMF extracted endmembers from simulated satellites with 1% noise and laboratory material spectra. Reported is the SKLD. The best matches are close to 0.

SKLD scores for CNMF are shown in Figure 4. Notice that the SKLD matching scores when the smoothness constraint is employed in CNMF are in general much smaller than those of NMF. In other words, the endmembers computed with CNMF reflect more accurately the “true” spectral signatures shown in Figure 1.

The method of selecting endmembers using the SKLD metric can allow one endmember to be the best match for multiple materials, which is reasonable given the difficulty of resolving some materials using either NMF or CNMF. This can be seen in the similarity of SKLD

	Aluminum	Mylar	Solar Cell	White Paint
50% Aluminum	0.0233	0.0124	0.4659	0.1321
50% Mylar	0.0165	0.0063	0.4009	0.1203
50% Solar Cell	0.0645	0.0292	0.0302	0.2863
50% White Paint	0.0460	0.0125	0.8560	0.1735
All Four	0.0280	0.0659	0.0130	0.0223

Figure 4: Confusion matrix for best matched CNMF extracted endmember candidates from simulated satellites and laboratory material spectra. Noise was 1% and $\alpha = 1$. Reported is the SKLD. Better matches are closer to 0.

scores between the laboratory spectra, particularly aluminum and mylar, shown in Figure 5.

	Aluminum	Mylar	Solar Cell	White Paint
Aluminum	0	0.0209	1.2897	0.3317
Mylar	0.0209	0	1.2719	0.2979
Solar Cell	1.2897	1.2719	0	2.5781
White Paint	0.3317	0.2979	2.5781	0

Figure 5: Confusion matrix for laboratory material spectra with each other. Reported is the SKLD.

Using the confusion matrix in Figure 5, reasonable selection thresholds can be defined for each of the four possible materials in the laboratory dataset L . The minimum SKLD from the confusion matrix in Figure 5 was chosen as the threshold, τ . In the case of aluminum and mylar, the SKLD between these two materials was rejected and the minimum of the remaining values was chosen. This decision has the result of including both materials in the matrix B whenever either is present, which could introduce errors when the fractional abundances are subsequently calculated if one were not present. However, choosing $\tau = 0.0209$ would have resulted in aluminum being excluded from B for the satellite dominated by aluminum, and aluminum and mylar being excluded from B for the dataset composed of all four satellites.

Another method for selecting τ is to use simulated satellites which do not have a given

material on their surfaces. Calculating the best SKLD for a set of CNMF-obtained endmembers also gives a reasonable τ . The same problem exists in selecting τ for aluminum and mylar as with the previous method. This method for choosing τ results in the same endmember selections for our data as the confusion matrix method. In either case, a suitable endmember is found for each of the four materials in D .

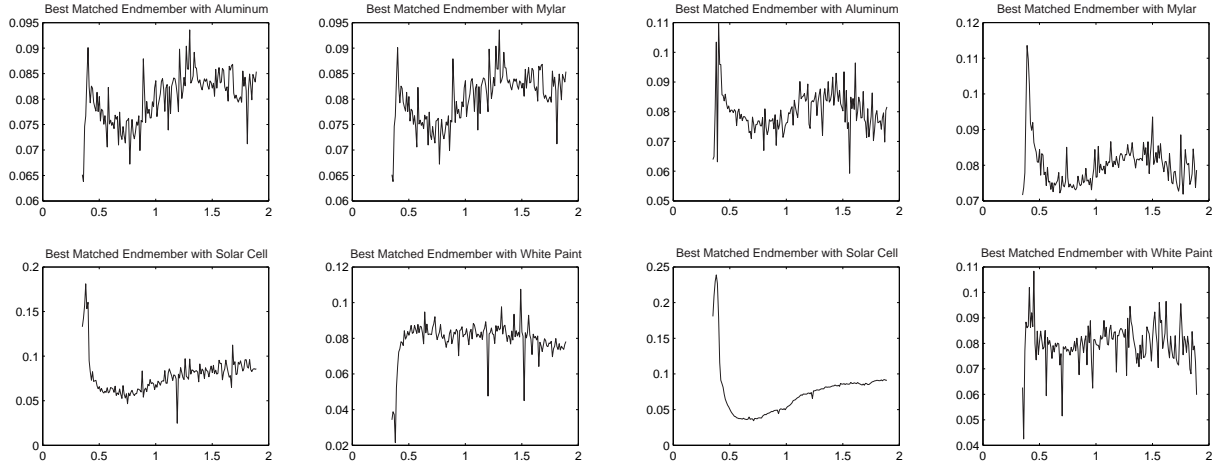
Figure 6 shows the best matched endmembers to each of the materials for three of the five datasets. Both Figures 4 and 6 indicate that the best matched endmembers are typically extracted for the material which dominates the satellite surface composition. The endmembers extracted from the dataset with all four satellites are all similar to the true materials. However, even in this case, CNMF and SKLD are unable to resolve aluminum and mylar.

The fractional abundances of the endmembers were then calculated using iterative approximation as described in Subsection 5. The results are shown in Figure 7. In most cases, our method is effective at estimating both the fractional abundances as well as following the trend over the course of the 100 observations of each satellite as the abundances are allowed to vary. Figure 7(c) shows again the difficulty of resolving aluminum, mylar, and white paint. The calculated fractional abundances for solar cell are quite good, while those for the other materials are less accurate, especially in the satellites dominated by aluminum and mylar.

6.3 Effect of Dataset size on Endmember Selection

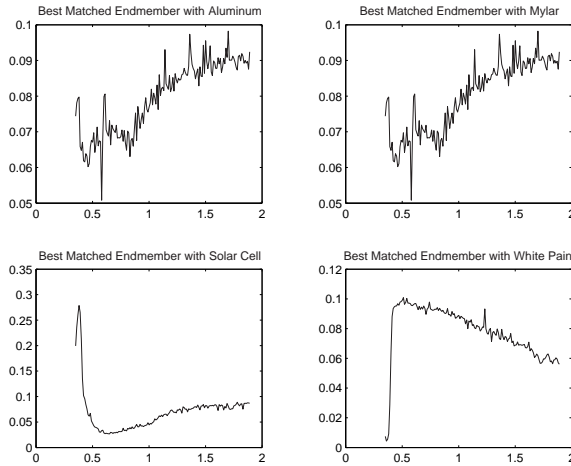
More accurate results are achieved with the dataset of size $n = 400$ composed of the four simulated satellites, when compared to datasets containing only 100 traces of one satellite. This suggests that dataset diversity and size have an impact on the extraction of endmembers using CNMF.

Datasets with 10, 25, 100, and 1000 simulated spectra were constructed using random additive mixtures of the four laboratory spectra shown in Figure 1. CNMF was then executed



(a) Satellite composed of 50% Aluminum, 17% Mylar, 17% Solar Cell, 16% White Paint.

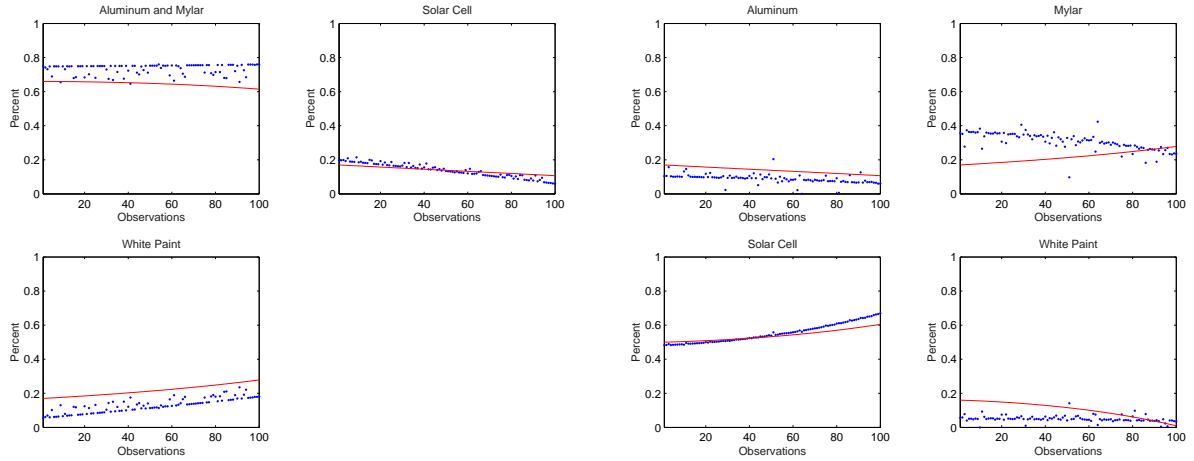
(b) Satellite composed of 17% Aluminum, 16% Mylar, 50% Solar Cell, 17% White Paint.



(c) All four simulated satellites.

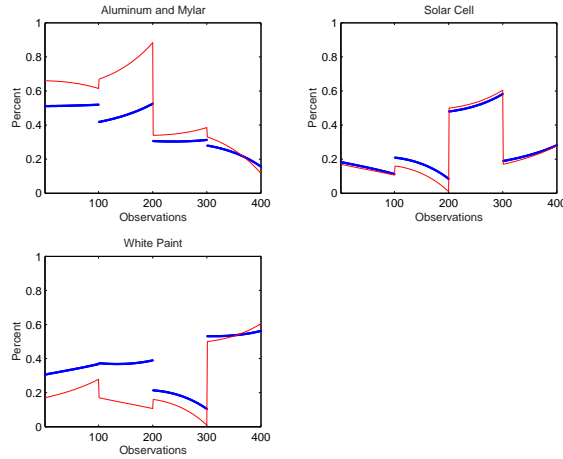
Figure 6: Best endmembers extracted using CNMF with 1% noise and the smoothness penalty term, $\alpha = 1$ with 6 bases executed 20 times. Each dataset consisted of 100 observations for the simulated satellite. The three dimensional nature of each satellite was modeled by allowing each of the four components to vary as defined by a sine wave with random amplitude, frequency, and phase shift. 1% noise was added to the spectra.

ten times on each of these datasets to extract a set of six proposed endmember spectra. These proposed spectra were then compared with each of the four laboratory spectra used in the



(a) Simulated satellite composed of 50% Aluminum, 17% Mylar, 17% Solar Cell, 16% White Paint.

(b) Simulated satellite composed of 17% Aluminum, 16% Mylar, 50% Solar Cell, 17% White Paint.



(c) All four simulated satellites.

Figure 7: Percent compositions obtained using PMRNSD on the basis spectra in Figure 6. True composition is represented in red. The result of the approximation using PMRNSD is in blue.

contruction of the data by calculating the cosine of the angle between the normalized spectra. There is a marked improvement in the matching of endmembers with the true component spectra as the number of simulated observations used in the CNMF algorithm increases, see

Figure 8.

CNMF was more successful at extracting endmembers when the data contained a larger number of spectra. CNMF, when executed on datasets with more distinct spectra, yielded visually smoother endmember spectra with higher matching indices with true endmember spectra. Datasets with fewer than 25 spectra resulted in noisier, more poorly resolved endmembers. More studies need to be made on the effect of dataset size on endmember selection.

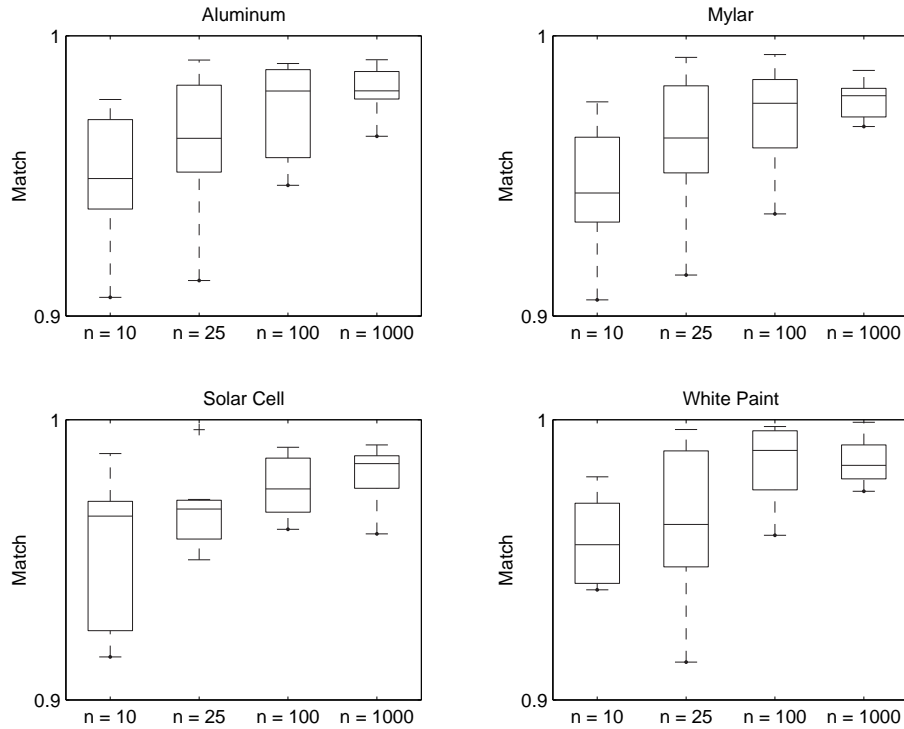


Figure 8: Box-and-whisker plots showing the average best matches between CNMF-extracted endmembers and the true laboratory-obtained component spectra for each of the four present materials. CNMF was executed ten times on each dataset. The size of the dataset is given by n . In each plot, the box represents the range from the lower quartile to the upper quartile, with the median line in between. The “whiskers” are 1.5 times the inter-quartile range.

We have recently received a large amount of real and laboratory space object reflectance data from Dr. Kira Abercromby at the NASA Johnson Space Center [1], which will be used for further numerical studies.

6.4 Spectral Feature Encoding

Other planned work on this ongoing space object identification and classification project includes efforts to develop techniques to more strongly differentiate between certain spectral signatures, especially between aluminum and mylar. These additional methods include spectral feature encoding using wavelets, an approach that has been successful in other spectral unmixing applications [13].

7 Conclusions and Future Work on General NMF

We have attempted to outline some of the major concepts related to nonnegative matrix factorization and to develop a novel and promising application to space object identification and classification. The smoothness constraint used in our CNMF algorithm clearly results in the computation of more accurate endmembers, as demonstrated in Figures 3 and 4.

Several open problem areas for future research remain for the general NMF problem, and we conclude the paper by listing just a few of them.

- Initializing the factors W and H . Methods for choosing, or seeding, the initial matrices W and H for various algorithms (see, e.g., [27]) is a topic in need of further research.
- Uniqueness. Sufficient conditions for uniqueness of solutions to the NNMF problem can be considered in terms of simplicial cones [2], and have been studied in [8]. Algorithms for computing the factors W and V generally produce local minimizers of $f(U, V)$, even when constraints are imposed. It would thus be interesting to apply global optimization algorithms to the NNMF problem.
- Updating the factors. Devising efficient and effective updating methods when columns are added to the data matrix Y in (1) also appears to be a difficult problem and one in need of further research.

Our plans are thus to continue the study of nonnegative matrix factorizations and develop further applications to spectral data analysis. Work on applications to air emission quality [6] and on text mining [23, 26] will also be continued.

References

- [1] K. Abercromby. “Personal Correspondence”, NASA Johnson Space Center, April, 2005.
- [2] A. Berman and R. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*, SIAM Press Classics Series, Philadelphia, 1994.
- [3] M. Catral, L. Han, M. Neumann and R. Plemmons. “Reduced Rank Nonnegative Factorization for Symmetric Nonnegative Matrices”, *Linear Algebra and Its Applications*, Special Issue on Positivity in Linear Algebra, Vol. 393, pp. 107-126, 2004.
- [4] M.-A. Cauquy, M. Roggemann and T. Schultz. “Approaches for Processing Spectral Measurements of Reflected Sunlight for Space Situational Awareness”, *Proc. SPIE Conf. on Defense and Security*, Orlando, FL, 2004.
- [5] C.-I Chang. “An Information Theoretic-based Approach to Spectral Variability, Similarity, and Discriminability for Hyperspectral Image Analysis”, *IEEE Trans. Information Theory*, Vol. 46, pp. 1927-1932, 2000.
- [6] M. T. Chu, F. Diele, R. Plemmons, and S. Ragni, “Optimality, computation, and interpretation of nonnegative matrix factorizations”, preprint, submitted 2004. See <http://www.wfu.edu/~plemmons>
- [7] M. Cooper and J. Foote, “Summarizing Video using Nonnegative Similarity Matrix Factorization”, *Proc. IEEE Workshop on Multimedia Signal Processing* St. Thomas, US Virgin Islands, 2002.

- [8] D. Donoho and V. Stodden. “When does Non-Negative Matrix Factorization Give a Correct Decomposition into Parts?”, preprint, Department of Statistics, Stanford University, 2003.
- [9] D. Guillamet, B. Schiele and J. Vitria. “Analyzing Non-Negative Matrix Factorization for Image Classification”, *16th International Conference on Pattern Recognition (ICPR’02)*, Vol. 2, Quebec City, Canada, 2002.
- [10] D. Guillamet and J. Vitria. “Determining a Suitable Metric when Using Non-Negative Matrix Factorization”, *16th International Conference on Pattern Recognition (ICPR’02)*, Vol. 2, Quebec City, QC, Canada, 2002.
- [11] P. Hoyer. “Non-Negative Sparse Coding”, *Neural Networks for Signal Processing XII (Proc. IEEE Workshop on Neural Networks for Signal Processing)*, Martigny, Switzerland, 2002.
- [12] A. Hyvriinen and P. Hoyer. “Emergence of Phase and Shift Invariant Features by Decomposition of Natural Images into Independent Feature Subspaces”, *Neural Computation*, Vol. 12, pp. 1705-1720, 2000.
- [13] N. Keshava and J. Mustard. “Spectral Unmixing”, *IEEE Signal Processing Magazine*, pp. 44-57, January 2002.
- [14] K. Jorgensen et al. “Squigley Lines and Why They are Important to You”, *NASA Lincoln Space Control Conference*, Lexington, MA, March 2002.
- [15] D. Lee and H. Seung. “Learning the Parts of Objects by Non-Negative Matrix Factorization”, *Nature*, Vol. 401, pp. 788-791, 1999.
- [16] D. Lee and H. Seung. “Algorithms for Non-Negative Matrix Factorization”, *Advances in Neural Processing*, 2000.

- [17] W. Liu and J. Yi. “Existing and New Algorithms for Non-negative Matrix Factorization”, preprint, Computer Sciences Dept., UT Austin, 2003.
- [18] K. Luu, C. Matson, J. Snodgrass, M. Giffin, K. Hamada and J. Lambert. “Object Characterization from Spectral Data”, *Proc. AMOS Technical Conference*, Maui, HI, 2003.
- [19] J. G. Nagy, K. Palmer, and L. Perrone. “Iterative Methods for Image Deblurring: A Matlab Object Oriented Approach”, *Numerical Algorithms*, Vol. 36, No. 1, pp. 73–93, 2004.
- [20] J. G. Nagy and Z. Strakos. “Enforcing Nonnegativity in Image Reconstruction Algorithms”, *Mathematical Modeling, Estimation, and Imaging*, D. C. Wilson, et al., Eds., Vol. 4121, pp. 182–190, 2000.
- [21] V. P. Pauca, R. J. Plemmons, M. Giffin, and K. Hamada. “Mining Scientific Data for Non-Imaging Identification and Classification of Space Objects”, *Proc. AMOS Tech Conf.*, 2004.
- [22] J. Piper, V. P. Pauca, R. J. Plemmons and M. Giffin. “Object Characterization from Spectral Data Using Nonnegative Matrix Factorization and Information Theory”, *Proc. AMOS Tech Conf.*, 2004.
- [23] V. P. Pauca, F. Shahnaz, M. Berry and R. Plemmons. “Text Mining using non-negative Matrix Factorizations”, *Proc. SIAM Inter. Conf. on Data Mining*, Orlando, April, 2004. See <http://www.wfu.edu/~plemmons>.
- [24] A. Plaza, P. Martinez, R. Perez and J. Plaza. “A Quantitative and Comparative Analysis of Endmember Extraction Algorithms from Hyperspectral Data”, *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 42, No. 3, pp. 650-663, 2004.

- [25] S. Prasad, T. Torgersen, P. Pauca, R. Plemmons, and J. van der Gracht. “Restoring Images with Space Variant Blur via Pupil Phase Engineering”, *Optics in Info. Systems, Special Issue on Comp. Imaging*, SPIE Int. Tech. Group Newsletter, Vol. 14, No. 2, pp. 4-5, 2003. See <http://www.wfu.edu/~plemmons>.
- [26] F. Shahnaz, M. Berry, P. Pauca, and R. Plemmons, “Document clustering using non-negative matrix factorization”, to appear in the *Journal on Information Processing and Management*, 2005. See <http://www.wfu.edu/~plemmons>.
- [27] S. Wild, Seeding non-negative matrix factorization with the spherical k-means clustering, M.S. Thesis, University of Colorado, 2002.