

Sparsity-Constrained NMF Algorithm Based on Evolution Strategy for Hyperspectral Unmixing

Ning ShangBin^{1,a}, Zuo FengChao¹

¹The School of Computer Science, Liaocheng University, 252000 LiaoCheng, Shan Dong, China

Abstract. As a powerful and explainable blind separation tool, non-negative matrix factorization (NMF) is attracting increasing attention in Hyperspectral Unmixing(HU). By effectively utilizing the sparsity priori of data, sparsity-constrained NMF has become a representative method to improve the precision of unmixing. However, the optimization technique based on simple multiplicative update rules makes its unmixing results easy to fall into local minimum and lack of robustness. To solve these problems, this paper proposes a new hybrid algorithm for sparsity constrained NMF by intergrating evolutionary computing and multiplicative update rules (MURs). To find the superior solution in each iteration, the proposed algorithm effectively combines the MURs based on alternate optimization technique, the coefficient matrix selection strategy with sparsity measure, as well as the global optimization technique for basis matrix via the differential evolution algorithm. The effectiveness of the proposed method is demonstrated via the experimental results on real data and comparison with representative algorithms.

1 Introduction

With the development of satellite technology, computer technology and sensor technology, hyperspectral remote sensing technology has made great progress, playing an increasingly important role in agriculture, meteorology, geology, military and other fields. The hyperspectral image produced by it has rich spectral and ground space information, but due to the complexity of the surface and the spatial resolution of the hyperspectral imager, a single pixel is often a multi-substance in a real hyperspectral image. In a real hyperspectral image, a single pixel is often a mixture of spectral properties' multiple substances, which is a mixed pixel. (1) Decompose it into the constituents of different substances (Endmember) and their proportion in this pixel (Abundance). This decomposition process is called hyperspectral image mixed pixel decomposition, referred to as HU. It can be divided into a Linear Mixed Model (LMM) and a Nonlinear Mixed Model (NLMM). (10) Each pixel in the LMM is a linear combination of the endmember spectrum and its abundance. Compared with the NLMM construction model, the physical meaning is easy to understand, so it becomes a widely used model.

In the existing unmixing methods, in general, it can be summarized into two categories: one type of algorithm is based on geometric methods, such as pixel purity index (PPI)(3), vertex component analysis (VCA)(4), and simplex growing algorithm. (SGA)(5), etc., such algorithms are based on the assumption that each end

element exists at least once, but in many hyperspectral data, the spatial resolution of the data is very low, this assumption may not hold; Another type of algorithm is based on statistics, the method considers the hyperspectral image unmixing problem as blind signal separation (BSS) problem, mainly including independent component analysis (ICA)(6) and NMF, due to the non-negative and Abundance Sum-to-One Constraint(ASC) constraint of hyperspectral data does not satisfy ICA. The method's original signal independence hypothesis limits the application of the ICA method. NMF does not require the existence of pure pixels in the image. The good interpretability of its partial composition is in line with the intuitive cognitive understanding, and the endmember and abundance can be estimated at the same time, making NMF very suitable for hyperspectral unmixing. Since the basic NMF does not make full use of the sparsity prior of the data, a non-negative matrix factorization method for increasing the desaturation constraint of the L0, L1, and L1/2 regularization terms for the abundance matrix is generated to improve the demixing precision, where L1/2 NMF method of sparsity constraint is a representative method to improve the accuracy of unmixing.

However, the optimization technique based on simple multiplication update rule lacks robustness in practical applications and is easy to fall into local minimum solution. To solve the above problems, this paper proposes a new sparse non-negative matrix factorization optimization algorithm based on the combination of evolutionary strategy and multiplication update. The first

* Corresponding author: *zane.ningwork@gmail.com

section of this paper introduces the basic background and related knowledge of hyperspectral mixing. The second section writes the basic NMF and L1/2NMF and the corresponding initialization; the evolution strategy and the local optimization process are described in detail in the third section; the fourth section analyzes and compares the results of the experiment; the fifth section summarizes and analyzes.

2 Preliminary

The hyperspectral image data is a data cube V , A and S represent the endmember matrix and the abundance matrix, respectively, and the LMM mathematical model can be expressed as:

$$V = AS + E \quad (1)$$

Where $V \in R^{M \times N}$ represents a hyperspectral image with M bands and N pixels, and E represents possible errors and noises. Since the information of V , A , and S is unknown during the decomposition process, Hyperspectral unmixing is a blind decomposition problem.

2.1 NMF

As a statistical method based blind source separation method, NMF has been widely used in signal processing, pattern recognition, computer vision and other fields. Specifically, NMF decomposes a given non-negative matrix $X \in R^{P \times Q}$ into two non-negative matrices $W \in R^{P \times U}$ and $H \in R^{U \times Q}$. And satisfy $U < \min(P, Q)$, which is designed to minimize the reconstruction error between X and WH . The mathematical formula can be described as:

$$\min_{W, H} \|X - WH\|_F^2, s. t. W, H \geq 0 \quad (2)$$

$\|\cdot\|_F$ stands for its Frobenius norm. When the error in the LMM model is ignored, $V=AS$ is transformed into a minimum value by the NMF method. NMF is well suited for hyperspectral linear demixing and can be effectively reduced for high spatial and spectral dimensions in hyperspectral data.

NMF has many solutions, the most famous of which is the MURs(7) proposed by Lee and Seung. The iterative solution method is as follows:

$$W \leftarrow W * (XH^T ./ WHH^T) \quad (3)$$

$$H \leftarrow H * (W^T X ./ W^T W H) \quad (4)$$

$(\cdot)^T$ represents the transpose of the matrix, "*" and "./" respectively represent the multiplication and division of the matrix.

2.2 L1/2-NMF

Since the basic NMF method does not make full use of the sparsity prior of the data, the sparsity limit of the regularization method is often added to the abundance matrix in the objective function:

$$F(W, H) = \frac{1}{2} \|X - WH\|_F^2 + \lambda \|H\|_{1/2} \quad (5)$$

Where $\|H\|_{1/2} = \sum_{t,q} (h_{p,n})^{1/2}$ as the abundance matrix Sparse constraints, $\lambda \in R^+$ as a regularization parameter. When $\frac{1}{2} \leq q < 1$ in the regularization matrix $L_q (0 < q < 1)$, the sparsity of L_q decreases as q

increases. When $1 \leq q < \frac{1}{2}$, the sparsity of L_q is basically stable with the change of q (8). Therefore, consider $q=1/2$ as the parameter of the regularization constraint. The iteration rules are as follows:

$$W \leftarrow W * (XH^T ./ WHH^T) \quad (6)$$

$$H \leftarrow H * (W^T X ./ W^T W H + \frac{\lambda}{2} H^{-\frac{1}{2}}) \quad (7)$$

The stopping criterion is usually used to set the maximum number of iterations or set the tolerance threshold for the error of the objective function. The stopping method used in this paper is to set the maximum number of iterations.

2.3 Abundance Sum-to-One Constraint

A pixel can be composed of several end elements, but the sum of the abundances of the end elements is 1, that is, Sum-to-One Constraint(ASC). Since the abundance matrix always satisfies the ASC constraint according to the LMM, we follow the method in the paper. (9)Merging ASC into NMF-based unmixing. The hyperspectral image matrix X and the end element matrix W become an augmented matrix:

$$X_a \leftarrow \begin{bmatrix} X \\ \delta 1_Q^T \end{bmatrix} \quad (8)$$

$$W_a \leftarrow \begin{bmatrix} W \\ \delta 1_U^T \end{bmatrix} \quad (9)$$

δ is used to control the influence of ASC in the loss function. 1_Q and 1_U are all 1 column vectors $(1, 1, \dots, 1)^T$, and H is updated with X_a and W_a in each iteration of the algorithm.

3 Proposed method

In order to improve the accuracy of demixing and increase the robustness of the algorithm, this paper proposes a new sparse non-negative matrix factorization optimization algorithm based on evolutionary strategy and multiplication update (ie EMNMF), aiming to find the global excellent solution and improve algorithm efficiency. In order to effectively utilize the given hyperspectral image data, multiple pairs of initial test sets $S_0 = \{(W_0^i, H_0^i)\}_{i=1}^m$ are set at initialization. The initial candidate set is clustered. $\{W_0^i\}_{i=1}^m, \{H_0^i\}_{i=1}^m$ is used as the seed matrix, and m represents the m set matrix or matrix pair in the candidate set. The t -th iteration uses $S_t = \{(W_t^i, H_t^i)\}_{i=1}^{m_t}$, and m_t represents the m -group matrix pair in the candidate set at the t -th iteration.

3.1. K-Means

The K-means clustering algorithm is a NP-hard optimization problem. It is still one of the most widely used clustering algorithms due to its easy implementation, simple, efficient and successful application cases and experience.

The K-means algorithm is an iterative process in order to minimize the square sum $J(C)$ of all samples in the

cluster domain to the cluster center. The algorithm flow includes 4 steps: 1) K in the selected data space objects act as the center of the initial cluster, and each object represents the center of a category. 2) For the data objects in the sample, according to their euclidean distance from these cluster centers, they are assigned to the class represented by the most similar cluster center according to the nearest distance criterion. 3) Calculate the mean of all objects in each category as the new cluster center of the category, and calculate the square sum of the distances of all samples to the cluster center of the category, ie the $J(C)$ value. 4) See if the cluster center and $J(C)$ values have changed, and if so, return to the second part, otherwise the clustering ends. After generating the endmember matrix, the corresponding abundance matrix is generated by least squares method:

$$H = \max(0, (W^T W)^{-1} W^T X) \quad (10)$$

3.2 Multiplication rule

The multiplication rule is to use the classic MURs to gradually obtain the minimum value of the objective function in an iterative manner and minimize the reconstruction error. In the appropriate way, the endmember matrix and the abundance matrix in $S_t = \{(W_t^i, H_t^i)\}_{i=1}^{m_t}$ are brought into the following formula for iteration:

$$W \leftarrow W.*(XH^T./WHH^T) \quad (11)$$

$$H \leftarrow H.*(W^T X./W^T W H + \frac{\lambda}{2} H^{-\frac{1}{2}}) \quad (12)$$

3.3 Sparse selection

Since the differential evolution algorithm is used to improve the endmember matrix and the better candidate set is selected, the abundance matrix is improved. Since the objective function is the N1/2 sparseness constraint NMF, the candidate is the most sparse abundance matrix is selected to select and the formula for evaluating the sparsity of the abundance matrix is as follows:

$$sparsity(H) = \frac{1}{\sqrt{U}} \sum_{U=1}^U \frac{\sqrt{Q} \cdot \frac{\| [H]_{U,:} \|_1}{\| [H]_{U,:} \|_2}}{\sqrt{Q-1}} \quad (13)$$

$[H]_U$ represents the abundance matrix of the u th end element.(10)

3.4 Differential Evolution (DE)

In order to avoid the algorithm falling into the local minimum, this paper introduces the strategy of differential evolution algorithm to expand the global search ability of the algorithm. DE is a heuristic random search algorithm based on group difference, which is robust to the optimization of real valued parameters.

Steps:

1) Initialization operation: first input the population size NP , cross factor $CR \in [0, 1]$, scaling factor $F \in [0, 2]$, evolution algebra T ;

2) Individual evaluation: bring each individual $S_i(t)$ into L1/2-NMF to calculate the objective function value $f(S_i(t))$;

3) Mutation operation: for each individual in the population, three different random integers $r_1, r_2, r_3 \in \{1, 2, \dots, NP\}$ are generated, and $r_1 \neq r_2 \neq r_3$ is satisfied, and the variant individual $Temp1$ is generated. :

$$Temp1 = W^{r_1} + F.*(W^{r_2} - W^{r_3}) \quad (14)$$

4) Cross operation: randomly generate a homomorphic matrix $Temp$ ($\text{rand}[0,1]$) of the endmember matrix by $\text{rand}[0,1]$ to indicate that each element of the generated matrix $Temp$ is a random number between $[0,1]$, Then get $U_i(t)$ by the formula (15):

$$U_i(t) = \begin{cases} Temp1 & \text{if } (Temp < CR) \\ Temp & \text{otherwise} \end{cases} \quad (15)$$

Where $Temp < CR$ represents the element smaller than CR in the $Temp$ matrix.

5) Select operation:

$$S_i(t+1) = \begin{cases} S_i(t) & \text{if } (S_i(t)) < U_i(t) \\ U_i(t) & \text{otherwise} \end{cases} \quad (16)$$

EMNMF Algorithm for Hyperspectral Unmixing

Input:

The hyperspectral image $\mathbf{X} \equiv [X_1, X_2, \dots, X_Q] \in R^{P \times Q}$;

The algorithm parameter ε ;

The punishment factor δ ;

The maximum number of algorithm iterations $MaxIter$;

Output:

The abundance matrix \mathbf{H} ;

The endmember matrix \mathbf{W} ;

1 : Initialize $S_0 = \{(W_0^i, H_0^i)\}_{i=1}^m$;

2 : Initialize regularization parameter λ ;

3 : K-Means ;

4 : for $i = 1$ to iteration do

5 : L1/2-NMF ;

6 : Sparsity ;

7 : DE

8 : for $i = 1$ to NP do

9 : (Mutation and Crossover) ;

10 : end ;

11 : (Greedy Seleccion) ;

12 : **repeat**

13 : end ;

14 : **until** the algorithm stopping criteria is satisfied.

4 Experiment analysis

In order to prove the performance of the proposed algorithm, this section compares the EMNMF algorithm with some representative algorithms, and adopts the Spectral Angle Distance (SAD) and the original hyperspectral image widely used in the field of unmixing research. SAD is used to indicate the degree of difference between spectra. Small SAD value means small spectral difference, and its expression formula is:

$$SAD_p = \cos^{-1} \left(\frac{w_p^T \hat{w}_p}{\|w_p\| \|\hat{w}_p\|} \right) \quad (17)$$

W_p and \hat{W}_p represent the estimated end element and the reference end element, respectively, and $\| \cdot \|$ represents the two norm of the vector or matrix.

Experimental environment: CPU model is Core i7 4790K, clocked at 4GHz, memory 8G.

Software platform: Win10 operating system ,Matlab R2014b .

Algorithm	SAD(1)	SAD(2)	SAD(3)	SAD(4)
NMF	0.1313	0.9207	0.3872	0.7813
L1/2NMF	0.1165	0.3744	0.6324	0.7277
EMNMF	0.0500	0.0782	0.0444	0.4087

It is found that the EMNMF experiment works best through the SAD indicator.

This paper also tests the performance of the algorithm in real hyperspectral scenarios by selecting real hyperspectral urban images.

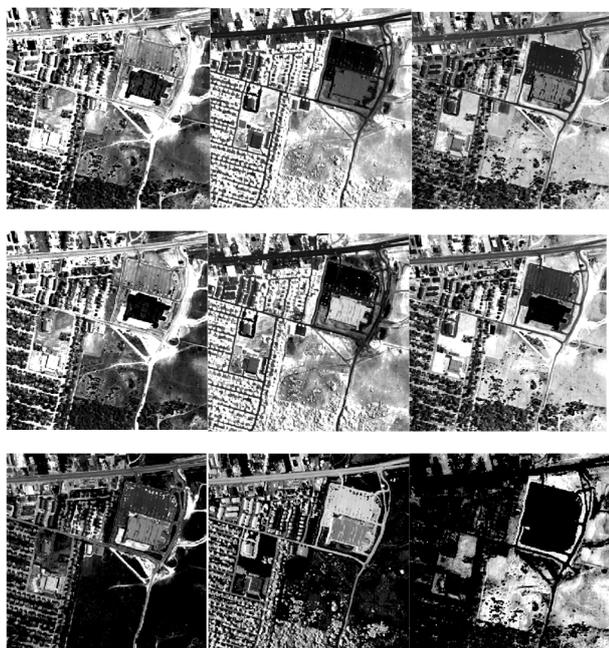


Figure 1. The urban's unmixing diagram of the 3 algorithms.

Figure 1. shows the abundance distribution of the three materials unmixed by NMF algorithm, L1/2NMF algorithm and EMNMF algorithm. By comparison, the EMNMF algorithm is improved compared to the first two algorithms.

5 Summary

The experimental design of the algorithm is not only effective on real data, but also reaches the ideal value in important experimental indicators, improves the accuracy and speed of hyperspectral demixing, finds approximate solutions better, and has better global optimality. You can also migrate your ideas to other areas of research to get more value. In addition, the algorithm has more room for improvement, and further research and improvement are still needed: in this paper, detailed research can be carried out on the parameter setting; only some actual data are tested, and more actual data can be tested to test the

robustness of the experiment; and the algorithm does not test the simulation data. In the future, simulation experiments can be added to test the performance of the algorithm in more simulation environments.

References

1. J. Bioucas-Dias *et al.*, *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp.354–379(2012)
2. N. Keshava, *LincolnLab.J.*,vol. 14, no. 1, pp. 55–77(2003)
3. J. Boardman, *AVIRISWorkshop*, vol.1,pp. 11–14. JPL Pub.93-26(1993)
4. J. Nascimento, J. Bioucas-Dias, *IEEE Trans. Geosci.Remote Sens.*, vol. 43, no. 4, pp. 898–910(2005)
5. C.I. Chang, C.C. Wu, W. Liu, and Y.C. Ouyang,*IEEE Trans.Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2804–2819, Oct(2006)
6. J. Nascimento , J. Bioucas-Dias, *IEEE Trans. Geosci.Remote Sens.*, vol. 43, no. 1, pp. 175–187, Jan(2005)
7. D. Lee , H. Seung, *Adv. Neural Inf. Process. Syst.*, vol. 13, pp. 556–562 (2001)
8. Y. Qian, S. Jia, J. Zhou, and A. Robles-Kelly, *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4282–4297, Nov (2011)
9. D. Heinz , C.I. Chang, *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no.3,pp.529–545,Mar. (2001)
10. N. Keshava , J. Mustard, *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan (2002)
11. E.W. Forgy, Cluster analysis of multivariate data: efficiency versus interpretability of classifications, *Biometrics* 21,768–769 (1965)
12. Z. Xu, X. Chang, F. Xu, and H. Zhang, *IEEE Trans. Neural Netw.Learn. Syst.*, vol. 23, no. 7, pp. 1013–1027, Jul. (2012)
13. Z. B. Xu, H. Zhang, Y. Wang, and X. Y. Chang, *Sci.China*, vol. 53, no. 6, pp. 1159–1169, (2010)
14. J. S. Zeng, J. Fang, and Z. B. Xu, *Sci. China Inf. Sci.*, vol. 55, no. 8, pp. 1755–1775(2012)
15. C. Xu, Z. M. Peng, and W. F. Jing, *Sci. China Inf. Sci.*, vol. 56, no. 042308,pp. 1–16(2013)
16. S. Cao, H. Qin, H.W. Deng, and Y.P. Wang, in *Proc. Int. Conf.Bioinformat. Comput. Biol. Biomed. Informat.*pp. 191–200(2013)
17. U. T. P. Paatero, *Environmetrics*, vol. 5, pp. 111–126(1994)