Non-negative Matrix Factorization for Hyperspectral Unmixing Using Prior Knowledge of Spectral Signatures

Wei Tang[a], Zhenwei Shi[a,b,c], Zhenyu An[a]

[a]Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, P.R. China

[b]Beijing Key Laboratory of Digital Media, Beihang University, Beijing 100191, P.R. China

[c]State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, P.R. China

Corresponding author: Dr. Zhenwei Shi, Tel.: +86-10-823-16-502; Fax: +86-10-823-38-798. E-mail: shizhenwei@buaa.edu.cn

Abstract

Hyperspectral unmixing is a process aiming at identifying the constituent materials and estimating the corresponding fractions from hyperspectral imagery of a scene. Non-negative matrix factorization (NMF), an effective linear spectral mixture model, has been applied in hyperspectral unmixing during recent years. As the data of hyperspectral imagery analyzed deeper, prior

knowledge of some signatures in the scene could be available. Besides, in several scenes such as mining areas, a few surface substances like copper and iron are easy to identify through field investigation. Thus, their spectral signatures can be used as prior knowledge to unmix hyperspectral data. In such a context, we propose a non-negative matrix factorization based framework for hyperspectral unmixing using such prior knowledge, referred to as NMFupk. Specifically, our algorithm supposes that some spectral signatures in the scene are known and then utilizes the prior knowledge of the spectral signatures to unmix the hyperspectral data. In a series of experiments, we test NMFupk and NMF without prior knowledge on both synthetic and real data. Results achieved demonstrate the efficacy of the proposed algorithm.

Indexing terms:Hyperspectral unmixing, non-negative matrix factorization (NMF), spectral signatures, prior knowledge.

1 Introduction

Hyperspectral remote sensing measures radiance of earth's surface materials at hundreds of narrow and contiguous wavelength bands. Most recently, hyperspectral sensors have been evolved to collect spectra extending from visible region through the infrared band. The major application categories of hyperspectral sensing include image fusion, anomaly detection, target recognition, and background characterization.[1–5] Frequently, however, due to the low spatial resolution of a sensor as well as the combination of distinct materials into a homogeneous mixture,[6] a single pixel in hyperspectral imagery is a mixture of several distinct substances. A great challenge is to unmix the given mixed pixels, or in other words, decompose the measured spectra of mixed pixels into a collection of constituent spectra (endmembers) and a set of corresponding fractions (abundances).

For the past few years, linear spectral unmixing model has been widely used to solve the problem.[6–8] It considers mixed pixel a linear combination of endmembers weighted by their corresponding abundance fractions. This linear model has practical advantages such as ease of implementations and flexibility in different applications.[9] Under this model, hyperspectral unmixing takes three main procedures, namely dimension reduction, endmember extraction, and inversion.[6] Dimension reduction, an optional step, intends to reduce the dimension of the data in the scene. Endmember extraction aims at estimating the set of distinct endmembers mixed in the pixels. A group of endmember extraction algorithms have been proposed. Some algorithms such as the pixel purity index (PPI),[10] N-FINDR,[11] and the vertex component anal-

ysis (VCA)[12,13] assume the presence of at least one pure pixel of each endmember in the data. So rigorous, however, the prerequisite could not be satisfied in some datasets. Other extraction algorithms like the minimum-volume enclosing simplex (MVES),[14] the minimum volume simplex analysis (MVSA)[15] and the iterated constrained endmembers (ICE)[16] are based on minimum volume. One main drawback of this kind of algorithm is the high computation cost, particularly as the number of endmembers increases.[17] Inversion, the last procedure, is oriented to estimating the fractional abundances of each mixed pixel from its spectrum and the endmember spectra.[6] Mature algorithms on inversion include least squares methods and minimum variance methods, etc.

In consideration of the non-negativity of both spectra and abundances, another method, non-negative matrix factorization (NMF),[18] has been applied in hyperspectral unmixing.[19] However, NMF, which decomposes the data into two non-negative matrices, could lead to non-unique solutions for the existence of local minima caused by the non-convexity of the objective function.[20] Even though, taking VCA as an initialization method for NMF, the performance can be improved dramatically, and the endmembers can be well extracted.

With the study of hyperspectral data progressing, prior knowledge of some signatures in the scene could be available. Besides, in several scenes such as mining areas, some materials like copper and iron are easy to identify through field investigation. Thus, their spectral signatures can be obtained from the spectral library. Such prior knowledge of endmembers has been used for classification and change detection of hyperspectral image.[21–23] In this paper, we propose an NMF based framework (N-

MFupk) to incorporate prior knowledge into hyperspectral data unmixing. In the proposed algorithm, we assume that one or two spectral signatures in the scene are known. NMFupk then makes use of the prior knowledge to unmix the hyperspectral data. Numerical experiments on both synthetic and real data confirm the efficacy of NMFupk.

The rest of the paper is organized as follows: in Section 2, we review NMF. The NMFupk algorithm is derived in Section 3. Experimental results appear in Section 4. Finally, we conclude in Section 5.

2 NMF for Hyperspectral Unmixing

In this section, we first briefly introduce the linear mixing model, which is the most widely used model in spectral unmixing. Secondly, NMF, which is the foundation of the proposed algorithm, is reviewed.

2.1 Linear Mixing Model

Spectral unmixing aims at identifying the constituent materials and estimating the corresponding fractions in each mixed pixel. The linear mixing model assumes that the observed spectrum of a mixed pixel can be expressed as a linear combination of the spectra of the endmembers weighted by their corresponding abundance fractions. Suppose $\mathbf{V} \in R^{L \times K}$ is the observed hyperspectral data matrix, where L is the number of spectral bands and K represents the number of pixels. Each column of matrix $\mathbf{V}$ denotes a material spectrum vector with L bands. $\mathbf{W} \in R^{L \times P}$ is the spectral signature

matrix, with each column corresponding to an endmember spectrum, and P is the number of endmembers in the scene. $\mathbf{H} \in R^{P\times K}$ indicates the abundance matrix. Each column of matrix $\mathbf{H}$, denoted by $\mathbf{h}_i$, is the fraction of the ith pixel. Then the linear mixing model for the pixel with coordinate (i, j) can be written as

$$\mathbf{V}_{ij} = \sum_{a=1}^{P} \mathbf{W}_{ia}\mathbf{H}_{aj} + \mathbf{N}_{ij} \tag{1}$$

where $\mathbf{N} \in R^{L\times K}$ signifies the additive observation noise. (1) can be rewritten in the matrix form as:

$$\mathbf{V} = \mathbf{WH} + \mathbf{N} \tag{2}$$

Since the fraction of each endmember for a mixed pixel cannot be negative and their sum should be one, the model has the following two constrains:

$$\sum_{p=1}^{P} \mathbf{H}_{pk} = 1 \tag{3}$$

$$\mathbf{H}_{ij} \geq 0, \forall i, j \tag{4}$$

which are called sum-to-one and non-negativity, respectively.

2.2 NMF

Typically, the NMF problem follows as:[18] given a non-negative matrix $\mathbf{V}$, find reduced rank non-negative matrices $\mathbf{W}$ and $\mathbf{H}$ to satisfy:

$$\mathbf{V} \approx \mathbf{WH}. \tag{5}$$

Here, $\mathbf{V} \in R^{L\times K}$ is a data matrix in which each observation of an object is stored as a column. $\mathbf{W} \in R^{L\times P}$ is often regarded as the source matrix, with each column

corresponding to an endmember. $\mathbf{H} \in R^{P \times K}$ is the abundance matrix associated with the data in $\mathbf{V}$.[19] We can regard $\mathbf{W}$ as a basis optimized for the linear approximation of the data matrix $\mathbf{V}$. As few basis vectors can represent many data vectors, good approximation can be obtained only if the basis vectors discover the latent structure in the data.[20] As with other methods of finding components such as PCA and ICA,[24] the created basis of NMF could have no physical meaning. However, in contrast to ICA and PCA, the non-negativity constrains make the representation purely additive (allowing no subtraction), which meets with the natural of both spectra and their fractional abundances. Besides, NMF is suitable for dealing with high-dimensional data and the dimension of hyperspectral data is very high. In this paper, we also use VCA for initialization, then NMF becomes more inclined to get the expected part-based representation of the data. Hence, NMF can be used to unmix hyperspectral data.

To unmix hyperspectral data, NMF can be performed by minimization of the following objective function which is based on Euclidean distance.[25]

$$F(\mathbf{W}, \mathbf{H}) = \frac{1}{2} \sum_i \sum_j \left(\mathbf{V}_{ij} - (\mathbf{WH})_{ij}\right)^2 = \frac{1}{2} \left\|\mathbf{V} - \mathbf{WH}\right\|^2 \tag{6}$$

Lee and Seung[18] proposed an NMF algorithm grounded on multiplicative update rules of $\mathbf{W}$ and $\mathbf{H}$. As mentioned earlier, for the existence of local minima caused by the non-convexity of the objective function in (6), NMF could lead to non-unique solutions. Hence, we can take VCA[12,13] as an initialization method to improve the performance of hyperspectral unmixing using NMF. The overall process is summarized in Algorithm 1.

Algorithm 1: Outline of hyperspectral unmixing using NMF with VCA initialization

Input: $L \times K$ matrix $\mathbf{V} \geq 0$, endmember number P

Initialize matrix $\mathbf{W}$ via VCA.

$\mathbf{H} \Leftarrow \mathbf{W}^{\dagger}\mathbf{V}$.

Set $\mathbf{W}$ and $\mathbf{H}$ positive.

Compute $F_{new} = \frac{1}{2}\|\mathbf{V} - \mathbf{WH}\|^2$.

repeat

  $F_{old} \Leftarrow F_{new}$.

  $\mathbf{W} \Leftarrow \mathbf{W}.*(\mathbf{VH}^T)./(\mathbf{WHH}^T + \varepsilon)$,

  $\mathbf{H} \Leftarrow \mathbf{H}.*(\mathbf{W}^T\mathbf{V})./(\mathbf{W}^T\mathbf{WH} + \varepsilon)$.

  $\mathbf{h}_i \Leftarrow \mathbf{h}_i/(\mathbf{1}_P^T\mathbf{h}_i), i = 1, 2 \ldots K$.

  Compute $F_{new} = \frac{1}{2}\|\mathbf{V} - \mathbf{WH}\|^2$.

until the maximum number of iterations has been reached or $\frac{|F_{old} - F_{new}|}{F_{new}} \leq \tau$.

Output: $\mathbf{W} \in R^{L \times P}$ and $\mathbf{H} \in R^{P \times K}$.

Here, .* and ./ denote elementwise multiplication and division, respectively. $\mathbf{W}^{\dagger}$ means pseudo inverse of $\mathbf{W}$. $\mathbf{h}_i$ denotes the ith column of matrix $\mathbf{H}$. $\mathbf{1}_P$ is P-vector with all the components equal to unity. In the process, a small positive quantity ε should be added to the denominators during the update of $\mathbf{W}$ and $\mathbf{H}$ at each iteration step. τ is a preset threshold. Clearly, given a non-negative initialization, the rules insure matrices non-negative.

3 NMF for hyperspectral unmixing using prior knowledge (NMFupk)

NMFupk is developed for hyperspectral unmixing with prior knowledge of spectral signatures in the scene. As we noted above, NMFupk assumes that some spectral signatures in the scene are known. NMFupk uses a decomposition model which explicitly distinguishes the known part and the unknown part of a spectral signature matrix as follows,

$$\begin{aligned}\mathbf{V} &= \mathbf{W}\mathbf{H} \\ &= \left[\mathbf{W}_1, \mathbf{W}_2\right] \begin{bmatrix}\mathbf{H}_1 \\ \\ \mathbf{H}_2\end{bmatrix} \\ &= \mathbf{W}_1\mathbf{H}_1 + \mathbf{W}_2\mathbf{H}_2 \end{aligned} \tag{7}$$

where $\mathbf{W}_1$ indicates the known signatures with $\mathbf{H}_1$ the corresponding abundance fractions, $\mathbf{W}_2$ and $\mathbf{H}_2$ represent the unknown endmembers and related abundance fractions, respectively. Figure 1 illustrates the NMFupk model. The most straight-

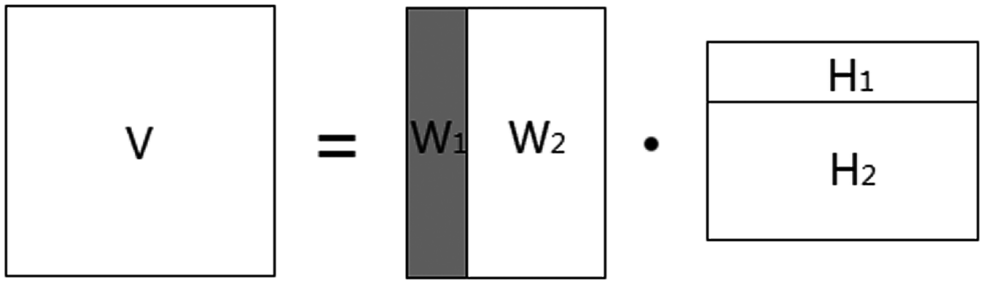

Figure 1: An illustration of NMFupk model. Spectral signature matrix $\mathbf{W}$ is treated as two parts. $\mathbf{W}_1$ is the known part with $\mathbf{H}_1$ the corresponding abundance fractions. $\mathbf{W}_2$ and $\mathbf{H}_2$ represent the unknown signatures and related abundance fractions, respectively.

forward way to use the additional knowledge is to replace the first part of $\mathbf{W}$ with the known signatures after initialization (in this paper, we use VCA for initialization), then use them in the NMFupk model as initial values. However, it could unexpectedly degrade the VCA initialization because the first few endmembers randomly extracted by VCA could not correspond with the known signatures in a good chance. Hence, by direct substitution, the valuable endmembers will decline in the number as well as get repetitive.

Instead of the simple initialization, we put forward a matching method which searches the known-signature-matched endmembers in $\mathbf{W}$ after VCA initialization and then supersedes them by the known ones. Specifically, if the angle between an extracted endmember and a known signature spectrum is less than a given threshold, then we consider them a matched pair. To keep pace with the NMFupk model, we also realign $\mathbf{W}$ via putting the known endmembers leftmost. Algorithm 2 embodies

Algorithm 2: Initialization of NMFupk

Input: $L \times K$ matrix $\mathbf{V} \geq 0$, endmember number P, and known-signature matrix $\mathbf{W}_1$

while not all known endmembers have been matched or iter $<$ *maxiter* **do**

  Initialize $\mathbf{W}$ via VCA.

  Search matched endmembers in $\mathbf{W}$.

end

Replace the matched endmembers in $\mathbf{W}$ by the known signatures and put them leftmost.

$\mathbf{H} \Leftarrow \mathbf{W}^{\dagger}\mathbf{V}$.

Set $\mathbf{W}$ and $\mathbf{H}$ positive.

Output: $\mathbf{W} \in R^{L \times P}$ and $\mathbf{H} \in R^{P \times K}$.

the process for the initialization of NMFupk. *maxiter* limits the number of iterations for the matching procedure.

Considering (7), the objective function in (6) can be transformed as follows,

$$F = \frac{1}{2} \left\| \mathbf{V} - \mathbf{W}_1\mathbf{H}_1 - \mathbf{W}_2\mathbf{H}_2 \right\|^2 . \tag{8}$$

If we can separate the gradient of the objective function as below,

$$\frac{\partial \mathbf{F}}{\partial \mathbf{W}} = \left[\frac{\partial \mathbf{F}}{\partial \mathbf{W}}\right]^{+} - \left[\frac{\partial \mathbf{F}}{\partial \mathbf{W}}\right]^{-} \tag{9}$$

where $\left[\frac{\partial \mathbf{F}}{\partial \mathbf{W}}\right]^{+} > 0$ and $\left[\frac{\partial \mathbf{F}}{\partial \mathbf{W}}\right]^{-} > 0$ denote the positive and negative parts of $\frac{\partial \mathbf{F}}{\partial \mathbf{W}}$ respectively. Then we can construct the multiplicative update factor by taking negative terms of the partial derivative as its numerator, while taking positive ones as its denominator:[26,27]

$$\mathbf{W} \Leftarrow \mathbf{W}. * \left[\frac{\partial \mathbf{F}}{\partial \mathbf{W}}\right]^{-} ./ \left[\frac{\partial \mathbf{F}}{\partial \mathbf{W}}\right]^{+} . \tag{10}$$

Obviously, $\mathbf{W}$ remains non-negative during the update.

Then, by partially differentiating the objective function (8) with respect to each matrix to be updated:

$$\frac{\partial \mathbf{F}}{\partial \mathbf{W}_2} = -\mathbf{V}\mathbf{H}_2^T + \mathbf{W}_2\mathbf{H}_2\mathbf{H}_2^T + \mathbf{W}_1\mathbf{H}_1\mathbf{H}_2^T \tag{11}$$

$$\frac{\partial \mathbf{F}}{\partial \mathbf{H}_1} = -\mathbf{W}_1^T\mathbf{V} + \mathbf{W}_1^T\mathbf{W}_1\mathbf{H}_1 + \mathbf{W}_1^T\mathbf{W}_2\mathbf{H}_2 \tag{12}$$

$$\frac{\partial \mathbf{F}}{\partial \mathbf{H}_2} = -\mathbf{W}_2^T\mathbf{V} + \mathbf{W}_2^T\mathbf{W}_2\mathbf{H}_2 + \mathbf{W}_2^T\mathbf{W}_1\mathbf{H}_1 \tag{13}$$

we can get the following multiplicative rules for NMFupk.

Theorem 1: The objective function (8) is nonincreasing under the update rules

$$\mathbf{W}_2 \Leftarrow \mathbf{W}_2 .* (\mathbf{V}\mathbf{H}_2^T)./(\mathbf{W}_2\mathbf{H}_2\mathbf{H}_2^T + \mathbf{W}_1\mathbf{H}_1\mathbf{H}_2^T) \tag{14}$$

$$\mathbf{H}_1 \Leftarrow \mathbf{H}_1 .* (\mathbf{W}_1^T\mathbf{V})./(\mathbf{W}_1^T\mathbf{W}_1\mathbf{H}_1 + \mathbf{W}_1^T\mathbf{W}_2\mathbf{H}_2) \tag{15}$$

$$\mathbf{H}_2 \Leftarrow \mathbf{H}_2 .* (\mathbf{W}_2^T\mathbf{V})./(\mathbf{W}_2^T\mathbf{W}_2\mathbf{H}_2 + \mathbf{W}_2^T\mathbf{W}_1\mathbf{H}_1). \tag{16}$$

The proof of the theorem can be found in the Appendix.

Fixing the known matrix $\mathbf{W}_1$ and updating the unknown ones iteratively, we can obtain all the endmembers and their corresponding abundance fractions finally. Based on the update rules, the algorithm for the hyperspectral unmixing using NMFupk is summarized in Algorithm 3.

4 Experiments

In this section, we test the proposed NMFupk and NMF without prior knowledge on both synthetic and real data. Through the experiments, we can verify the validity of the NMFupk model for hyperspectral unmixing as well as test whether prior

Algorithm 3: Outline of hyperspectral unmixing using NMFupk

Input: $L \times K$ matrix $\mathbf{V} \geq 0$, endmember number P, and known-signature matrix $\mathbf{W}_1$

Initialize $\mathbf{W}$ and $\mathbf{H}$ via **Algorithm 2**

$\mathbf{h}_i \Leftarrow \mathbf{h}_i / (\mathbf{1}_P^T \mathbf{h}_i), i = 1, 2 \ldots K.$

Compute $F_{new} = \frac{1}{2} \left\| \mathbf{V} - \mathbf{W}_1 \mathbf{H}_1 - \mathbf{W}_2 \mathbf{H}_2 \right\|^2$.

repeat

  $F_{old} \Leftarrow F_{new}$.

  Update $\mathbf{W}_2$, $\mathbf{H}_1$ and $\mathbf{H}_2$ using (14) - (16).

  $\mathbf{h}_i \Leftarrow \mathbf{h}_i / (\mathbf{1}_P^T \mathbf{h}_i), i = 1, 2 \ldots K.$

  Compute $F_{new} = \frac{1}{2} \left\| \mathbf{V} - \mathbf{W}_1 \mathbf{H}_1 - \mathbf{W}_2 \mathbf{H}_2 \right\|^2$.

until the maximum number of iterations has been reached or $\frac{|F_{old} - F_{new}|}{F_{new}} \leq \tau$.

Output: $\mathbf{W} \in R^{L \times P}$ and $\mathbf{H} \in R^{P \times K}$.

knowledge contributes to a better performance. One or two signatures in NMFupk are given known, so we only compare the remained unknown endmembers between the two algorithms.

Three distance metrics are considered to evaluate the performances of the two algorithms. One is the spectral information divergence (SID),[28] which is frequently used to measure the quasi-distance or directed difference between the pth true endmember signature $\mathbf{W}_p$ and its estimate $\hat{\mathbf{W}}_p$. It is defined as:

$$\mathrm{SID}_p = D(\mathbf{W}_p||\hat{\mathbf{W}}_p) + D(\hat{\mathbf{W}}_p||\mathbf{W}_p). \tag{17}$$

Here $D(\mathbf{W}_p||\hat{\mathbf{W}}_p)$ is the relative entropy of $\mathbf{W}_p$ with respect to $\hat{\mathbf{W}}_p$ given by

$$D(\mathbf{W}_p||\hat{\mathbf{W}}_p) = \sum_{l=1}^{L} p_l \log(\frac{p_l}{q_l}) \tag{18}$$

where $p_l = \mathbf{W}_{lp} / \sum_{j=1}^{L} \mathbf{W}_{jp}$, $q_l = \hat{\mathbf{W}}_{lp} / \sum_{j=1}^{L} \hat{\mathbf{W}}_{jp}$.

The second metric is the L_∞ norm, it measures the maximum difference between corresponding spectral components. For a vector $\mathbf{x} \in R^L$, the L_∞ norm is defined as:

$$\|\mathbf{x}\|_\infty = \max_{1 \le i \le L} |\mathbf{x}_i| \tag{19}$$

where $\mathbf{x}_i$ is the ith element of $\mathbf{x}$. If two spectra have low difference on average, but they differ significantly at a specific wavelength, where one spectrum has a spike and the other does not, then they are really two different spectra. This is specially important for the identification of some specific material whose spectrum is characterized by peaks in certain frequencies.

The third metric is the root mean square error (RMSE):[29]

$$\mathrm{RMSE}_p = \sqrt{\frac{1}{K} \sum_{k=1}^{K} (\mathbf{H}_{pk} - \hat{\mathbf{H}}_{pk})^2}. \tag{20}$$

Here, $\mathbf{H}$ denotes the true abundances and $\hat{\mathbf{H}}$ represents the estimated ones. RMSE is commonly used to measure the similarity of two abundances.

Mean values of all the metrics, signified as *sid*, l_∞ and *rmse* respectively, will be computed to assess the performances of the algorithms. And, generally speaking, the smaller *sid*, l_∞ or *rmse* is, the more the estimation approximates the truth. Meanwhile, the standard deviations of all the metrics, denoted by σ_{SID}, σ_{L_∞}, σ_{RMSE} respectively, are reported. We also adopt the Monte Carlo method in our experiments. Each algorithm is run 50 times, then the median values and standard deviations are computed.

In all the experiments, the preset threshold τ and maximum number of iterations for Algorithm 1 and Algorithm 3 are set to be 0.001 and 200, respectively. The *maxiter* for Algorithm 2 is set to be 20. We set the threshold of angle in Algorithm 2 by comparing the endmember of a known material and the endmember of the same material obtained by VCA, then the angle is set to be (a little larger than) the angle between these two spectral signatures. Specifically, in the synthetic data experiment, it is set to be 10 degrees while in the real data experiment, it is set to be 20 degrees.

4.1 Evaluation with Synthetic Data

Data Creation: we choose five mineral spectral signatures (shown in Figure 2) from United States Geological Survey (USGS)[30] digital spectral library to simulate real endmember spectral signatures. The synthetic data are created as follows:

(1) Divide the scene, whose size is $z^2 \times z^2$ ($z = 8$), into $z \times z$ regions. Initialize

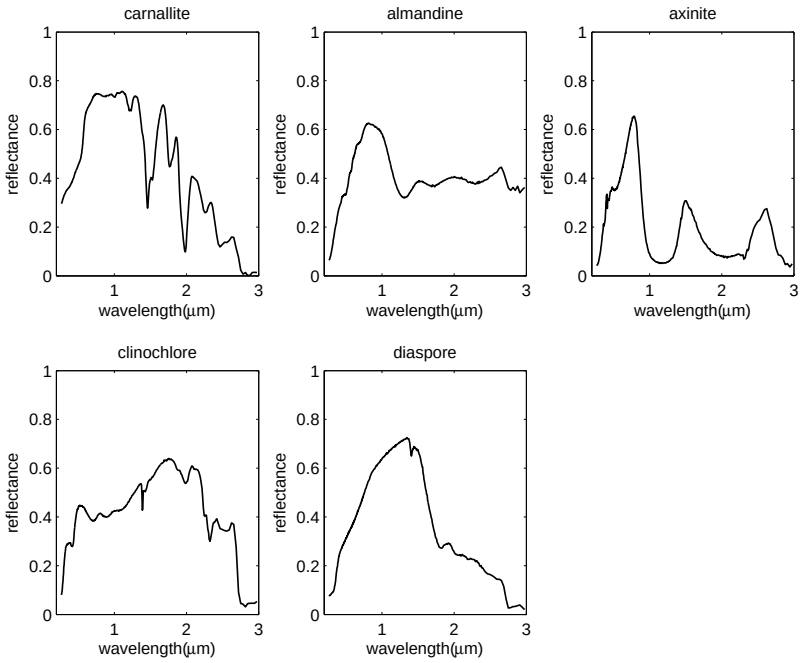

Figure 2: Selected spectral signatures from USGS. The title of each subimage denotes the mineral corresponding to the signature.

each region with the same type of ground cover, randomly selected from the endmember class. The number of endmembers is 5 ($P = 5$).

(2) Generate mixed pixels through a simple $(z + 1) \times (z + 1)$ spatial low-pass filter.

(3) Replace all the pixels in which the abundance of a single endmember is larger than 70% with a mixture made up of only two endmembers (the abundances of the two endmembers both equal 50%) so as to further remove pure pixels and represent the sparseness of abundances at the same time; After these three steps, we obtain the distribution of five endmembers in the scene and specific values are stored in $\mathbf{H}$ with a size of of $P \times K$ ($K = z^2 \times z^2$).

(4) From the USGS digital spectral library, we selected 470 wave bands ($L = 470$).

The size of spectral signatures $\mathbf{W}$ is $L \times P$.

(5) Use linear spectral mixing model $\mathbf{V} = \mathbf{W} \times \mathbf{H}$ to generate hyperspectral data, add white Gaussian noise at the same time. The signal to noise ratio (SNR) of the noise is 20 db. The size of hyperspectral data $\mathbf{V}$ is $L \times K$.

As mentioned earlier, we run NMFupk with one or two signatures known, both of which are selected randomly. For simplicity, we indicate the two scenarios by NMFupk(1) and NMFupk(2). The numbers in the parentheses count the given endmembers. Axinite is given as the known signature in NMFupk(1), while Axinite and Clinochlor are given in NMFupk(2). We calculate *sid* and l_∞ only of the unknown signatures for both algorithms and *rmse* of all abundances.

We can conclude from Table 1 and Table 2 that NMFupk behaves better than NMF without prior knowledge on *sid*, l_∞ and *rmse* in both scenarios. Comparing *rmse*s in the two tables, we find that the more signatures are given, the more estimation of abundances approximates the truth. Meanwhile considering the standard deviations of all the metrics, NMFupk also has a better performance than NMF in both scenarios. Prior knowledge of some endmembers in the scene does contribute to better extraction of the other endmembers and estimation of abundances. Figure 3 shows comparison of the extracted endmembers among different algorithms on synthetic data and Figure 4 shows their corresponding estimated abundance maps as well as the true ones.

Table 1: Comparison between NMF without prior knowledge and NMFupk(1) on synthetic data

algorithm	*sid*	l_∞	*rmse*	σ_{SID}	σ_{L_∞}	σ_{RMSE}
NMF	0.0160	0.1383	0.1182	0.1073	0.0558	0.0377
NMFupk(1)	0.0132	0.1319	0.1060	0.0808	0.0452	0.0344

Table 2: Comparison between NMF without prior knowledge and NMFupk(2) on synthetic data

algorithm	*sid*	l_∞	*rmse*	σ_{SID}	σ_{L_∞}	σ_{RMSE}
NMF	0.0242	0.1282	0.1169	0.1556	0.0516	0.0296
NMFupk(2)	0.0180	0.1239	0.0961	0.0781	0.0349	0.0273

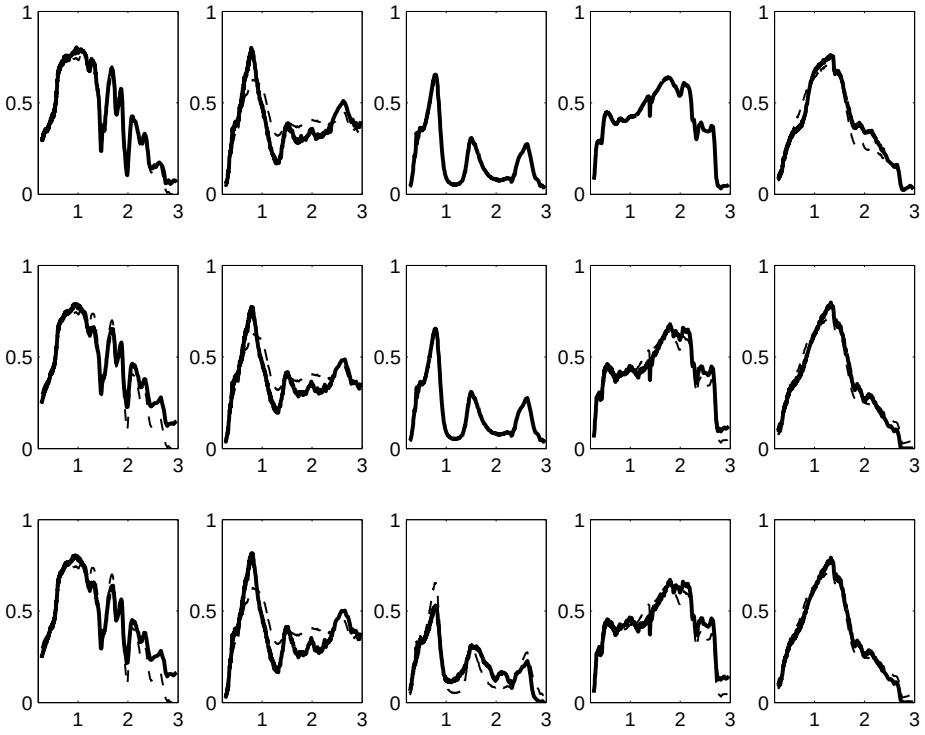

Figure 3: Comparison of the extracted endmembers among different algorithms on synthetic data. Spectral signatures are estimated using NMFupk(2) ($sid = 0.0187$), NMFupk(1) ($sid = 0.0206$), NMF without prior knowledge ($sid = 0.0262$) from top row to bottom row, respectively. Each column, from left to right, corresponds to the same endmember: Carnallite, Almandine, Axinite, Clinochlore and Diaspore. The thin lines are estimated spectral signatures, the thick lines are standard spectral signatures. The x-coordinate stands for wavelength (μm), and y-coordinate represents reflectance.

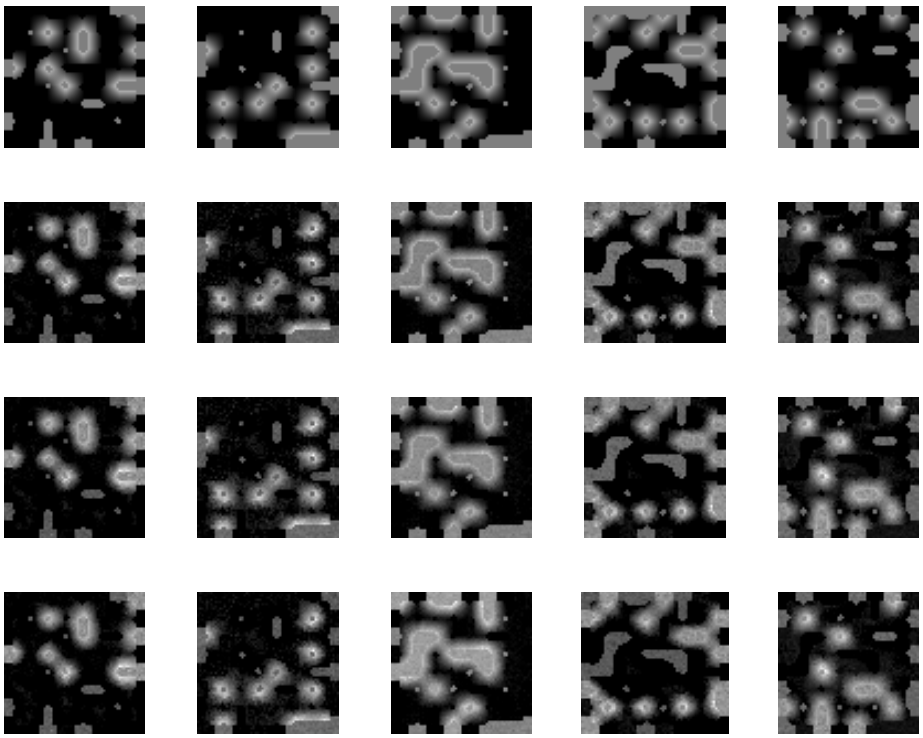

Figure 4: Comparison of abundance maps on synthetic data. True abundance maps, abundance maps obtained by NMFupk(2) ($rmse = 0.0616$), NMFupk(1) ($rmse = 0.0681$), NMF without prior knowledge ($rmse = 0.0739$) from top row to bottom row, respectively.

4.2 Evaluation with Real Data

In this section, we evaluate the performance of the proposed NMFupk algorithm using the AVIRIS data taken over the Cuprite Nevada site.[31] In our experiment, a 188×191 subimage of the hyperspectral data is considered as our region of interest, with 224 spectral bands. Due to strong noise or dense water vapor content, the bands $1-2$, $104-113$, $148-167$ and $221-224$ are removed, leaving a total of 188 spectral bands. As true total number of endmembers and the associated minerals are yet to be accurately identified, we can only refer to previous works[12,13,32–35] done by others to preset our known signatures and endmember number. And there is no way to calculate *rmse* for a lack of true abundance maps. Finally, we choose Montmorillonite and Sphene for NMFupk(2) and Sphene for NMFupk(1). They can be extracted by VCA.[13] The signatures of these two minerals can be obtained from the USGS library. In order to estimate the number of endmembers present in the processed area, we resort to the VCA, proposed in,[13] and take $P = 8$.

In Table 3, for both algorithms, we calculate *sid* and l_∞ on average of all the extracted endmembers except Sphene. Likewise, we calculate *sid* and l_∞ with Montmorillonite and Sphene excluded in Table 4. Obviously, NMFupk outperforms NMF without prior knowledge in both scenarios. Besides, considering the standard deviations of errors, we conclude that NMFupk has a more predictable behavior than NMF without prior knowledge. Figure 5, Figure 6 and Figure 7 show spectral signatures estimated by NMFupk(2), NMFupk(1) and NMF without prior knowledge, respectively. Comparing the three figures, we find that endmembers extracted by the

Table 3: Comparison between NMF without prior knowledge and NMFupk(1) on real data

algorithm	*sid*	l_∞	σ_{SID}	σ_{L_∞}
NMF	0.0284	0.1853	0.0131	0.0522
NMFupk(1)	0.0219	0.1817	0.0077	0.0475

Table 4: Comparison between NMF without prior knowledge and NMFupk(2) on real data

algorithm	*sid*	l_∞	σ_{SID}	σ_{L_∞}
NMF	0.0302	0.2094	0.0118	0.0506
NMFupk(2)	0.0222	0.1683	0.0049	0.0369

algorithms are different. Thus, we refer to the mineral map[23,36] produced by USGS in 1995. The Tricorder 3.3 software product was used to map different minerals in the Cuprite mining district. Among the spectral signatures extracted by NMFupk(2), two minerals, namely Dry Long Grass and Quartz, do not appear in the map. As for the endmembers extracted by NMFupk(1), three minerals cannot be found in the map. They are Dry Long Grass, Ammonio-Smectite and Microcline. By contrast, NMF without prior knowledge behaves worst, with only three minerals (Kaolin/Smect #1, Kaolin/Smect #3 and Alunite) appearing in the map. Hence, we conclude that NMFupk can extract endmembers more accurate than NMF without prior knowledge and the more prior knowledge is given, the better result is obtained.

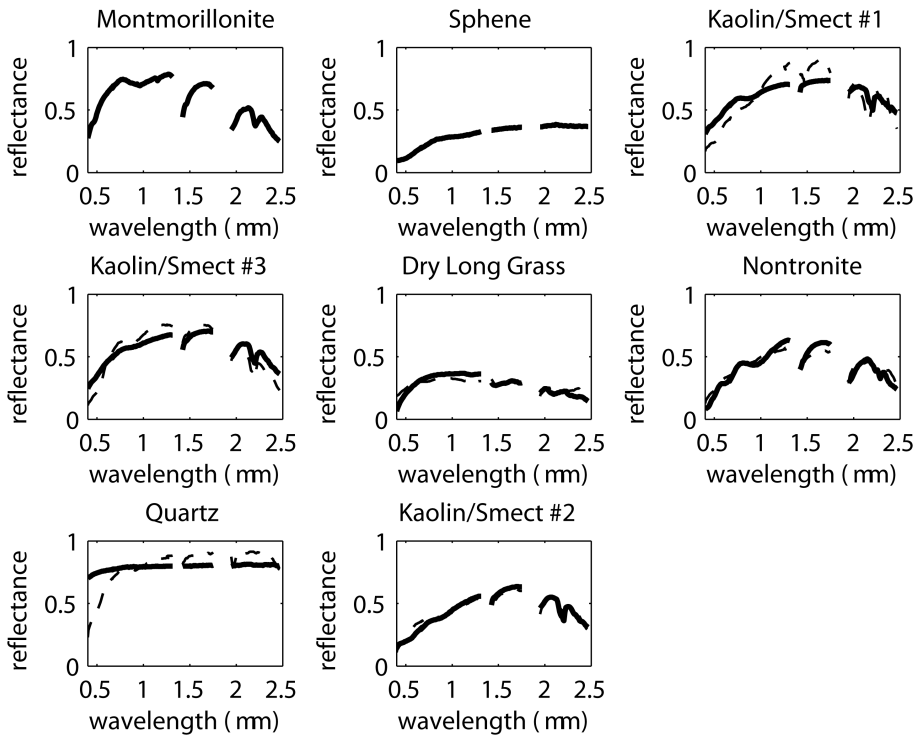

Figure 5: Spectral signatures estimated by NMFupk(2) ($sid = 0.0203$). The spectral signatures of Montmorillonite and Sphene are given known. The title of each subimage denotes the mineral corresponding to the signature. The thin lines are estimated spectral signatures, the thick lines are standard spectral signatures.

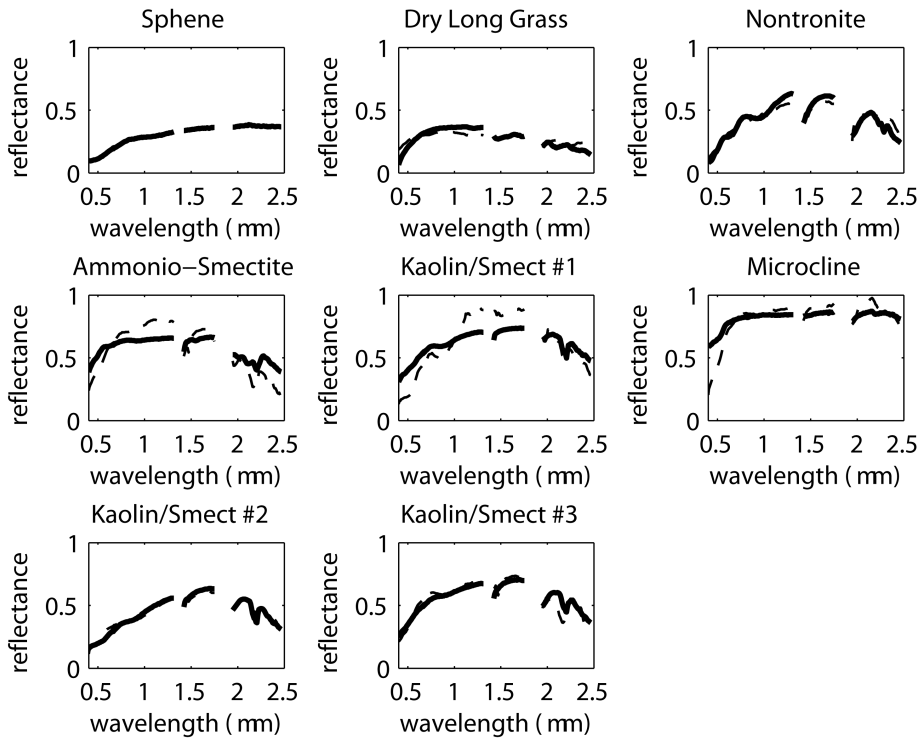

Figure 6: Spectral signatures estimated by NMFupk(1) ($sid = 0.0225$). The spectral signature of Sphene is given known. The title of each subimage denotes the mineral corresponding to the signature. The thin lines are estimated spectral signatures, the thick lines are standard spectral signatures.

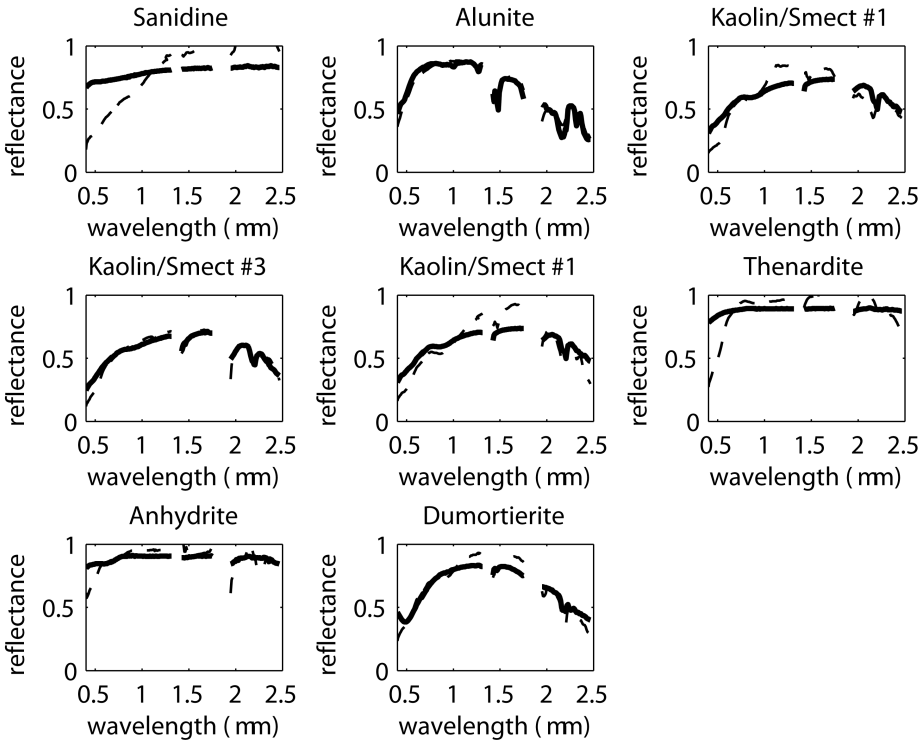

Figure 7: Spectral signatures estimated by NMF without prior knowledge (sid = 0.0338). The title of each subimage denotes the mineral corresponding to the signature. The thin lines are estimated spectral signatures, the thick lines are standard spectral signatures.

5 Conclusion

In this paper, we present a novel non-negative matrix factorization based framework for hyperspectral unmixing using prior knowledge of spectral signatures (NMFupk). The NMFupk algorithm is based on multiplicative update rules and makes a use of prior knowledge of spectral signatures in the scene. Numerous experiments both on synthetic and real data lead to the conclusion that NMFupk is a useful model to solve the hyperspectral unmixing problem with such prior knowledge and behaves better than NMF without prior knowledge.

6 Acknowledgment

The work was supported by the National Natural Science Foundation of China under the Grants 60975003 and 91120301, the 973 Program under the Grant 2010CB327904, the open funding project of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University (Grant No. BUAA-VR-12KF-07), the Program for New Century Excellent Talents in University of Ministry of Education of China under the Grant NCET-11-0775, and the Beijing Natural Science Foundation (Non-negative Component Analysis for Hyperspectral Imagery Unmixing) under the Grant 4112036.

Appendix: Proof of Theorem 1

First, we prove convergence of the update rule for $\mathbf{H}_1$ in (15) (here, we use $\widetilde{\mathbf{H}}$ for a replacement to avoid confusion by subscript later), and then convergence of the other two update rules could be proved in the similar way. And, we will employ auxiliary

function similar to that used in.[20]

Rewrite $\mathbf{W}$ and $\mathbf{H}$ in a partitioned form:

$$\mathbf{W} = \left[\widetilde{\mathbf{W}}, \overline{\mathbf{W}}\right], \mathbf{H} = \begin{bmatrix}\widetilde{\mathbf{H}} \\ \overline{\mathbf{H}}\end{bmatrix}. \tag{21}$$

The objective function for a single row, denoted $\widetilde{\mathbf{h}}$, can be shown as

$$F(\widetilde{\mathbf{h}}) = \frac{1}{2}\left\|\mathbf{v} - \widetilde{\mathbf{W}}\widetilde{\mathbf{h}} - \overline{\mathbf{W}}\bar{\mathbf{h}}\right\|^2 \tag{22}$$

$$= \frac{1}{2}\sum_i (\mathbf{v}_i - \sum_a \widetilde{\mathbf{W}}_{ia}\widetilde{\mathbf{h}}_a - \sum_p \overline{\mathbf{W}}_{ip}\bar{\mathbf{h}}_p)^2 \tag{23}$$

where $\mathbf{v}$, $\widetilde{\mathbf{h}}$ and $\bar{\mathbf{h}}$ indicate separate rows of $\mathbf{V}$, $\widetilde{\mathbf{H}}$ and $\overline{\mathbf{H}}$ respectively. We define an auxiliary function $G(\widetilde{\mathbf{h}}, \widetilde{\mathbf{h}}^t)$ for $F(\widetilde{\mathbf{h}})$ satisfying the conditions

$$G(\widetilde{\mathbf{h}}, \widetilde{\mathbf{h}}^t) \geq F(\widetilde{\mathbf{h}}), G(\widetilde{\mathbf{h}}, \widetilde{\mathbf{h}}) = F(\widetilde{\mathbf{h}}). \tag{24}$$

Then, $F(\widetilde{\mathbf{h}})$ is nonincreasing under the update

$$\widetilde{\mathbf{h}}^{t+1} = \arg\min_{\widetilde{\mathbf{h}}} G(\widetilde{\mathbf{h}}, \widetilde{\mathbf{h}}^t). \tag{25}$$

The proof is given in.[20]

Given $\mathbf{K}(\widetilde{\mathbf{h}}^t)$ is the diagonal matrix

$$\mathbf{K}_{ab}(\widetilde{\mathbf{h}}^t) = \delta_{ab}(\widetilde{\mathbf{W}}^T\widetilde{\mathbf{W}}\widetilde{\mathbf{h}}^t + \widetilde{\mathbf{W}}^T\overline{\mathbf{W}}\bar{\mathbf{h}}^t)_a/\widetilde{\mathbf{h}}^t_a, \tag{26}$$

we define function G as

$$G(\widetilde{\mathbf{h}}, \widetilde{\mathbf{h}}^t) = F(\widetilde{\mathbf{h}}^t) + (\widetilde{\mathbf{h}} - \widetilde{\mathbf{h}}^t)^T\nabla F(\widetilde{\mathbf{h}}^t) + \frac{1}{2}(\widetilde{\mathbf{h}} - \widetilde{\mathbf{h}}^t)^T\mathbf{K}(\widetilde{\mathbf{h}}^t)(\widetilde{\mathbf{h}} - \widetilde{\mathbf{h}}^t) \tag{27}$$

where function $\delta(\mathbf{x})$ means the diagonalization of vector $\mathbf{x}$. Obviously, $G(\widetilde{\mathbf{h}}, \widetilde{\mathbf{h}}) = F(\widetilde{\mathbf{h}})$. On the other hand, $F(\widetilde{\mathbf{h}})$ can be expanded as

$$F(\widetilde{\mathbf{h}}) = F(\widetilde{\mathbf{h}}^t) + (\widetilde{\mathbf{h}} - \widetilde{\mathbf{h}}^t)^T \nabla F(\widetilde{\mathbf{h}}^t) + \frac{1}{2}(\widetilde{\mathbf{h}} - \widetilde{\mathbf{h}}^t)^T (\widetilde{\mathbf{W}}^T \widetilde{\mathbf{W}})(\widetilde{\mathbf{h}} - \widetilde{\mathbf{h}}^t). \tag{28}$$

Comparing (27) with (28), we find that $G(\widetilde{\mathbf{h}}, \widetilde{\mathbf{h}}^t) \geq F(\widetilde{\mathbf{h}})$ is equivalent to

$$(\widetilde{\mathbf{h}} - \widetilde{\mathbf{h}}^t)^T [\mathbf{K}(\widetilde{\mathbf{h}}^t) - \widetilde{\mathbf{W}}^T \widetilde{\mathbf{W}}](\widetilde{\mathbf{h}} - \widetilde{\mathbf{h}}^t) \geq 0. \tag{29}$$

Rescaling the components of $\mathbf{K}(\widetilde{\mathbf{h}}^t) - \widetilde{\mathbf{W}}^T \widetilde{\mathbf{W}}$, we can get the matrix

$$\mathbf{M}_{ab}(\widetilde{\mathbf{h}}^t) = \widetilde{\mathbf{h}}_a^t (\mathbf{K}(\widetilde{\mathbf{h}}^t) - \widetilde{\mathbf{W}}^T \widetilde{\mathbf{W}})_{ab} \widetilde{\mathbf{h}}_b^t. \tag{30}$$

Then $\mathbf{K}(\widetilde{\mathbf{h}}^t) - \widetilde{\mathbf{W}}^T \widetilde{\mathbf{W}}$ is semipositive definite if and only if $\mathbf{M}$ is, and

$$\begin{aligned}
\mathbf{v}^T \mathbf{M} \mathbf{v} &= \sum_{ab} \mathbf{v}_a \mathbf{M}_{ab} \mathbf{v}_b && (31)\\
&= \sum_{ab} \mathbf{v}_a \widetilde{\mathbf{h}}_a^t (\mathbf{K}(\widetilde{\mathbf{h}}^t) - \widetilde{\mathbf{W}}^T \widetilde{\mathbf{W}})_{ab} \widetilde{\mathbf{h}}_b^t \mathbf{v}_b && (32)\\
&= \sum_{ab} \delta_{ab} (\widetilde{\mathbf{W}}^T \widetilde{\mathbf{W}} \widetilde{\mathbf{h}}^t + \widetilde{\mathbf{W}}^T \overline{\mathbf{W}} \overline{\mathbf{h}}^t)_a \widetilde{\mathbf{h}}_b^t \mathbf{v}_a^2 - \mathbf{v}_a \widetilde{\mathbf{h}}_a^t (\widetilde{\mathbf{W}}^T \widetilde{\mathbf{W}})_{ab} \widetilde{\mathbf{h}}_b^t \mathbf{v}_b && (33)\\
&\geq \sum_{ab} \delta_{ab} (\widetilde{\mathbf{W}}^T \widetilde{\mathbf{W}} \widetilde{\mathbf{h}}^t)_a \widetilde{\mathbf{h}}_b^t \mathbf{v}_a^2 - \mathbf{v}_a \widetilde{\mathbf{h}}_a^t (\widetilde{\mathbf{W}}^T \widetilde{\mathbf{W}})_{ab} \widetilde{\mathbf{h}}_b^t \mathbf{v}_b && (34)\\
&= \sum_{ab} \widetilde{\mathbf{h}}_a^t (\widetilde{\mathbf{W}}^T \widetilde{\mathbf{W}})_{ab} \widetilde{\mathbf{h}}_b^t \mathbf{v}_a^2 - \mathbf{v}_a \widetilde{\mathbf{h}}_a^t (\widetilde{\mathbf{W}}^T \widetilde{\mathbf{W}})_{ab} \widetilde{\mathbf{h}}_b^t \mathbf{v}_b && (35)\\
&= \sum_{ab} (\widetilde{\mathbf{W}}^T \widetilde{\mathbf{W}})_{ab} \widetilde{\mathbf{h}}_a^t \widetilde{\mathbf{h}}_b^t [\frac{1}{2}\mathbf{v}_a^2 + \frac{1}{2}\mathbf{v}_b^2 - \mathbf{v}_a \mathbf{v}_b] && (36)\\
&= \frac{1}{2} \sum_{ab} (\widetilde{\mathbf{W}}^T \widetilde{\mathbf{W}})_{ab} \widetilde{\mathbf{h}}_a^t \widetilde{\mathbf{h}}_b^t (\mathbf{v}_a - \mathbf{v}_b)^2 && (37)\\
&\geq 0. && (38)
\end{aligned}$$

Thus, $G(\widetilde{\mathbf{h}}, \widetilde{\mathbf{h}}^t)$ is an auxiliary function for $F(\widetilde{\mathbf{h}})$. In that case, the update rule can

be built by replacing $G(\widetilde{\mathbf{h}}, \widetilde{\mathbf{h}}^t)$ in (24) by (27):

$$\widetilde{\mathbf{h}}^{t+1} = \widetilde{\mathbf{h}}^t - \mathbf{K}(\widetilde{\mathbf{h}}^t)^{-1}\nabla F(\widetilde{\mathbf{h}}^t). \tag{39}$$

And, by partially differentiating the objective function in (22) with respect to $\widetilde{\mathbf{h}}$, we have

$$\nabla F(\widetilde{\mathbf{h}}) = -\widetilde{\mathbf{W}}^T\mathbf{v} + \widetilde{\mathbf{W}}^T\widetilde{\mathbf{W}}\widetilde{\mathbf{h}} + \widetilde{\mathbf{W}}^T\overline{\mathbf{W}}\bar{\mathbf{h}}. \tag{40}$$

Then, writing the components of (39) explicitly, we get

$$\widetilde{\mathbf{h}}_a^{t+1} = \widetilde{\mathbf{h}}_a^t \frac{(\widetilde{\mathbf{W}}^T\mathbf{v})_a}{(\widetilde{\mathbf{W}}^T\widetilde{\mathbf{W}}\widetilde{\mathbf{h}}^t + \widetilde{\mathbf{W}}^T\overline{\mathbf{W}}\bar{\mathbf{h}}^t)_a}. \tag{41}$$

Hence, the convergence of update rule for $\mathbf{H}_1$ in (15) has been proved. And, convergence of the other two update rules in (14) and (16) can be proved in the same way.

References

[1] Goetz, Alexander F.H., "Measuring the earth from above: 30 years (and Counting) of hyperspectral imaging," *Photonics Spectra* **45**(6) (Jun. 2011).

[2] J. M. Bioucas-Dias and Antonio Plaza, "An overview on hyperspectral unmixing: geometrical, statistical, and sparse regression based approaches," *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International* 1135-1138 (Jul. 2011).

[3] Ershad Sharifahmadian and Shahram Latifi, "Advanced Hyperspectral Remote Sensing for Target Detection," *Systems Engineering (ICSEng)* 200-205 (Aug. 2011).

[4] Zhenwei Shi, Zhenyu An, and Zhiguo Jiang, "Hyperspectral image fusion by the similarity measure based variational method," *Optical Engineering* **50**(7) (Mar. 2011)

[5] Zhenwei Shi, Shuo Yang, and Zhiguo Jiang, "Hyperspectral target detection using regularized high-order matched filter," *Optical Engineering* **50**(5) (May 2011)

[6] Nirmal Keshava and John F. Mustard, "Spectral Unmixing," S*ignal Processing Magazine, IEEE* **19**(1), 44-57 (Jan. 2002).

[7] Maria Petrou and Patricia G. Foschi, "Confidence in Linear Spectral Unmixing of Single Pixels," *IEEE Transactions on Geoscience and Remote Sensing* **37**(1), 624-626 (Jan. 1999).

[8] Y. H. Hu, H. B. Lee, and F. L. Scarpace, "Optimal Linear Spectral Unmixing," *IEEE Transactions on Geoscience and Remote Sensing* **37**(1), 639-644 (Jan. 1999).

[9] Mario Parente and Antonio Plaza, "Survey of geometric and statistical unmixing algorithms for hyperspectral images," *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2010 2nd Workshop on* 1-4 (Jun. 2010).

[10] J. Boardman, "Automating spectral unmixing of AVIRIS data using convex geometry concepts," *Proc. Summ. 4th Annu. JPL Airborne Geoscience Workshop* **1**, 11-14 (1993).

[11] M. E. Winter, "N-findr: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," *Proc. SPIE Conf. Imaging Spectrometry V* 266-275 (1999).

[12] J. M. Nascimento and J. M. Bioucas-Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing* **43**(4), 898-910 (Apr. 2005).

[13] J. M. Peixoto Nascimento, "Unsupervised Hyperspectral Unmixing," *Universidade techica de lisboa instituto superior tecnico* (2006).

[14] T. H. Chan, C. Y. Chi, Y. M. Huang, and W. K. Ma, "A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing* **47**(11), 4418-4432 (2009).

[15] J. Li and J. Bioucas-Dias, "Minimum volume simplex analysis: a fast algorithm to unmix hyperspectral data," *IEEE Geoscience and Remote Sensing Symposium- IGARSS08* **3**, 250-253 (2008).

[16] M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Dunne, and J. F Huntington, "ICE: a statistical approach to identifying endmembers in hyperspectral

images," *IEEE Transactions on Geoscience and Remote Sensing* **42**(10), 2085-2095 (2004).

[17] Junmin Liu and Jiangshe Zhang, "A New Maximum Simplex Volume Method Based on Householder Transformation for Endmember Extraction," *Geoscience and Remote Sensing, IEEE Transactions on* **50**(1), 104-108 (Jun. 2011).

[18] D. D. Lee and H. S. Seung, "Learning the parts of objects by nonnegative matrix factorization," *Nature* **401**(6755), 788-791 (Oct. 1999).

[19] V. Paul Pauca, J. Piper, and Robert J. Plemmons, "Nonnegative matrix factorization for spectral data analysis," *Linear Algebra Appl.* **416**(1), 29-47 (Jul. 2006).

[20] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Adv. Neural Inform. Process. Syst.* **13**, 556-562 (2002).

[21] S. Tompkins, K. McNaron-Brown, J. Sunshine and J. Burt, "Landcover Change over Central Virginia: Comparison of Endmember Fractions in Hyperspectral Data," in *Proceedings of SPIE.* **4381**, 477-488 (2001).

[22] J. B. Adams, D. E. Sabol, V. Kapos, R. A. Filho, D. A. Roberts, M. O. Smith and A. R. Gillespie, "Classification of Multispectral Images Based on Fractions of Endmembers: Application to Land-Cover Change in the Brazilian Amazon," in *Remote Sens. Environ.* **52**, 137-154 (1995).

[23] R. N. Clark, G. A. Swayze, K. E. Livo, R. F. Kokaly, S. J. Sutley, J. B. Dalton, R. R. McDougal, and C. A. Gent, "Imaging spectroscopy: Earth and planetary

remote sensing with the USGS Tetracorder and expert systems," *Journal of Geophysical Research* **108[E12]**, 5131 (2003).

[24] Maria Petrou and Costas Petrou, *Image Processing: The Fundamentals, 2nd Edition.* Wiley.

[25] T. M. Cover and J. A. Thomas, *Elements of information theory*, John Wiley (1991).

[26] Minje Kim, Jiho Yoo, Kyeongok Kang and Seungjin Choi, "Blind rhythmic source separation: Nonnegativity and repeatability," *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)* 2006-2009 (Mar. 2010).

[27] Jiho Yoo, Minje Kim, Kyeongok Kang, and Seungjin Choi, "Nonnegative matrix partial co-factorization for drum source separation," *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)* 1942-1945 (Mar. 2010).

[28] C.-I. Chang, *Hyperspectral imaging: techniques for spectral detection and classification*, Kluwer Academic and Plenum Publishers (2003).

[29] A. Plaza, P. Martinez, R. Perez, and J. Plaza, "A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing* **42**(3), 650-663 (Mar. 2004).

[30] R. N. Clark, G. A. Swayze, A. J. Gallagher, T. V. V. King, and W. M.Calvin, "The U.S. geological survey digital spectral library: Version 1: 0.2 to 3.0 microns," *U.S. Geological Survey, Denver, CO, Open File Rep.* 93-592 (1993).

[31] AVIRIS Data Products. [Online]. Available: http://aviris.jpl.nasa.gov/html/aviris.freedata.html

[32] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Transactions on Geoscience and Remote Sensing* **45**(3), 765-777 (Mar. 2007).

[33] T.-H. Chan, C.-Y. Chi, Y.-M. Huang, and W.-K. Ma, "A convex analysis based minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Transactions Signal Processing* **57**(11), 4418-4432 (Nov. 2009).

[34] Tech. Rep. [Online]. Available: http://speclab.cr.usgs.gov/cuprite.html

[35] ArulMurugan Ambikapathi, Tsung-Han Chan, Wing-Kin Ma, and Chong-Yung Chi, "Chance-Constrained Robust Minimum-Volume Enclosing Simplex Algorithm for Hyperspectral Unmixing," *IEEE Transactions on Geoscience and Remote Sensing* **49**(11), 4194 - 4209 (Nov. 2011).

[36] R. N. Clark, J. B. Dalton, J. Mustard, F. Kruse, Cindy Ong, C. Pieters, and G. A. Swayze, "Mineral Mapping and Applications of Imaging Spectroscopy," *Geoscience and Remote Sensing Symposium, 2006. IGARSS 2006. IEEE International Conference on* 1986-1989 (2006).

A Biographies

Wei Tang: He is currently pursuing his M.S. degree in the Image Processing Center, School of Astronautics, Beihang University. His research interest is hyperspectral image processing. E-mail: tangwei125125@gmail.com

Zhenwei Shi: He received his Ph.D. degree in Mathematics from Dalian University of Technology, Dalian, China, in 2005. He was a Postdoctoral Researcher in the Department of Automation, Tsinghua University, Beijing, China, from 2005 to 2007. From 2008, he is an associate professor in the Image Processing Center, School of Astronautics, Beijing University of Aeronautics and Astronautics. His research interests include hyperspectral remote sensing, blind signal processing, image processing, pattern recognition and machine learning. E-mail: shizhenwei@buaa.edu.cn

Zhenyu An: He is currently pursuing his Ph.D. degree in the Image Processing Center, School of Astronautics, Beihang University. His research interest is hyperspectral image processing. E-mail: anzhenyu_155@126.com