A CONSTRAINED NON-NEGATIVE MATRIX FACTORIZATION APPROACH TO UNMIX HIGHLY MIXED HYPERSPECTRAL DATA

Lidan Miao and Hairong Qi

University of Tennessee
Department of Electrical and Computer Engineering
Knoxville, TN 37996

ABSTRACT

This paper presents a blind source separation method to unmix highly mixed hyperspectral data, i.e., each pixel is a mixture of responses from multiple materials and no pure pixels are present in the image due to large sampling distance. The algorithm introduces a minimum volume constraint to the standard non-negative matrix factorization (NMF) formulation, referred to as the minimum volume constrained NMF (MVC-NMF). MVC-NMF explores two important facts: first, the spectral data are non-negative; second, the constituent materials occupy the vertices of a simplex, and the simplex volume determined by the actual materials is the minimum among all possible simplexes that circumscribe the data scatter space. The experimental results based on both synthetic mixtures and a real image scene demonstrate that the proposed method outperforms several state-of-the-art approaches.

Index Terms— Hyperspectral imagery, spectral unmixing, endmember extraction, non-negative matrix factorization, linear mixture model.

1. INTRODUCTION

Hyperspectral sensors capture electromagnetic energy from 400 to 2,500 nanometers with hundreds of narrow wavelength intervals. The increased spectral information provides fine details that allow to discriminate among different materials. However, due to the large ground sampling distance, numerous disparate materials additively contribute to the spectrum measured at a single pixel [1], resulting in the so-called *mixed pixels*. A challenge in analyzing these mixed pixels is to identify the individual constituent materials (*endmembers*) in the mixture, and infer their relative proportions (*abundances*) in terms of the spatial coverage within a single pixel, a process referred to as *spectral unmixing*.

In practical applications, the mixture formation is widely modelled as a linear process due to its effectiveness and simplicity. The basic formation model is expressed as

$$\mathbf{X} = \mathbf{AS} + \mathbf{E} \tag{1}$$

where the columns of $\mathbf{X} \in \mathbf{R}^{l \times n}$ denote the observation vectors of n pixels measured at l spectral bands. $\mathbf{A} \in \mathbf{R}^{l \times c}$ is the material signature matrix whose columns, $\{\mathbf{a}_j\}_{j=1}^{c} \in \mathbf{R}^{l}$, correspond to the spectral signatures of c endmembers. The abundance vectors are represented by the columns of $\mathbf{S} \in \mathbf{R}^{c \times n}$, which satisfy two physical constraints: first, each element of $\mathbf{S}$ is non-negative; second, the sum of column elements equals 1; that is, $\sum_{j=1}^{c} s_{ji} = 1, i = 1, \ldots, n$. $\mathbf{E}$ denotes the possible noise and errors.

Given this formation model and assume the information of endmembers (i.e., $\mathbf{A}$) is known, the spectral unmixing reduces to a linear regression problem. In this paper, we focus on *blind* unmixing; that is, both $\mathbf{A}$ and $\mathbf{S}$ are unknown. Various techniques have been investigated to perform this estimation, varying from the popular independent component analysis (ICA) [2], the iterative constrained least squares [3], to the convex hull analysis [4]. The convex geometry-based algorithms exploit the fact that the observations in a hyperspectral scene are within a simplex whose vertices correspond to the endmembers [4, 5]. But, these approaches are either computationally prohibitive or based on the assumption that the image contains *pure pixels*, which is not realistic for hyperspectral images with low spatial resolutions. Recently, non-negative matrix factorization (NMF) [6], as a blind source separation method, has been applied to hyperspectral unmixing [7, 8]. However, the simple sum-to-one constraint [7] and the endmember smoothness constraint [8] are not effective to yield a reliable estimate. Moreover, both algorithms are sensitive to algorithm initializations.

In this paper, we present a new constrained NMF method, which integrates the least squares analysis and the convex geometry model by incorporating a volume constraint into the NMF formulation, referred to as the minimum volume constrained NMF method (MVC-NMF). The objective is to find a simplex that circumscribes the data space, whose vertices are mostly close to the real endmembers. The initial simplex is constructed by randomly choosing c data points. The learning algorithm then expands the simplex under the effect of two forces: the *external* force (minimizing the approximation error) drives the estimation to move outward of the data cloud; and the *internal* force (minimizing the simplex volume) acts

This work is supported in part by Office of Naval Research under grant no. N00014-04-1-0797.

in the opposite direction by forcing the endmembers to be as close to each other as possible. Through experimental validations, we observe that the balance between these two forces effectively guides the learning process to converge to the true endmember locations. In the next section, we will give a detailed description of the proposed MVC-NMF algorithm. The experimental results will be demonstrated in Sec. 3, and Sec. 4 concludes the paper.

2. THE MVC-NMF ALGORITHM

NMF is a popular matrix factorization method [6]. Mathematically, given a non-negative matrix $\mathbf{X}$, NMF looks for two matrices $\mathbf{A}$ and $\mathbf{S}$ with non-negative elements such that $\mathbf{X} \approx \mathbf{AS}$. Geometrically, for the data in a hyperspace, NMF with the sum-to-one constraint aims at constructing a simplex enclosing the data, whose vertices correspond to the columns of $\mathbf{A}$, and $\mathbf{S}$ consists of the weight coefficients of these vertices in generating each data point. It is apparent that the solution is not unique. For spectral unmixing, we assume the data cloud retains certain shape which allows the reconstruction of simplex. Then, the endmembers correspond to the most compact simplex among all possible simplexes that circumscribe the data space [4]. In addition, the noise normally results in a bigger data cloud, which thereby leads to a simplex with larger volume than the clean data. Based on these observations, we propose to introduce the volume constraint to the standard NMF formulation.

2.1. Problem Formulation

Combining the goal of minimum approximation error with the volume constraint, we arrive at the following constrained optimization problem

$$\begin{aligned} &\text{minimize} \quad f(\mathbf{A},\mathbf{S}) = \frac{1}{2}\|\mathbf{X}-\mathbf{AS}\|_F^2 + \lambda J(\mathbf{A}) \\ &\text{subject to} \quad \mathbf{A} \succeq \mathbf{0},\ \mathbf{S} \succeq \mathbf{0},\ \mathbf{1}_c^T\mathbf{S} = \mathbf{1}_n^T \end{aligned} \tag{2}$$

where the operator $\|\cdot\|_F$ denotes the Frobenius norm, and $J(\mathbf{A})$ is the penalty function, calculating the simplex volume determined by the estimated endmembers. The regularization parameter $\lambda \in \mathbf{R}$ is used to control the tradeoff between the accurate reconstruction and the volume constraint. The symbol $\succeq$ denotes *componentwise inequality*, i.e., $\mathbf{A} \succeq \mathbf{0}$ means $a_{ij} \geq 0$ for $i = 1,\ldots,l, j = 1,\ldots,c$. $\mathbf{1}_c$ ($\mathbf{1}_n$) is a c (n)-dimensional column vector of all 1s. The first term of the objective serves as the external force to drive the search to move outward, so that the generated simplex contains all data points with relatively small errors. The second term serves as the internal force, which constrains the simplex volume to be small. A solution is found when these two forces balance each other.

The simplex volume is calculated based on the connection between the volume and the determinant, which leads to

$$J(\mathbf{A}) = \frac{1}{2(c-1)!}\det^2\left(\begin{bmatrix} \mathbf{1}_c^T \\ \tilde{\mathbf{A}} \end{bmatrix}\right) \tag{3}$$

where the matrix $\tilde{\mathbf{A}} \in \mathbf{R}^{(c-1)\times c}$ is a low dimensional transform of $\mathbf{A}$ given by

$$\tilde{\mathbf{A}} = \mathbf{U}^T(\mathbf{A} - \boldsymbol{\mu}\mathbf{1}_c^T) \tag{4}$$

The matrix $\mathbf{U} \in \mathbf{R}^{l\times(c-1)}$ is formed by the $c-1$ most significant principal components of $\mathbf{X}$ through principal component analysis. The column vector $\boldsymbol{\mu}$ is the data mean. To formulate the penalty term as a function of $\mathbf{A}$, we define two constant matrices

$$\mathbf{C} \triangleq \begin{bmatrix} \mathbf{1}_c^T \\ \underline{\mathbf{0}} \end{bmatrix},\ \mathbf{B} \triangleq \begin{bmatrix} \mathbf{0}_{c-1}^T \\ \underline{\mathbf{I}} \end{bmatrix} \tag{5}$$

with $\underline{\mathbf{0}}$ being a $(c-1)\times c$ zero matrix and $\underline{\mathbf{I}}$ an identity matrix of size $(c-1)\times(c-1)$. We next define

$$\mathbf{Z} \triangleq \begin{bmatrix} \mathbf{1}_c^T \\ \tilde{\mathbf{A}} \end{bmatrix} = \mathbf{C} + \mathbf{B}\mathbf{U}^T(\mathbf{A} - \boldsymbol{\mu}\mathbf{1}_c^T) \tag{6}$$

Then, the objective function is finalized as

$$f(\mathbf{A},\mathbf{S}) = \frac{1}{2}\|\mathbf{X}-\mathbf{AS}\|_F^2 + \frac{\tau}{2}\det^2(\mathbf{Z}) \tag{7}$$

with $\tau = \frac{\lambda}{(c-1)!}$.

2.2. Optimization Algorithm

Minimizing the objective function in Eq. 7 with respect to both $\mathbf{A}$ and $\mathbf{S}$ is a combinatorial optimization problem. We here resort to the *alternating non-negative least squares* technique [8, 9]. This technique treats the original optimization problem as two sub-problems with the iterative update,

$$\mathbf{S}^{k+1} = \arg\min_{\mathbf{S}} f(\mathbf{A}^k,\mathbf{S}),\ \mathbf{A}^{k+1} = \arg\min_{\mathbf{A}} f(\mathbf{A},\mathbf{S}^{k+1})$$

that is, we alternatively update one matrix, holding the other one fixed. The initial $\mathbf{A}$ is formed by randomly choosing c points from the given data as its columns, and the initial $\mathbf{S}$ is selected as a zero matrix.

To incorporate the non-negative constraint, we adopt a projective method, which follows the standard update rule. If the new estimate is outside the feasible set, a projective function is used to project the point back to the feasible region. For the non-negative constraint, the following learning rule will be used,

$$\mathbf{A}^{k+1} = \max(0, \mathbf{A}^k + \alpha^k\mathbf{D}_A^k), \mathbf{S}^{k+1} = \max(0, \mathbf{S}^k + \beta^k\mathbf{D}_S^k)$$

where α^k and β^k denote the stepsize, and $\mathbf{D}_A^k$ and $\mathbf{D}_S^k$ are the descent directions selected using the conjugate gradient method [10]. The central idea of conjugate direction stems from the problem associated with steepest descent, which often takes a step almost in the same direction as the earlier step. By taking conjugate directions, this drawback can be avoided. The descent direction $\mathbf{D}_A^k$ is given by

$$\mathbf{D}_A^k = -\nabla_A f(\mathbf{A}^k,\mathbf{S}) + \zeta^k\mathbf{D}_A^{k-1} \tag{8}$$

with ζ^k calculated based on

$$\zeta^k = \frac{\nabla_A f(\mathbf{A}^k, \mathbf{S})^T \left(\nabla_A f(\mathbf{A}^k, \mathbf{S}) - \nabla_A f(\mathbf{A}^{k-1}, \mathbf{S})\right)}{\nabla_A f(\mathbf{A}^{k-1}, \mathbf{S})^T \nabla_A f(\mathbf{A}^{k-1}, \mathbf{S})}$$

The initial conjugate direction is simply chosen as the negative gradient $\mathbf{D}_A^0 = -\nabla_A f(\mathbf{A}^0, \mathbf{S})$. To make sure the generated search direction is conjugate to the previous direction, the stepsize has to be carefully selected. The line minimization rule is normally used, that is,

$$\alpha^k = \arg\min_{\alpha} f(\mathbf{A}^k + \alpha \mathbf{D}_A^k, \mathbf{S}) \tag{9}$$

The descent direction $\mathbf{D}_S^k$ and the stepsize β^k are selected in the same ways as in Eqs.8 and 9. The derivative $\nabla_S f(\mathbf{A}, \mathbf{S})$ and $\nabla_A f(\mathbf{A}, \mathbf{S})$ can be derived as

$$\nabla_S f(\mathbf{A}, \mathbf{S}) = \mathbf{A}^T(\mathbf{AS} - \mathbf{X})$$
$$\nabla_A f(\mathbf{A}, \mathbf{S}) = (\mathbf{AS} - \mathbf{X})\mathbf{S}^T + \tau \det^2(\mathbf{Z})\mathbf{U}\mathbf{B}^T(\mathbf{Z}^{-1})^T$$

To take care of the sum-to-one constraint, we augment the matrices $\mathbf{X}$ and $\mathbf{A}$ by a row of constant, denoted by

$$\tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X} \\ \delta \mathbf{1}_n^T \end{bmatrix}, \quad \tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A} \\ \delta \mathbf{1}_c^T \end{bmatrix} \tag{10}$$

where δ is a positive number to control the effect of the sum-to-one constraint. The learning of abundance $\mathbf{S}$ takes these two augmented matrices as inputs. The resulting estimate will approach the constraint as δ increases.

3. EXPERIMENTAL RESULTS

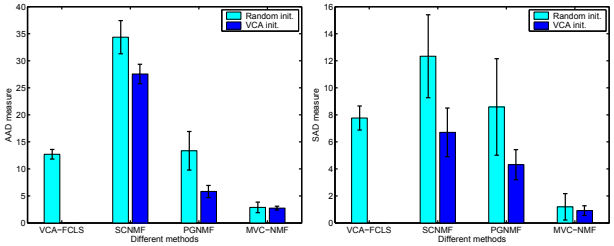
In this section, we evaluate the proposed MVC-NMF using both synthetic images and a real hyperspectral scene. The performance comparison is conducted from two perspectives, the evaluation of the identified endmember signatures and the evaluation of the estimated abundances. We use two metrics, spectral angle distance (SAD) and abundance angle distance (AAD), expressed as

$$SAD = \cos^{-1}\left(\frac{\mathbf{a}^T\hat{\mathbf{a}}}{\|\mathbf{a}\|\|\hat{\mathbf{a}}\|}\right), \; AAD = \cos^{-1}\left(\frac{\mathbf{s}^T\hat{\mathbf{s}}}{\|\mathbf{s}\|\|\hat{\mathbf{s}}\|}\right)$$

to measure the shape similarity between the estimated and the actual endmembers and abundances, respectively.

3.1. Evaluation with Synthetic Images

The proposed method is compared with two NMF-based approaches, the smoothness constrained NMF (SCNMF) [8] and the projected gradient NMF (PGNMF) [9], and one convex geometry-based method based on the pure pixel assumption, vertex component analysis (VCA) [5]. Since VCA only identifies endmembers, to find the abundances, we adopt the popular fully constrained least squares method (FCLS) [3]. The synthetic images are generated using the spectral reflectances selected from the USGS digital spectral library [11]. We divide the entire image into units of 8×8 blocks. The pixels within each block are pure and have the same type of cover, randomly selected as one of the endmember classes. The resulting image is then degraded by a spatial low pass filter to simulate an image with mixed pixels. To further remove pure pixels, we replace all the pixels whose abundance is larger than 80% with a mixture made up of all endmembers of equal abundances; that is, each endmember has an abundance of $\frac{1}{c}$ in the mixture. To simulate possible errors and sensor noise, zero mean white Gaussian noise is added to the mixture data.



(a) (b)

Fig. 1. Performance comparison with random and VCA initializations (SNR = 20dB) (a) AAD (b) SAD

As mentioned earlier, unconstrained NMF methods are sensitive to initializations. To study the effectiveness of the proposed volume constraint, we use two methods for the endmember initialization, i.e., random initialization and the use of the endmembers identified by VCA. The learning is terminated when the iteration number of successive increase of the objective value is greater than 5, or the maximum iteration number, 100, is reached. Considering the low convergence of SCNMF, we allow the maximum number of iteration to be 300. The regularization parameter of MVC-NMF is selected as $\tau = 0.015$.

Fig. 1 displays the mean (bar) and the standard deviation (error bar) of 20 random tests in terms of SAD and AAD, respectively. One immediate observation is that no matter which initialization method is used, the proposed MVC-NMF always generates the smallest means and standard deviations, indicating that this approach produces the most accurate and stable estimates. In addition, PGNMF outperforms VCA-FCLS when initialized with the endmembers estimated by VCA, which however, is not the case when using random initializations. Another important observation is that MVC-NMF is less sensitive to the selection of initial points compared to SCNMF and PGNMF. This observation leads to the conclusion that the volume constraint effectively confines the solution space and converts the original ill-posed problem to a well-posed one. Fig. 2 displays the simplex volume as the learning progresses. The constant profile corresponds to the result calculated using the real endmember set. It can be seen that during the first few iterations, both PGNMF and

MVC-NMF expand the simplex rapidly. After that, the introduced volume constraint in MVC-NMF effectively confines the simplex volume to be close to the true value. However, the PGNMF learning keeps increasing the simplex size, resulting in a much larger volume size than its actual value.

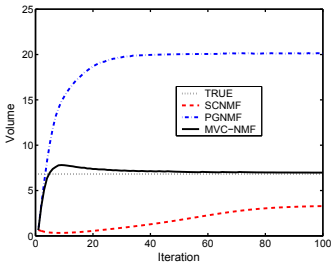


Fig. 2. Comparison of the simplex volume.

3.2. Evaluation with Real AVIRIS Data

In this experiment, we use the hyperspectral data captured by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over Cuprite, Nevada, as the test site. This site has many well exposed minerals, and some of them are prevalent, while others are highly mixed in a small set of pixels. The test image is a subscene containing 100 lines and 100 pixels per line. To improve the unmixing performance, we have removed the low SNR bands as well as the water vapor absorption bands (including bands 1-2, 104-113, 148-167 and 221-224) from the original 224-band data cube. A total of 188 bands are used in the experiment.

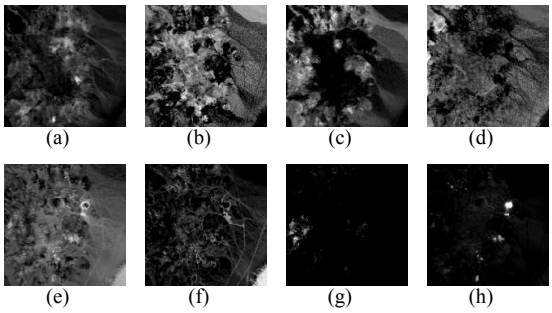


Fig. 3. Abundance maps of different minerals. (a) Alunite (b) Chalcedony (c) Kaolinite (d) Sphene (e) Nontronite (f) Montmorillonite (g) Buddingtonite (h) Muscovite

Fig. 3 illustrates the estimated abundance maps. It can be seen that these estimations present high level of similarity to the published results [5, 12]. The last two maps are of particular interest, in which the abundance maps present positive values in a small area and zero everywhere else. This observation confirms that MVC-NMF is able to identify highly mixed endmembers, even though they are only present in a small set of pixels.

4. CONCLUSIONS

This paper addressed an unmixing method for highly mixed hyperspectral data without the pure pixel assumption. The experimental results with synthetic mixtures showed that the introduced volume constraint effectively stabilizes the solution and results in more accurate estimates. The proposed MVC-NMF outperforms three advanced approaches compared. The evaluation using a real hyperspectral scene collected by the AVIRIS sensor indicated that MVC-NMF has the potential of identifying highly mixed endmembers.

5. REFERENCES

[1] N. Keshava, "A survey of spectral unmixing algorithms," *Lincoln Laboratory Journal*, vol. 14, no. 1, pp. 55–78, 2003.

[2] J. M. P. Nascimento and J. M. B. Dias, "Does independent component analysis play a role in unmixing hyperspectral data," *IEEE Trans. Geosci. Remote Sensing*, vol. 43, no. 1, pp. 175–184, Jan. 2004.

[3] D. C. Heinz and C.-I Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, no. 3, pp. 529–545, Mar. 2001.

[4] J. W. Boardman, F. A. Kruse, and R. O. Green, "Mapping target signatures via partial unmixing of aviris data," in *Summaries of Fifth Annual JPL Airborne Geoscience Workshop*, R. O. Green, Ed., 1995, vol. 1, pp. 23–26.

[5] J. M. P. Nascimento and J. M. B. Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sensing*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

[6] D. Lee and S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems*. 2001, vol. 13, pp. 556–562, MIT Press.

[7] B. Wang, H. Zhou, and L. Zhang, "Blind decomposition of mixed pixels using constrained non-negative matrix factorization," in *Proc. IEEE International on Geoscience and Remote Sensing Symposium*, July 2005, vol. 6, pp. 3757 – 3760.

[8] V. P. Paura, J. Piper, and R. J. Plemmons, "Nonnegative matrix factorization for spectral data analysis," *To appear in Linear Algebra and Applications*, 2006.

[9] C. J Lin, "Projected gradient methods for non-negative matrix factorization.," Technical report, Department of Computer Science, National Taiwan University, 2005.

[10] M. R. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," *J. Research Nat. Bur. Standards*, vol. 49, pp. 409–436, 1952.

[11] R. N. Clark, G. A. Swayze, A. Gallagher, T. V. King, and W. M. Calvin, "The u.s. geological survey digital spectral library: Version 1:0.2 to 3.0 μm," Open file rep. 93-592, U.S. Geol. Surv., 1993.

[12] G. Swayze, *The hydrothermal and structural history of the Cuprite Mining District, southwestern Nevada: An integrated geological and geophysical approach*, Phd dissertation, University of Colorado, Boulder, 1997.