

# Investigation on Constrained Matrix Factorization for Hyperspectral Image Analysis

Qian Du<sup>1</sup>, Ivica Kopriva<sup>2</sup>, Harold Szu<sup>2,3</sup>

<sup>1</sup>Department of Electrical and Computer Engineering, Mississippi State University, MS 39762, USA

<sup>2</sup>Department of Electrical and Computer Engineering, The George Washington University, DC 20052, USA

<sup>3</sup>Office of Naval Research, Arlington, VA22217, USA

**Abstract** — Matrix factorization is applied to unsupervised linear unmixing for hyperspectral imagery. The algorithm, called non-negative matrix factorization, is used. It imposes a constraint on the non-negativity of the amplitudes of the recovered endmember spectral signatures as well as their fractional abundances. This ensures the recovery of physically meaningful endmembers and their abundances. This algorithm is further modified to include the sum-to-one constraint such that the sum of the fractional abundances for each pixel is unity. Several practical implementation issues in hyperspectral image unmixing are discussed. Some preliminary results from AVIRIS experiments are presented. We also discuss the advantages and possible limitations of this method in hyperspectral image analysis.

**Keywords:** matrix factorization; nonnegative matrix factorization; linear mixture model; unsupervised linear unmixing; hyperspectral imagery.

## I. INTRODUCTION

Linear unmixing analysis is a well-known technique in remote sensing image analysis. It is based on the fact that the rough spatial resolution permits different materials present in the area covered by a single pixel. The linear mixture model says that a pixel reflectance in a visible-near infrared multispectral or hyperspectral image is the linear mixture from all independent pure materials (endmembers) present in an image scene [1].

Let  $L$  be the number of spectral bands and  $\mathbf{r}$  an  $L \times 1$  column pixel vector in a multispectral or hyperspectral image. Assume that there are  $P$  objects/materials (i.e., endmembers) present in an image scene, which construct an  $L \times P$  signature matrix  $\mathbf{M} = [\mathbf{m}_1 \mathbf{m}_2 \cdots \mathbf{m}_P]$ , where  $\mathbf{m}_j$  represents the  $j$ -th endmember. Assume that  $\boldsymbol{\alpha} = (\alpha_1 \alpha_2 \cdots \alpha_P)^T$  is a  $p \times 1$  abundance vector associated with  $\mathbf{r}$ , where  $\alpha_j$  denotes the abundance fraction of the  $\mathbf{m}_j$  in  $\mathbf{r}$ . In the linear mixture model,  $\mathbf{r}$  is considered as the linear mixture of  $\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_P$  as

$$\mathbf{r} = \mathbf{M}\boldsymbol{\alpha} + \mathbf{n} \quad (1)$$

where  $\mathbf{n}$  is included to account for either measurement or model error [1]. If  $\mathbf{M}$  is assumed to be known and keeps to be

the same for all the pixels, then the problem is to estimate  $\boldsymbol{\alpha}$  which is changed with pixel.

A typical method to estimate  $\boldsymbol{\alpha}$  is the least squares approach. The estimate from the least squares solution is the one that minimizes the estimation residual

$$\min_{\boldsymbol{\alpha}} (\mathbf{r} - \mathbf{M}\boldsymbol{\alpha})^T (\mathbf{r} - \mathbf{M}\boldsymbol{\alpha}). \quad (2)$$

In order for the estimated abundance vector  $\boldsymbol{\alpha}$  to faithfully represent an image pixel vector  $\mathbf{r}$ , two constraints are generally imposed on  $\boldsymbol{\alpha}$  in Eq. (1): (a) abundance sum-to-one constraint,

referred to as ASC,  $\sum_{p=1}^P \alpha_p = 1$ ; and (b) abundance non-

negativity constraint, referred to as ANC,  $\alpha_p \geq 0$  for all  $1 \leq p \leq P$ . There is no closed-form solution to such a constrained linear unmixing problem. So an iterative method generally is used.

In real applications, endmember signatures in  $\mathbf{M}$  may be unknown and difficult to obtain. Then the task is much more challenging since both  $\mathbf{M}$  and  $\boldsymbol{\alpha}$  need to be estimated. Intuitively, the  $\mathbf{M}$  matrix should be nonnegative. Otherwise, the solution may not be physically realistic.

## II. NON-NEGATIVE MATRIX FACTORIZATION

The linear unmixing problem in Eq. (1) can also be solved by matrix factorization. Recently, non-negative matrix factorization (NMF) is developed [2-7]. Given a non-negative data matrix  $\mathbf{X}$  of size  $L \times IJ$ , NMF can find an approximate factorization  $\mathbf{X} \approx \mathbf{A}\mathbf{S}$  into non-negative factors  $\mathbf{A}$  of size  $L \times P$  and  $\mathbf{S}$  of size  $P \times IJ$ . The non-negativity constraint imposed on  $\mathbf{A}$  and  $\mathbf{S}$  makes them purely additive, in contrast to other factor analysis techniques such as principal component analysis (PCA) and independent component analysis (ICA). Let  $IJ$  be the number of pixels in an image of size  $I \times J$ , and let  $\mathbf{X} = \{\mathbf{r}_i\}_{i=1}^{IJ}$  be the data matrix including all the pixel vectors. Then the linear mixture model in Eq. (1) can be represented as

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{N} \quad (2)$$

# Report Documentation Page

*Form Approved*  
*OMB No. 0704-0188*

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE <b>25 JUL 2005</b>	2. REPORT TYPE <b>N/A</b>	3. DATES COVERED <b>-</b>	
4. TITLE AND SUBTITLE <b>Investigation on Constrained Matrix Factorization for Hyperspectral Image Analysis</b>		5a. CONTRACT NUMBER	
		5b. GRANT NUMBER	
		5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)		5d. PROJECT NUMBER	
		5e. TASK NUMBER	
		5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) <b>Department of Electrical and Computer Engineering, Mississippi State University, MS 39762, USA</b>		8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)	
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT <b>Approved for public release, distribution unlimited</b>			
13. SUPPLEMENTARY NOTES <b>See also ADM001850, 2005 IEEE International Geoscience and Remote Sensing Symposium Proceedings (25th) (IGARSS 2005) Held in Seoul, Korea on 25-29 July 2005.</b>			
14. ABSTRACT			
15. SUBJECT TERMS			
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT
a. REPORT <b>unclassified</b>	b. ABSTRACT <b>unclassified</b>	c. THIS PAGE <b>unclassified</b>	<b>UU</b>
			18. NUMBER OF PAGES <b>3</b>
			19a. NAME OF RESPONSIBLE PERSON

where  $\mathbf{A}$  corresponds to the endmember matrix  $\mathbf{M}$ ,  $\mathbf{S} = \{\mathbf{a}_i\}_{i=1}^{IJ}$  is the abundance matrix with the  $i$ -th column vector representing the abundances for the  $i$ -th pixel, and  $\mathbf{N} = \{\mathbf{n}\}_{i=1}^{IJ}$ . Eq. (2) will be used thereafter.

The formulation of the NMF algorithm of Lee and Seung in [4] is reviewed as follows. Assume noise is Gaussian distributed. The maximum likelihood estimation of  $\mathbf{A}$  and  $\mathbf{S}$  are

$$\begin{aligned} \hat{\mathbf{A}}, \hat{\mathbf{S}} &= \arg \max_{\mathbf{A}, \mathbf{S}} p(\mathbf{X} | \mathbf{A}, \mathbf{S}) \\ &= \arg \min_{\mathbf{A}, \mathbf{S}} (-\log p(\mathbf{X} | \mathbf{A}, \mathbf{S})) \\ &= \arg \min_{\mathbf{A}, \mathbf{S}} \|\mathbf{X} - \mathbf{AS}\|^2 \\ &\text{subject to } \mathbf{A} \geq 0, \mathbf{S} \geq 0 \end{aligned} \quad (3)$$

Define  $F = \|\mathbf{X} - \mathbf{AS}\|^2$ . The gradients with respect to  $\mathbf{A}$  and  $\mathbf{S}$  are given by

$$\begin{aligned} \frac{\partial F}{\partial A_{l,p}} &= -2 \left( (\mathbf{XS}^T)_{l,p} - (\mathbf{ASS}^T)_{l,p} \right) \\ \frac{\partial F}{\partial S_{p,i}} &= -2 \left( (\mathbf{A}^T \mathbf{X})_{p,i} - (\mathbf{A}^T \mathbf{AS})_{p,i} \right) \end{aligned} \quad (4)$$

where  $l = 1, \dots, L$ ,  $p = 1, \dots, P$ , and  $i = 1, \dots, IJ$  are indexes for  $\mathbf{A}$  and  $\mathbf{S}$ . Then the adaptation functions can be constructed as

$$\begin{aligned} A_{l,p} &\leftarrow A_{l,p} + \delta_{l,p} \left( (\mathbf{XS}^T)_{l,p} - (\mathbf{ASS}^T)_{l,p} \right) \\ S_{p,i} &\leftarrow S_{p,i} + \eta_{p,i} \left( (\mathbf{A}^T \mathbf{X})_{p,i} - (\mathbf{A}^T \mathbf{AS})_{p,i} \right) \end{aligned} \quad (5)$$

where  $\delta_{l,p}$  and  $\eta_{p,i}$  are learning rates. In [4] it is shown that by choosing them as  $\delta_{l,p} = A_{l,p} / (\mathbf{ASS}^T)_{l,p}$  and  $\eta_{p,i} = S_{p,i} / (\mathbf{A}^T \mathbf{AS})_{p,i}$ , the adaptation functions become

$$\begin{aligned} A_{l,p} &\leftarrow A_{l,p} \frac{(\mathbf{XS}^T)_{l,p}}{(\mathbf{ASS}^T)_{l,p}} \\ S_{p,i} &\leftarrow S_{p,i} \frac{(\mathbf{A}^T \mathbf{X})_{p,i}}{(\mathbf{A}^T \mathbf{AS})_{p,i}} \end{aligned} \quad (6)$$

which is a multiplicative update rule.

The update rule in Eq. (6) is proved to converge to at least a local optimal maximum likelihood solution [5]. We can see that the advantage of using such a multiplicative update rule in Eq. (6) over the additive rule in Eq. (5) is the guarantee of non-negativity of  $\mathbf{A}$  and  $\mathbf{S}$ , provided that they are initiated non-negative and the data matrix  $\mathbf{X}$  is non-negative. Also, it is unnecessary to choose any learning rate.

### III. CONSTRAINED MATRIX FACTORIZATION FOR HYPERSPECTRAL IMAGE LINEAR UNMIXING

The NMF algorithm has been applied to some real applications [8-10]. Here we discuss several practical implementation issues when it is applied to linear unmixing of hyperspectral imagery.

#### Sum-to-one Constraint

The NMF needs to be modified such that the sum-to-one constraint can be satisfied. This can be achieved by adding one more row of the data matrix  $\mathbf{X}$  and  $\mathbf{A}$  as 1, i.e.,

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{1}^T \end{bmatrix} = \begin{bmatrix} \mathbf{A} \\ \tilde{\mathbf{1}}^T \end{bmatrix} \mathbf{S} + \mathbf{N} = \tilde{\mathbf{A}} \mathbf{S} + \mathbf{N} \quad (6)$$

where  $\mathbf{1}$  and  $\tilde{\mathbf{1}}$  are column vectors of size  $IJ \times 1$  and  $P \times 1$ , respectively, with all the elements equal to 1. Then the last row of  $\tilde{\mathbf{A}}$  will not participate the adaptation.

#### Estimation of the Number of Factors

The number of factors  $P$  ought to be used is unknown, which is the number of endmembers present in an image scene. If the value of  $P$  is changed, then the final results of  $\mathbf{A}$  and  $\mathbf{S}$  are obviously different. This is a common problem for any factor analysis based technique.

The hypothesis testing based eigen-thresholding method in [11] can be applied to estimate the number of distinctive signals in an image scene, and this number is used as  $P$ .

#### Algorithm Initiation

The algorithm may be sensitive to the initial condition. There are different ways to initialize the  $L \times P$  non-negative  $\mathbf{A}$  matrix: 1)  $\mathbf{A}$  is randomly created within the dynamic range of the image; 2)  $P$  pixel vectors are uniformly selected from the image scene as an initial  $\mathbf{A}$ ; 3) Perform vector quantization (VQ) first and the  $P$  clusters are used to construct the  $\mathbf{A}$ . In our research, we use the first method to randomly initialize the  $\mathbf{A}$ . After  $\mathbf{A}$  is initiated,  $\mathbf{S}$  is initiated using the estimate from the fully constrained least squares linear unmixing method in [12].

#### Data Pre-processing

All the water absorption bands and low signal-to-noise ratio bands are removed before linear unmixing. Due to sensor noise, some pixel elements may have negative values, which are replaced with 0.

To summarize, the steps for estimating the endmember matrix  $\mathbf{A}$  and abundance matrix  $\mathbf{S}$  are listed as follows.

- 1) Construct the data matrix  $\tilde{\mathbf{X}}$  by removing bad bands, bad pixel points, and adding the last row as 1.
- 2) Estimate the number of factors  $P$ .
- 3) Initiate  $\tilde{\mathbf{A}}$  as an  $L \times P$  matrix and adding the last row as 1, and initiate  $\mathbf{S}$  by finding the FCLSLU estimate.

4) Update  $\tilde{\mathbf{A}}$  and  $\mathbf{S}$  using Eq. (6) with the last row of  $\tilde{\mathbf{A}}$  unchanged.

#### IV. EXPERIMENT

The AVIRIS Lunar Lake image data was used in computer simulation. The original  $200 \times 200$  subimage was shown in Figure 1. According to the prior information, there are five materials in this scene. So in this preliminary experiment we set  $P$  equal to 5. Then five endmember signatures were randomly initialized. After about 1000 iterations, the five fractional abundance images were generated as shown in Figure 2. The algorithm actually converged very fast, because after about 50~100 iterations there were no obviously difference in the classified images although the estimation error  $F = \|\mathbf{X} - \mathbf{AS}\|^2$  continued to gradually decrease.

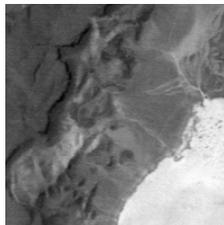


Figure 1. A band subimage of AVIRIS Lunar Lake scene.

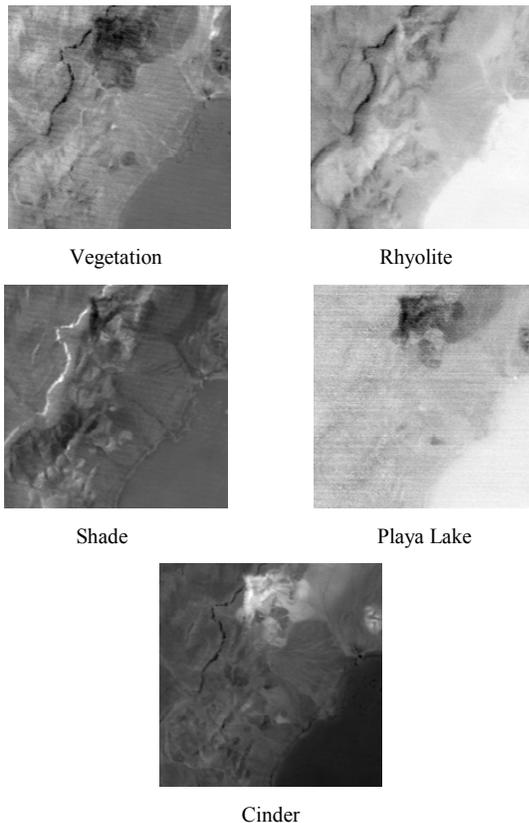


Figure 2. Classification results using constrained matrix factorization.

Compared to the unsupervised least squares result [12], we can see that “Vegetation”, “Shade” and “Cinder” were correctly classified, but “Rhyolite” and “Playa lake” were not well separated.

#### VII. DISCUSSION

The preliminary experiment shows that the constrained matrix factorization algorithm may be feasible to linear-unmixing based hyperspectral image classification, but its performance needs to be further improved. The major problem is that the algorithm is sensitive to the initial conditions. How to find an appropriate  $\mathbf{A}$  initial matrix should be investigated.

The algorithm in this paper basically is maximum likelihood estimation with the assumption that noise is Gaussian distributed. But we know that noise may not be well modeled by Gaussian distribution. So a better noise model is required to improve the overall performance of this algorithm for hyperspectral image analysis.

#### REFERENCE

- [1] J. B. Adams and M. O. Smith, “Spectral mixture modeling: a new analysis of rock and soil types at the Viking lander 1 suite,” *J. Geophysical Res.*, vol. 91, pp. 8098-8112, 1986.
- [2] P. Paatero and U. Tapper, “Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values,” *Environmetrics*, vol. 5, pp. 111-126, 1994.
- [3] K. G. Paterson, J. L. Sagady, and D. L. Hooper, “Analysis of air quality data using positive matrix factorization,” *Environmental Science & Technology*, vol. 33, pp. 635-641, 1999.
- [4] D. D. Lee and H. S. Seung, “Learning the parts of objects by non-negative matrix factorization,” *Nature*, vol. 40, pp. 788-791, 1999.
- [5] D. D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization,” *Advances in Neural Information Processing*, pp. 556-562, 2001.
- [6] D. Donoho and V. Stodden, “When does non-negative matrix factorization give a correct decomposition into parts,” *Proceedings of NIPS 2003 workshop on ICA*, 2003.
- [7] P. O. Hoyer, “Non-negative matrix factorization with sparseness constraints,” *Journal of Machine Learning Research*, vol. 5, pp. 1457-1469, 2004.
- [8] P. Sajda, S. Du, and L. Parra, “Nonnegative matrix factorization for rapid recovery of constituent spectra in magnetic resonance chemical shift imaging of the brain,” *IEEE Transactions on Medical Imaging*, vol. 23, no. 12, 2004.
- [9] M. Bernhardt, J. P. Heather, and M. I. Smith, “New models for hyperspectral anomaly detection and unmixing,” *Proceeding of SPIE Conference on Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XI*, March 2005, Orlando, Florida (to appear).
- [10] Y. M. Masalmah, M. Velez-Reyes, and S. Rosario-Torres, “An algorithm for unsupervised unmixing of hyperspectral imagery using positive matrix factorization,” *Proceeding of SPIE Conference on Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XI*, March 2005, Orlando, Florida (to appear).
- [11] C.-I Chang and Q. Du, “Estimation of number of spectrally distinct signal sources in hyperspectral imagery,” *IEEE Trans. Geoscience Remote Sensing*, vol. 42, pp. 608-619, 2004.
- [12] D. Heinz and C.-I Chang, “Fully constrained least squares linear mixture analysis for material quantification in hyperspectral imagery,” *IEEE Trans. on Geoscience and Remote Sensing*, vol. 39, no. 3, pp. 529-545, 2001.