

ENDMEMBER EXTRACTION OF HIGHLY MIXED DATA USING L_1 SPARSITY-CONSTRAINED MULTILAYER NONNEGATIVE MATRIX FACTORIZATION

H. Fang¹, A. H. Li^{1,*}, H. X. Xu², T. Wang¹, K. Jiang¹, Y. Z. Su¹, H. W. Chang¹

¹ The Xi'an Research Institute of High Technology, 710025 Xi'an, China - (1037723204, 1418525651)@qq.com, (fanghao_502, tech_commu, changhongweiathjj)@163.com, jugglerchn@hotmail.com

² Dept. of Electronic Information, Huanggang Normal University, 438600 Huanggang, China - 2026585882@qq.com

Commission III, WG III/4

KEY WORDS: Hyperspectral Imagery, Nonnegative Matrix Factorization, Multilayer Nonnegative Matrix Factorization, Sparsity Constraint, Endmember Extraction

ABSTRACT:

Due to the limited spatial resolution of remote hyperspectral sensors, pixels are usually highly mixed in the hyperspectral images. Endmember extraction refers to the process identifying the pure endmember signatures from the mixture, which is an important step towards the utilization of hyperspectral data. Nonnegative matrix factorization (NMF) is a widely used method of endmember extraction due to its effectiveness and convenience. While most NMF-based methods have single-layer structures, which may have difficulties in effectively learning the structures of highly mixed and complex data. On the other hand, multilayer algorithms have shown great advantages in learning data features and been widely studied in many fields. In this paper, we presented a L_1 sparsity-constrained multilayer NMF method for endmember extraction of highly mixed data. Firstly, the multilayer NMF structure was obtained by unfolding NMF into a certain number of layers. In each layer, the abundance matrix was decomposed into the endmember matrix and abundance matrix of the next layer. Besides, to improve the performance of NMF, we incorporated sparsity constraints to the multilayer NMF model by adding a L_1 regularizer of the abundance matrix to each layer. At last, a layer-wise optimization method based on NeNMF was proposed to train the multilayer NMF structure. Experiments were conducted on both synthetic data and real data. The results demonstrate that our proposed algorithm can achieve better results than several state-of-art approaches.

1. INTRODUCTION

Hyperspectral remote sensing images generally contain abundant spatial and spectral information of the covered areas, which can be useful in the applications of earth monitoring, land cover classification, mineral exploration, military surveillance, etc. (Tong et al., 2016). These images are usually captured by space-borne sensors or air-borne sensors. Due to the long observation distance and the low spatial resolution of hyperspectral sensors, pixels in the acquired hyperspectral images are usually a mixture of several ground cover spectrum. These pure spectra are known as endmembers and their proportions in the pixels are called the corresponding abundance fractions (Zhu et al., 2014). Before utilizing the hyperspectral images, we need to decompose these mixed pixels into a set of endmembers and their corresponding abundances first. This process is called hyperspectral unmixing, which contains two procedures: endmember extraction and abundance estimation (Miao et al., 2007).

Many algorithms have been proposed for endmember extraction, such as Vertex component analysis (VCA) (Nascimento et al., 2005). With the pure-pixel assumption, VCA projects all the pixels into a random direction and regard the pixel with the largest projection as endmember. Recently NMF has been widely studied. The major advantages of NMF are that it does not need to assume the existence of pure pixels in the

hyperspectral images and can decompose the target matrix into two nonnegative matrices simultaneously (Wang et al., 2013). However, it may has a huge solution space since the cost function of traditional NMF is nonconvex.

Many efforts have been made to improve the unmixing performance of NMF. Among them, based on the assumption that most pixels are mixtures of only a few of the endmembers in the images, sparsity constraints based on regularization methods have acquired extensive attention. Though many L_q ($0 < q < 1$) regularizers have been incorporated to generate sparse results, they may suffer from numerical stability since they are nondifferentiable at 0. The L_1 regularizer is generally the most popular choice for achieving sparse constraints due to its stability and effectiveness (Qian et al., 2011). Besides, for those highly-mixed and complex data, single-layered NMF may have poor performance since it is difficult to effectively decompose the complicated data (Cichocki et al., 2007). Multilayer structure can decompose the data layer by layer and help to extract the tiny features hidden in the data. So multilayer structure are generally more effective than single layer in learning feature representations of highly mixed data. Multilayer algorithms has been also widely studied in many areas such as image classification (Chen et al., 2015), hyperspectral unmixing (Rajabi et al., 2015), attribute representations (Trigeorgis et al., 2017), etc.

* tech_commu@163.com

In this paper, we proposed a L_1 sparsity-constrained multilayer NMF (L_1 -MLNMF) unmixing algorithm for highly mixed hyperspectral data. The multilayer NMF structure is acquired by iteratively decomposing the target matrix into the endmember matrix and the abundance matrix in each layer. Due to the intrinsic sparsity of the abundance matrix, we added sparse constraint to each layer with the L_1 regularizer of the abundance matrix, which is different from traditional multilayer NMF algorithms. Besides, the major contribution of our paper is that a novel optimization method was proposed to train the structure layer-wisely. To improve the approximation accuracy and accelerate the optimization process, NeNMF (Guan et al., 2012) was adopted to solve the NMF problem in each layer.

2. METHODOLOGY

The linear mixing model (LMM) has been widely used in hyperspectral unmixing due to its simplicity and satisfactory performance in most scenes. In this paper, the proposed algorithm is built based on LMM. So we first introduce LMM in this section. Then NeNMF is further presented, which is the foundation of our optimization procedure. At last, the formulation of our method is described.

2.1 Linear Mixing Model

For the hyperspectral data $\mathbf{X} \in R^{m \times n}$ with m bands and n pixels, suppose the number of endmembers in the image is c , then the mathematical model of LMM can be expressed as

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{E} \quad (1)$$

where $\mathbf{A} \in R^{m \times c}$ is the endmember matrix
 $\mathbf{S} \in R^{c \times n}$ is the abundance matrix
 \mathbf{E} is the noise matrix

Both the endmember matrix and the abundance matrix should satisfy some physical constraints. The elements in the endmember matrix should be nonnegative since they denote the observed energy in certain bands. The abundance matrix should subject to the following two constraints: the abundance nonnegative constraint, $s_{ij} \geq 0, \forall s_{ij} \in \mathbf{S}$, and the abundance sum-to-one constraint, $\sum_{i=1}^c s_{ij} = 1$.

2.2 NeNMF

NeNMF is an efficient NMF solver based on Nesterov's optimal gradient method (Nesterov, 2005), which has been proven to be superior to the multiplicative update rule (MUR) and the projected gradient method (PG) in terms of efficiency as well as approximation accuracy.

A common cost function of NMF can be written as

$$\min \frac{1}{2} \|\mathbf{X} - \mathbf{W}\mathbf{H}\|_F^2 \quad (2)$$

where $\mathbf{X} \in R^{m \times n}$ is the target matrix
 $\mathbf{W} \in R^{m \times c}$ and $\mathbf{H} \in R^{c \times n}$ are the two decomposed nonnegative matrices

Since (2) is nonconvex, it is impractical to obtain the optimal solution. The block coordinate descent method is usually used to solve it by the following two sub-problems

$$\mathbf{H}^{t+1} = \arg \min_{\mathbf{H} \geq 0} F(\mathbf{W}^t, \mathbf{H}) = \frac{1}{2} \|\mathbf{X} - \mathbf{W}^t \mathbf{H}\|_F^2 \quad (3)$$

$$\mathbf{W}^{t+1} = \arg \min_{\mathbf{W} \geq 0} F(\mathbf{W}, \mathbf{H}^{t+1}) = \frac{1}{2} \|\mathbf{X} - \mathbf{W} \mathbf{H}^{t+1}\|_F^2 \quad (4)$$

where t is the iteration counter

In the method, it alternatively solve the above two sub-problems until convergence. NeNMF was built based on the block coordinate descent method, which uses Nesterov's optimal gradient algorithm to alternatively minimize them. However, before Nesterov's method can be used, the target function is required to be convex and the gradient of the function should be Lipschitz continuous. Fortunately, both (3) and (4) meet these requirements. They can be solved similarly since they are symmetric. Take (3) for example, the method construct two sequences $\{\mathbf{Y}_k\}$ and $\{\mathbf{H}_k\}$, and alternatively update them in each iteration as follows:

$$\mathbf{H}_k = \arg \min_{\mathbf{H} \geq 0} \{ \phi(\mathbf{Y}_k, \mathbf{H}) = F(\mathbf{W}^t, \mathbf{Y}_k) + \langle \nabla_{\mathbf{H}} F(\mathbf{W}^t, \mathbf{Y}_k), \mathbf{H} - \mathbf{Y}_k \rangle + \frac{LC}{2} \|\mathbf{H} - \mathbf{Y}_k\|_F^2 \} \quad (5)$$

$$\mathbf{Y}_{k+1} = \mathbf{H}_k + \frac{\alpha_k - 1}{\alpha_{k+1}} (\mathbf{H}_k - \mathbf{H}_{k-1}) \quad (6)$$

$$\alpha_{k+1} = \frac{1 + \sqrt{4\alpha_k^2 + 1}}{2} \quad (7)$$

where \mathbf{Y}_k is the k th search point
 \mathbf{H}_k is the k th approximate solution
 $\phi(\mathbf{Y}_k, \mathbf{H})$ is the approximation of $F(\mathbf{W}^t, \mathbf{H})$ on \mathbf{Y}_k
 LC is the corresponding Lipschitz constant
 α_k is the combination coefficient

By using the Lagrange multiplier method and the Karush-Kuhn-Tucker (K.K.T.) conditions of (5), it can be solved as

$$\mathbf{H}_k = P \left(\mathbf{Y}_k - \frac{1}{LC} \nabla_{\mathbf{H}} F(\mathbf{W}^t, \mathbf{Y}_k) \right) \quad (8)$$

where $P(\bullet)$ can project all the negative elements to zero

By alternatively updating the two sequences, \mathbf{H}_k can be optimized towards the direction of reducing (3). Until convergence, the optimal solution of (3) can be obtained. The above procedure can be used to solve (4) in the same way. Using Nesterov's optimal gradient method to alternatively minimize (3) and (4), it can reach the local optimum of (2).

2.3 The Proposed L_1 -MLNMF Method

In L_1 -MLNMF, the proposed multilayer NMF structure is used to decompose the hyperspectral data \mathbf{X} . In the first layer, the

data is decomposed into the basic endmember matrix and the abundance matrix, namely \mathbf{W}_l and \mathbf{H}_l . For the l layer, the data matrix \mathbf{X}_l which is directly obtained from \mathbf{H}_{l-1} , is decomposed into \mathbf{W}_l and \mathbf{H}_l . Suppose the number of layers is L . So the mathematical representation of the multilayer structure can be expressed as

$$\mathbf{X} \approx \mathbf{W}_1 \mathbf{W}_2 \cdots \mathbf{W}_L \mathbf{H}_L \quad (9)$$

Then sparsity constraint is added to the model with the L_l regularizer of the abundance matrix in each layer. Then the total cost function can be constructed as

$$\mathbf{F} = \frac{1}{2} \|\mathbf{X} - \mathbf{W}_1 \mathbf{W}_2 \cdots \mathbf{W}_L \mathbf{H}_L\|_F^2 + \sum_{l=1}^L \mu_l \|\mathbf{H}_l\| \quad (10)$$

where μ_l is the penalty factor in layer l .

The penalty factor μ_l is used to estimate the sparsity of the abundance matrix in the l th layer, which is calculated using the estimator (Hoyer, 2004) as

$$\mu_l = \frac{1}{\sqrt{c}} \sum_{i=1}^c \frac{\sqrt{n} - \|\mathbf{X}_{l,i}\|_1 / \|\mathbf{X}_{l,i}\|_2}{\sqrt{n-1}} \quad (11)$$

where $\mathbf{X}_{l,i}$ is the i th band of \mathbf{X}_l .

The function in (10) is a multi-factor NMF problem. Traditional algorithms usually decompose it into two-factor NMFs, which is easy to accumulate cumulative error. In this paper, we solve it layer-by-layer. In l th layer, the cost function can be expressed as

$$F(\mathbf{W}_l, \mathbf{H}_l) = \frac{1}{2} \|\mathbf{X} - \boldsymbol{\Phi}_{l-1} \mathbf{W}_l \mathbf{H}_l\|_F^2 + \mu_l \|\mathbf{H}_l\| \quad (12)$$

where $\boldsymbol{\Phi}_{l-1} = \mathbf{W}_1 \mathbf{W}_2 \cdots \mathbf{W}_{l-1}$.

Here we use NeNMF to optimize (12). The gradients of $F(\mathbf{W}_l, \mathbf{H}_l)$ and corresponding Lipschitz constants can be calculated as

$$\nabla_{\mathbf{W}_l} F(\mathbf{W}_l, \mathbf{H}_l) = \boldsymbol{\Phi}_{l-1}^T \boldsymbol{\Phi}_{l-1} \mathbf{W}_l \mathbf{H}_l \mathbf{H}_l^T - \boldsymbol{\Phi}_{l-1}^T \mathbf{X} \mathbf{H}_l^T \quad (13)$$

$$\nabla_{\mathbf{H}_l} F(\mathbf{W}_l, \mathbf{H}_l) = \mathbf{W}_l^T \boldsymbol{\Phi}_{l-1}^T \boldsymbol{\Phi}_{l-1} \mathbf{W}_l \mathbf{H}_l - \mathbf{W}_l^T \boldsymbol{\Phi}_{l-1}^T \mathbf{X} + \mu_l \mathbf{E} \quad (14)$$

$$LC_{\mathbf{W}_l} = \|\boldsymbol{\Phi}_{l-1}^T \boldsymbol{\Phi}_{l-1}\|_2 \|\mathbf{H}_l^T \mathbf{H}_l\|_2 \quad (15)$$

$$LC_{\mathbf{H}_l} = \|\mathbf{W}_l^T \boldsymbol{\Phi}_{l-1}^T \boldsymbol{\Phi}_{l-1} \mathbf{W}_l\|_2 \quad (16)$$

Several issues should be considered in the optimization process. Firstly, we adopt VCA (Nascimento et al., 2002) and the fully constrained least squares method (FCLS) (Heinz et al., 2001) to estimate endmembers and abundance fractions for all layers, separately. Besides, for the abundance sum-to-one constraint, an effective method is adopted to augment the target matrix and the endmember matrix as

$$\widetilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X} \\ \delta \mathbf{1}^T \end{bmatrix} \quad \widetilde{\mathbf{W}} = \begin{bmatrix} \mathbf{W} \\ \delta \mathbf{1}^T \end{bmatrix} \quad (17)$$

where δ is a positive constant.

At last, the optimization will be stopped when the number of iterations exceeds the maximum number of iterations T_{\max} or the cost function changes less than the threshold, set to 10^{-5} in this paper, for more than 20 successive times. Our proposed algorithm can be summarized as in Algorithm 1.

Algorithm 1: L_l -MLNMF

Input: $\mathbf{X} \in \mathbf{R}^{m \times n}$, c , δ , L , T_{\max}

Output: $\mathbf{W} \in \mathbf{R}^{m \times c}$, $\mathbf{H} \in \mathbf{R}^{c \times n}$

$\mathbf{X}_1 = \mathbf{X}$

for all layers **do**

 initialize \mathbf{W}_l , \mathbf{H}_l using VCA and FCLS

repeat

 calculate μ_l using (11)

$\boldsymbol{\Phi}_{l-1} = \mathbf{W}_1 \mathbf{W}_2 \cdots \mathbf{W}_{l-1}$

 update \mathbf{W}_l with Nesterov's method

$\boldsymbol{\Phi}_l = \boldsymbol{\Phi}_{l-1} \mathbf{W}_l$

 augment \mathbf{X} , $\boldsymbol{\Phi}_l$ similarly using (17)

 update \mathbf{H}_l with Nesterov's method

until stopping criterion is satisfied

$\mathbf{X}_{l+1} = \mathbf{H}_l$

end for

$\mathbf{W} = \mathbf{W}_1 \mathbf{W}_2 \cdots \mathbf{W}_L$, and $\mathbf{H} = \mathbf{H}_L$

3. EXPERIMENTS AND RESULTS

In this section, several experiments are conducted on both synthetic data and real data to compare our method with several existing state-of-art approaches, namely the vertex component analysis method (VCA-FCLS), the projected gradient NMF method (PGNMF) (Lin, 2007), the minimum volume constrained NMF method (MVC-NMF) (Miao et al., 2007), and the multilayer NMF method (MLNMF) (Rajabi et al., 2015). Spectra from the USGS spectral library are used to generate the synthetic data with the method proposed by (Miao et al., 2007). To get highly-mixed data, the number of extracted spectra is set to 6 and the size of the low pass filter is set to 9×9 . The parameter concerning the mixing degree of the data, η , is set to 0.8 to further remove pure pixels. 20-dB white Gaussian noise is also added to the data.

In our method, we set $L = 10$, $\delta = 20$, and $T_{\max} = 1000$. The parameters of the other four approaches are consistent with their original work, except that the layer of MLNMF is kept the same with our method. Four widely used metrics, including the spectral angle distance (SAD), the spectral information divergence (SID), the abundance angle distance (AAD) and the abundance information divergence (AID) are used to evaluate the performance of the methods. Two experiments are conducted on the synthetic data. The metrics in each experiment is acquired by 20 random tests.

3.1 Overall Comparison of Different Methods

In this experiment, we set $\eta = 0.6$ to generate higher mixing degree of the synthetic data. Figure 1 shows the averaging experimental results. As can be seen, L_1 -MLNMF can generate the smallest metrics among all the five tested algorithms in terms of AAD, AID, SAD, and SID. In particular, it can also achieve better performance than MLNMF since they both have multilayer structures. This is mainly owing to the carefully-designed multilayer model and the effective optimization method. In general, our proposed method is superior to the other four approaches.

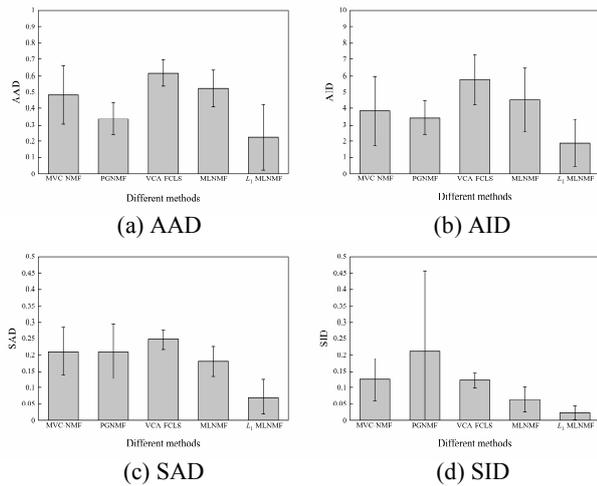


Figure 1. Performance comparison of different algorithms

3.2 Performance Analysis of Different Number of Layers

In this experiment, the effect of different number of layers on L_1 -MLNMF is analysed. η is remained to be 0.8. Figure 2 shows the calculated results. It can be seen that all the four metrics decrease with the number of layers increases. Especially, when the number of layer is less than 3, all the metrics have a big decline. When the number of layers is close to 10, the metrics decrease slowly, demonstrating that the unmixing effect is reaching its limitation. In general, more layers can help to improve the unmixing performance in a certain range.

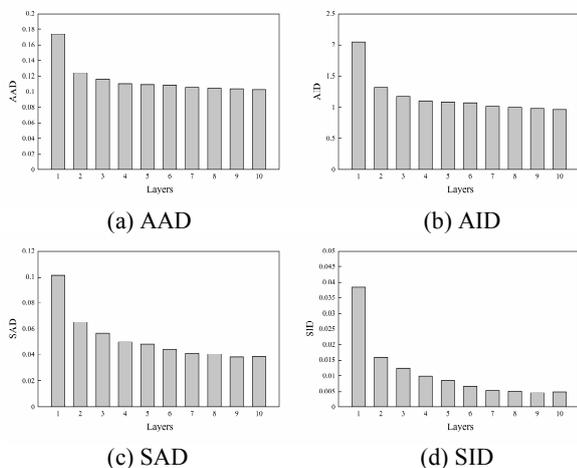


Figure 2. Performance comparison of L_1 -MLNMF under different number of layers

3.3 Performance Comparison on Real Data

In this experiment, our method is tested on the Cuprite dataset, which has 224 bands and a spectral resolution of 10 nm, covering the wavelength range from 0.4 to 2.5 μm . Figure 3 shows the 30th band of the sub image used in this experiment. After removing the low SNR and water-vapor absorption bands (including bands 1-3, 105-113, 148-167, and 221-224), a total of 188 bands are used in this experiment. The number of endmembers is defined as 12, according to the study in (Nascimento et al., 2002). Table 1 summarizes the comparison results of the five algorithms. Figure 4 shows the estimated abundance maps by our proposed method. It can be seen that our method could achieve smaller SAD than other methods and the estimated abundance maps presented high similarity to the published results.

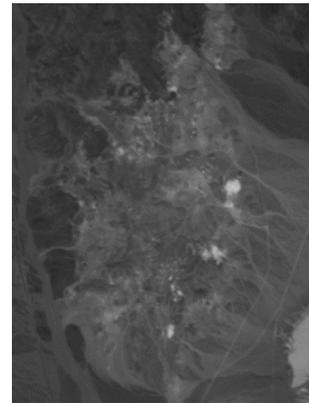


Figure 3. Band 30 of the experimental subimage

	L_1 - MLNMF	MVC- NMF	PGNM F	VCA- FCLS	MLN MF
Alunite	0.0807	0.0806	0.0954	0.0723	0.1234
Nontronite#1	0.0827	0.1452	0.1132	0.0717	0.1409
Buddingtonite	0.1188	0.0860	0.2090	0.0690	0.1030
Montmorillonite	0.0551	0.0923	0.1217	0.0727	0.0992
Kaolinite	0.2299	0.2097	0.2536	0.2148	0.2573
Chalcedony	0.0975	0.1456	0.2384	0.1391	0.1476
Pyrope	0.0749	0.0976	0.0876	0.1145	0.1041
Dumortierite	0.1092	0.1404	0.0843	0.0658	0.0951
Muscovite	0.0744	0.0842	0.0824	0.0783	0.0816
Andradite	0.0687	0.0957	0.2259	0.1025	0.0676
Nontronite#2	0.1086	0.0992	0.1257	0.1175	0.1135
Sphene	0.0521	0.0526	0.0648	0.1267	0.1458
Mean	0.0960	0.1810	0.1418	0.1037	0.1233

TABLE 1. Comparison results of different methods in terms of SAD

4. CONCLUSIONS

For the endmember extraction of highly mixed data, we have proposed a L_1 sparsity-constrained multilayer NMF method in this paper. The proposed multilayer NMF model is formulated by unfolding NMF into a certain number of layers and adding the L_1 regularization terms of the abundance matrices. The cost function is built on all the matrices in each layer. And a novel optimization method based on NeNMF is further presented to solve the multi-factor NMF problem, since it is difficult to simultaneously optimize all the endmember matrices and abundance matrices. The experimental results demonstrates that when the number of layers is less than 10, more layers can help

to improve the unmixing performance. Due to the carefully-designed multilayer structure, our method can also achieve better performance than several other state-of-art unmixing approaches on both synthetic data and real data.

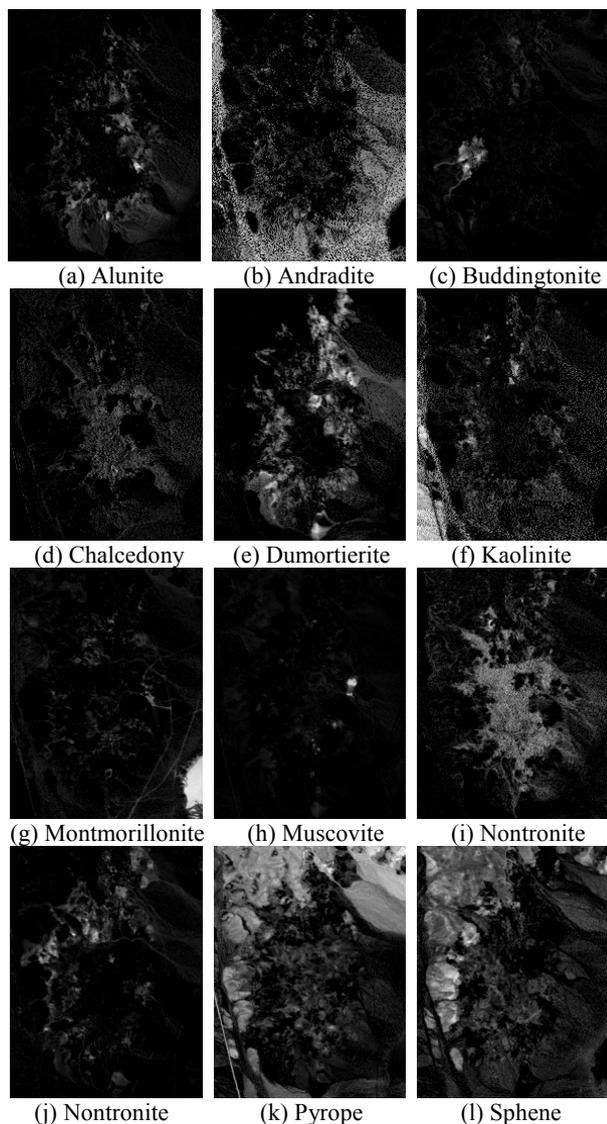


Figure 4. The estimated abundance maps by L_1 -MLNMF

REFERENCES

- Chen Y., Zhao X., Jia X., 2015. Spectral-spatial classification of hyperspectral data based on deep belief network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6), pp. 2381-2392.
- Cichocki, A., and R. Zdunek., 2007. Multilayer nonnegative matrix factorization using projected gradient approaches. *International Journal of Neural Systems*, 17(6), pp. 431-446.
- Guan, N., Tao D., Luo Z. et al., 2012. NeNMF: an optimal gradient method for nonnegative matrix factorization. *IEEE Transactions on Signal Processing*, 60(6), pp. 2882-2898.
- Heinz, D. C. and Chang C.-I., 2001. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 39(3), pp. 529-545.
- Hoyer, P. O., 2004. Non-negative matrix factorization with sparseness constraints. *Journal of machine learning research*, 5, pp. 1457-1469.
- Lin, C.-J., 2007. Projected gradient methods for non-negative matrix factorization. *Neural Computation*, 19(10), pp. 2756-2779.
- Miao, L., H. Qi., 2007. Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization. *IEEE Transactions on Geoscience and Remote Sensing*, 45(3), pp. 765-777.
- Nascimento J. M. P., Dias J. M. B., 2005. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(4), pp. 898-910.
- Nesterov, Y., 2005. Smooth minimization of non-smooth functions. *Mathematical Programming*, 103(1), pp. 127-152.
- Qian, Y., Jia S., Zhou J., et al., 2011. Hyperspectral Unmixing via $L_{1/2}$ Sparsity-Constrained Nonnegative Matrix Factorization, *IEEE Transactions on Geoscience and Remote Sensing*, 49(11), pp. 4282-4297.
- Rajabi R. and Ghassemian H., 2015. Spectral unmixing of hyperspectral imagery using multilayer NMF. *IEEE Geoscience and Remote Sensing Letters*, 12(1), pp. 38-42.
- Tong L., Zhou J., Qian Y., et al., 2016. Nonnegative-Matrix-Factorization-Based Hyperspectral Unmixing With Partially Known Endmembers. *IEEE Transactions on Geoscience and Remote Sensing*, 54(11), pp. 6531-6543.
- Trigeorgis G., Bousmalis K., Zafeiriou S., et al., 2017. A deep matrix factorization method for learning attribute representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(3), pp. 417-429.
- Wang, N., Du B., and Zhang L. P., 2013. An Endmember Dissimilarity Constrained Non-Negative Matrix Factorization Method for Hyperspectral Unmixing. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(2), pp. 554-569.
- Zhu, F. Y., Wang Y., Xiang S. M., et al., 2014. Structured Sparse Method for Hyperspectral Unmixing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 88, pp. 101-118.