HIERARCHICAL ALTERNATING LEAST SQUARES ALGORITHM WITH SPARSITY CONSTRAINT FOR HYPERSPECTRAL UNMIXING

Sen Jia\(^1\), Yuntao Qian\(^2\), Jiming Li\(^2\), Yan Li\(^1\), Zhong Ming\(^1\)

\(^1\)Shenzhen City Key Laboratory of Embedded System Design, College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China
\(^2\)College of Computer Science, Zhejiang University, Hangzhou, China

ABSTRACT

In this paper, we not only extend the temporal hierarchical alternating least squares (HALS) to spatial domain, but also incorporate two necessary characteristics of material abundances, full additivity and sparsity, to unmix hyperspectral data. The new algorithm is abbreviated as HALSSC (HALS with Sparsity Constraint). Different from the other endmember extraction approaches, the proposed algorithm does not need the existence assumption of pure pixel of each endmember in the scene. Experimental results on highly mixed synthetic data and real hyperspectral data from Washington DC mall confirm the accuracy of the developed algorithm.

Index Terms— Hyperspectral unmixing, nonnegative matrix factorization (NMF), hierarchical alternating least squares (HALS), sparsity

1. INTRODUCTION

Hyperspectral sensors collect multichannel contiguous narrow spectral band imagery, spanning from the visible to the infrared portion of the electromagnetic spectrum. Due to the low spatial resolution of the sensor, disparate substances may contribute to the spectrum measured from a single pixel, leading to the wide-existence of “mixed” pixel in hyperspectral imagery. Hence, hyperspectral unmixing, which decomposes a mixed pixel into a collection of constituent spectra, or endmembers, and their corresponding fractional abundances indicating the proportion of each endmember, is often used to preprocess the hyperspectral data [1].

The linear mixing model is classically used to model the spectrum of a pixel in the observed scene. Conventional hyperspectral unmixing generally includes two steps: endmember extraction and mixed pixel decomposition. Several endmember extraction algorithms have been proposed by projecting the data into a new space, and choosing the extreme as the endmember spectra, such as pixel purity index (PPI), N-FINDR, vertex component analysis (VCA) etc. Nevertheless, the non-completeness of ground object spectral library and the existence assumption of pure pixel of each endmember lead to the unsatisfactory unmixing results. On the other hand, based on the linear spectral mixing model and constrained conditions of mixed pixel, unsupervised hyperspectral unmixing techniques obtain the endmember spectra and their corresponding fractional abundances directly from the hyperspectral imagery without knowing the knowledge of endmembers.

According to the nonnegativity of both spectra and abundances, nonnegative matrix factorization (NMF) [2], which decomposes the data into two nonnegative matrices (i.e., the elements are constrained to be nonnegative), has recently been applied to hyperspectral unmixing [3]. From signal processing point of view and data analysis, NMF are very attractive because they take into account temporal correlation between variables and usually provide a parts-based representation of the data, making the resultant matrices more intuitive and interpretable [4]. However, since the minimization of the cost function is not convex, only local minima could be achieved.

In this paper, we extend an algorithm referred as hierarchical alternating least squares (HALS) [5] to unmix hyperspectral imagery. Different from the commonly adopted multiplicative updating rules proposed by Lee and Seung [6], HALS performs sequential constrained minimization on a set of squared Euclidean distances. Except the full additivity property (given in the next section) of endmember abundances caused by the physical constraints in the data acquisition process, sparsity is another important property of hyperspectral imagery [7, 8]. Specifically, most of the pixels are mixed by only several but not all the endmembers in the scene, implying that the abundance of each endmember is localized, with a degree of sparsity. By incorporating the penalty terms into the local squared Euclidean norms, we develop a new algorithm called HALSSC (HALS with Sparsity Constraint), which is able to achieve sparse representation of the desired solution, and alleviate the problem of getting stuck in local minima. Experimentations on synthetic and real data validate the efficiency of the proposed approach.

The remainder of this paper is organized as follows.
Section 2 presents a brief resume of linear spectral mixture model. Section 3 develops the HALSSC algorithm. Experimental results on synthetic and real hyperspectral data are reported in Section 4. Section 5 concludes the paper.

2. LINEAR SPECTRAL MIXTURE MODEL

For the 3D hyperspectral imagery, \( I, J \) and \( B \) are respectively used to denote the spatial dimensions and spectral bands. Let \( X \) be the image cube with each spectrum \( X_{ij} \) being an \( B \times 1 \) pixel vector. Let \( S \) be an \( B \times K \) spectral signature matrix that each column vector \( S_k \) corresponds to an endmember spectrum and \( K \) is the number of endmembers in the image. Let \( A \) be the abundance cube (the length of each dimension is \( I, J \) and \( K \) respectively) and every column \( A_{ij} \) be a \( K \times 1 \) abundance vector associated with \( X_{ij} \). Accordingly, the simplified linear spectral mixture model for the pixel with coordinate \((i, j)\) can be written as [9]

\[
X_{ij} = SA_{ij} + N \tag{1}
\]

where \( N \) is noise that can be interpreted as receiver electronic noise. Meanwhile, endmember spectra and fractional abundances are subject to

\[
S_{bk} \geq 0 \ (1 \leq b \leq B), \quad A_{ijk} \geq 0, \quad \sum_{k=1}^{K} A_{ijk} = 1 \tag{2}
\]

which are called nonnegativity and full additivity respectively.

3. HALS ALGORITHM WITH SPARSITY CONSTRAINT FOR HYPERSPECTRAL UNMIXING

Before introducing the HALSSC algorithm, we provide a brief background on the alternating least squares (ALS) algorithm, which performs NMF by minimization of the following squared Euclidean distance objective function:

\[
E(A, S) = \frac{1}{2} \|X - AS^T\|^2 = \frac{1}{2} \sum_{i,j,b} (X_{ij} - (SA_{ij})_{b})^2 \tag{3}
\]

where the entries in \( S \) and \( A \) are nonnegative. The ALS algorithms take advantage of the fact that the problem has bilinear structure, that is for fixed \( A \) the problem is linear in \( S \), and for fixed \( S \) the problem is linear in \( A \). However, the decomposition model defined by (3) does not impose any constraints and therefore minimizing the objective function can hardly yield a factorization which reveals local features in the data. For hyperspectral imagery, through taking into account the full additivity and sparsity constraints for the abundance cube \( A \), together, the cost function can be reformulated as

\[
D(A, S) = E(A, S) + \alpha J_\alpha(A) + \beta J_\beta(A) \tag{4}
\]

\[
J_\alpha(A) = \| \sum_k A_k - 1 \|^2 = \sum_{i,j} (\sum_k A_{ijk} - 1)^2
\]

\[
J_\beta(A) = -\|A\|^2 = -\sum_{i,j,k} A_{ijk}^2
\]

where \( A_k \) is the \( I \times J \times 1 \) abundance fraction of endmember \( k \), \( 1 \) is the matrix with all element being 1, \( \alpha > 0 \) and \( \beta > 0 \) are regularization parameters controlling level of full additivity and sparsity, respectively.

The basic idea of HALS is to define residues:

\[
X^{(k)} = X - \sum_{p \neq k} A_p S_p^T = X - \alpha S_k^T + A_k S_k^T \tag{5}
\]

for \( k = 1, 2, \ldots, K \). Substituting equation (5) into equation (4) yields a set of new cost functions:

\[
D_k^{(k)}(A_k, S_k) = \frac{1}{2} \|X^{(k)} - A_k S_k^T\|^2 + \frac{\alpha}{2} \| \sum_k A_k - 1 \|^2 - \frac{\beta}{2} \|A_k\|^2 \tag{6}
\]

The gradients of the cost functions (6) with respect to the unknown \( A_k \) and \( S_k \) are expressed by

\[
\frac{\partial D^{(k)}(A_k, S_k)}{\partial A_k} = \|S_k\|^2 A_k - \sum_b X^{(k)k} S_k + \alpha (\sum_k A_k - 1) - \beta A_k \tag{7}
\]

\[
\frac{\partial D^{(k)}(A_k, S_k)}{\partial S_k} = \|A_k\|^2 S_k - \sum_{i,j} X^{(k)k} A_{ij} \tag{8}
\]

By equating equations (7) and (8) to zero and using a simple “half-rectifying” nonlinear projection to enforce the nonnegativity constraints, a new set of sequential learning rules are obtained:

\[
A_k = \frac{1}{\|S_k\|^2 + \alpha - \beta (\sum_b X^{(k)k} S_k + \alpha (1 - \sum_p A_p + A_k))} \tag{9}
\]

\[
S_k = \frac{1}{\|A_k\|^2} [\sum_{i,j} X^{(k)k} A_{ij}]_+ \tag{10}
\]

for \( k = 1, 2, \ldots, K \), where \([\delta]_+ = \max\{\epsilon, \delta\}\), and \( \epsilon \) is a small constant to enforce positive entries (typically, \( 10^{-10} \)). We refer to these update rules as the HALSSC algorithm. Obviously, as long as the initial values of \( A \) and \( S \) are all chosen strictly positive, and \( \alpha \) is greater than \( \beta \), the update rules guarantee that the elements of the two matrices stay nonnegative. In this paper, the number of endmembers \( K \) present in the scene are assumed to be known in advance, which could be estimated by HySime algorithm [10]. The detailed pseudocode of HALSSC algorithm is summarized as follows.

1. Initialize \( A \) and \( S \) using the standard multiplicative update algorithm, then rescale \( A \) to satisfy the full additivity constraint
2. \( E = X - AS^T \)
3. repeat
   for \( k = 1 \) to \( K \)
   (a) \( X^{(k)} \leftarrow E \) + \( A_k S_k^T \)
   (b) Update \( A_k \) and \( S_k \) by (9) and (10), respectively
   (c) \( E \leftarrow X^{(k)} - A_k S_k^T \)
   end for
until convergence criterion is reached

At last, it is worth to mention that \( A \) and \( S \) are initialized from multiple different starting points in order to decrease the sensitivity of the developed algorithm on initial values.

4. EXPERIMENTAL RESULTS

In this section, the proposed HALSSC algorithm are compared with SCNMF, which extends the standard multiplicative algorithm by enforcing the smoothness constraints in spectra and abundances [3]. Meanwhile, to evaluate the performance of the algorithms, the spectral angle distance (SAD) is used to compare the similarity of the \( k \)th true endmember signature \( S_k \) and its estimate \( \hat{S}_k \), which is defined as

\[
SAD_k = \arccos \left( \frac{S_k^T \hat{S}_k}{\| S_k \| \| \hat{S}_k \|} \right) 
\]

SAD is invariant to unknown multiplicative scalings of spectra. On the other hand, root mean square error (RMSE) is taken to evaluate the similarity of true versus estimated abundances,

\[
RMSE_k = \left( \frac{1}{T \times J} \sum_{i,j} (A_{ijk} - \hat{A}_{ijk})^2 \right)^{\frac{1}{2}} 
\]

4.1. Synthetic Data

The spectral signatures are chosen from the United States Geological Survey (USGS) digital spectral library [11]. Fig. 1 shows five of these endmember signatures. In order to represent the sparsity of spatial abundances, it is created through three steps: firstly, the scene, with a size of \( z^2 \times z^2 \) \((z \in \mathbb{Z}^+)\)

pixels, is divided into \( z \times z \) regions. Each region is initialized with the same type of ground cover, randomly selected as one of the endmember classes. Secondly, we use a simple \((z + 1) \times (z + 1)\) spatial low pass filter to generate mixed pixels. Thirdly, to further remove pure pixels, and represent the sparseness of abundances at the same time, we replace all the pixels whose abundance is larger than 70% with a mixture made up of only two endmembers. Hence, there is not a single pure signature for each endmember. Additionally, the added noise item for every endmember in (1) is chosen to be 5% Gaussian, that is \( \| N_k \| = 0.05 \times \| X_k \| / \| R_k \| \) where the entries in \( \| R_k \| \) are randomly chosen from a Gaussian distribution \( N(0, 1) \).

The true and estimated signatures of endmembers are illustrated in Figure 2 (the estimated abundances and the results of SCNMF are omitted due to the length of the paper). Obviously, all the five materials are well extracted. Table 1 quantifies the unmixing results using the two performance metrics (the first line is SAD, while the second is RMSE).
4.2. Real data

Figure 3 shows a subscene of size $30 \times 30$ extracted from the Washington D.C. data set. After low signal-to-noise ratio (SNR) bands are removed, only 191 bands remain (i.e., $B=191$). According to the ground truth [12], there are four materials of interest: grass, trail, tree, and water, so the number of endmembers is equal to 4.

Figure 4 and 5 display the separated abundance fractions of each endmember by SCNMF and HALSSC, respectively. The parameter values are the same as those used in the above experiment. As can be seen from the figure, except the abundance of water, the other three material abundances extracted by SCNMF are still mixtures. Contrarily, HALSSC successfully separates out all the four endmembers. Table 2 quantifies the obtained results using the SAD similarity and RMSE error scores.

5. CONCLUSION

In this paper, we have presented a HALS with sparsity constraint algorithm (named HALSSC) for hyperspectral unmixing. Firstly, the algorithm extends the HALS to spatial domain. Secondly, two characteristics of endmember abundances, full additivity and sparsity, are introduced. Its effectiveness has been tested by comparison to SCNMF with synthetic and real data. The experimental results show that HALSSC could achieve more accurate estimates of both endmember spectra and abundance maps.

6. REFERENCES