$L_{1/2}$ Sparsity Constrained Nonnegative Matrix Factorization for Hyperspectral Unmixing

Yuntao Qian[1], Sen Jia[2]*, Jun Zhou[3,4], Antonio Robles-Kelly[3,4]
[1]College of Computer Science, Zhejiang University, Hangzhou, China
[2]Shenzhen City Key Laboratory of Embedded System Design
College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China
[3]Canberra Research Laboratory, NICTA†, PO BOX 8001, Canberra, ACT 2601, Australia
[4]College of Engineering and Computer Science, The Australian National University, Canberra, ACT 0200, Australia

Abstract

Hyperspectral unmixing is a crucial preprocessing step for material classification and recognition. In the last decade, nonnegative matrix factorization (NMF) and its extensions have been intensively studied to unmix hyperspectral imagery and recover the material end-members. As an important constraint, sparsity has been modeled making use of L_1 or L_2 regularizers. However, the full additivity constraint of material abundances is often overlooked, hence, limiting the practical efficacy of these methods. In this paper, we extend the NMF algorithm by incorporating the $L_{1/2}$ sparsity constraint. The $L_{1/2}$-NMF provides more sparse and accurate results than the other regularizers by considering the end-member additivity constraint explicitly in the optimisation process. Experiments on the synthetic and real hyperspectral data validate the proposed algorithm.

1. Introduction

Hyperspectral data is acquired by high spectral-resolution imaging sensors, containing hundreds of contiguous narrow spectral band images. Due to the low spatial resolution of the sensor, disparate substances may contribute to the spectrum for a single pixel, leading to the existence of "mixed" spectra in hyperspectral imagery. Hence, hyperspectral unmixing, which decomposes a mixed pixel into a collection of constituent spectra, or *end-members*, and their corresponding fractional abundances, is often adopted to preprocess hyperspectral data [14]. Several hyperspectral unmixing methods have been proposed in recent years, which include N-FINDR [30], vertex component analysis [20], Independent component analysis [29], the minimum volume enclosing simplex algorithm [2] and flexible similarity measures [4].

Most of these methods assume a linear spectral mixture model to be applied to the unmixing problem. Two constraints need to be considered when solving such linear mixture model [3]. The first of these is nonnegativity of both spectra and their fractional abundances. This is natural as the contribution from end-members should be larger than or equal to zero. The second constraint is the additivity of the fractional abundances, which guarantees the addition of the proportional contribution from the end members matches the mixed observation.

Nonnegative matrix factorization (NMF) [24, 15], which commonly adopts a Euclidean loss function and decomposes the data into two nonnegative matrices, is a natural solution to the nonnegativity constraint [25, 26]. From the data analysis point of view, NMF is very attractive because it usually provides a part-based representation of the data, making the decomposition matrices more intuitive and interpretable [5, 12]. However, since the full additivity constraint of hyperspectral data is not considered, the solution space is too large. This, added to the fact that the minimization of the cost function is not convex can make the algorithm prone to noise corruption and computationally demanding.

To reduce the space of solutions, sparsity constraints are attractive since they allow exploiting the notion that most of the pixels are a mixture of only a few of the end-members in the scene [13, 33]. This implies that a number of entries in the abundance matrix are zeros, in some cases with a large degree of sparsity. Regularization methods are usually uti-

*Corresponding Author. Sen Jia is supported by National Natural Science Foundation of China (60902070) and Doctor Starting Project of Natural Science Foundation of Guangdong Province, China (9451806001002287).

†NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program

978-0-7695-4271-3/10 $26.00 © 2010 IEEE
DOI 10.1109/DICTA.2010.82

lized to define the sparsity constraint on the abundance of the end-members. Along these lines, the L_0 regularizer accounts for the number of zero elements in an abundance matrix so as to yield the most sparse result given a cost function. However, the application of the L_0 regularizer is an NP hard optimization problem that cannot be solved in practice [22]. The L_2 regularizer, on the other hand, generates smooth but not sparse results [1]. In general, the L_1 regularizer is the most popular one for achieving sparsity of the abundance matrix [9, 10, 23, 32, 18].

There are two limitations for the application of the L_1 regularizer. Firstly, enforcing a sufficiently sparse solution is not straightforward since the penalty imposed upon the cost function is a linear one with respect to the deviations of the abundance matrix from zero. Second, and more severely, it conflicts with the full additivity unmixing constraint. Recall that the full additivity constraint requires that the sum over the L_1 regularization function be a constant. This is often conflicts with the sparsity constraint imposed by the regularizer which favours smaller, not constant, summations over the terms of the regularization function.

Recently, Xu *et al* have proposed fractional regularizers, i.e. $L_q(0 \leq q \leq 1)$ [31]. The authors have shown that the $L_{1/2}$ regularizer is a good choice to enforce the sparsity constraint due to the fact that its an unbiased one. In this paper, we introduce the $L_{1/2}$ regularization function into NMF so as to characterize the sparsity of abundances. The $L_{1/2}$-NMF presented here is effected through the multiplicative update algorithm by Lee and Seung [16] by the applications of a rescaled gradient descent approach which, in turn, ensures convergence. In our approach, the full additivity constraint is considered in the parameter update process. The experiments on synthetic and real hyperspectral data demonstrate the effectiveness of the $L_{1/2}$-NMF approach for unmixing hyperspectral data.

2. NMF with $L_{1/2}$ Sparsity Constraint

To commence, let the linear mixing model, which is classically used to model the spectrum of a pixel in the observed scene [27], be expressed as

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{e} \tag{1}$$

where $\mathbf{x}$ denotes the $L \times 1$ vector of the observed spectral pixel, $\mathbf{s}$ is the $K \times 1$, vector of the abundance fractions of each end-member, $\mathbf{e}$ is a $L \times 1$ vector of an additive noise representing the measurement errors and $\mathbf{A}$ is the $L \times K$ nonnegative spectral signature matrix whose columns corresponding to an end-member spectrum. In the dimensionalities above, L is the number of bands and K is the number of end-members.

Using matrix notation, the mixing model for the N pixels in the image can be written as

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{E} \tag{2}$$

where the matrices $\mathbf{X} \in \mathbb{R}_+^{L\times N}$, $\mathbf{S} \in \mathbb{R}_+^{K\times N}$ and $\mathbf{E} \in \mathbb{R}^{L\times N}$ represent, respectively, the hyperspectral data, the end-member abundances and the additive noise. Note that, in general, only $\mathbf{X}$ is known in advance, while the other two matrices, $\mathbf{A}$ and $\mathbf{S}$ are our aims of computation. With these ingredients, the Euclidean distance-based loss function of NMF is as follows

$$\mathcal{C}(\mathbf{A}, \mathbf{S}) = \frac{1}{2}\|\mathbf{X} - \mathbf{A}\mathbf{S}\|_2^2 \tag{3}$$

When the regularization term is added to control the sparsity, we get

$$\mathcal{C}(\mathbf{A}, \mathbf{S}) = \frac{1}{2}\|\mathbf{X} - \mathbf{A}\mathbf{S}\|_2^2 + \lambda\|\mathbf{S}\|_q \tag{4}$$

where λ is the regularization parameter and $\|\mathbf{S}\|_q = (\sum_{k,n=1}^{K,N} \mathbf{S}_{kn}^q)^{1/q}$. In the case of $L_{1/2}$ regularization, $q = \frac{1}{2}$.

2.1. Rescaled Gradient Descent Algorithm

Based on the cost function described in Equation (4), and, making use of an auxiliary function similar to that used in the Expectation-Maximization algorithm [9], we have developed a rescaled gradient descent algorithm by iterating the following multiplicative update rules. The objective (4) is nonincreasing under the update rules:

$$\mathbf{A} \leftarrow \mathbf{A}. * \mathbf{X}\mathbf{S}^T./\mathbf{A}\mathbf{S}\mathbf{S}^T \tag{5}$$

$$\mathbf{S} \leftarrow \mathbf{S}. * \mathbf{A}^T\mathbf{X}./(\mathbf{A}^T\mathbf{A}\mathbf{S} + \frac{\lambda}{2}\mathbf{S}^{-\frac{1}{2}}) \tag{6}$$

where $(\cdot)^T$ denotes the transpose of the matrix, .* and ./ denote element-wise multiplication and division, respectively. Likewise, $\mathbf{S}^{-\frac{1}{2}}$ is the negative element-wise square root for each entry in the matrix $\mathbf{S}$.

The update rule for $\mathbf{A}$ in (5) is the same as that in [16]. For the sake of brevity, we focus our attention on the update rule for $\mathbf{S}$ in (6). To make our ellaboration more clear, we focus on each column of $\mathbf{S}$ alone. We can do this without any loss of generality since the objective function (4) is separable in the columns of $\mathbf{S}$. Let these columns be denoted $\mathbf{s}$ for convenience. Similarly, the corresponding row of $\mathbf{X}$ is denoted $\mathbf{x}$. The column-wise objective function becomes

$$\mathcal{C}(\mathbf{s}) = \frac{1}{2}\|\mathbf{x} - \mathbf{A}\mathbf{s}\|_2^2 + \lambda\|\mathbf{s}\|_{\frac{1}{2}} \tag{7}$$

We define an auxiliary function $G(\mathbf{s}, \mathbf{s}^t)$ satisfying the conditions $G(\mathbf{s}, \mathbf{s}) = \mathcal{C}(\mathbf{s})$ and $G(\mathbf{s}, \mathbf{s}^t) \geq \mathcal{C}(\mathbf{s})$ such that

$\mathcal{C}(\mathbf{s})$ is nonincreasing when updated using the following equation

$$\mathbf{s}^{(t+1)} = \arg\min_{\mathbf{s}} G(\mathbf{s}, \mathbf{s}^t) \tag{8}$$

This is guaranteed by

$$\mathcal{C}(\mathbf{s}^{(t+1)}) \leq G(\mathbf{s}^{(t+1)}, \mathbf{s}^t) \leq G(\mathbf{s}^t, \mathbf{s}^t) = \mathcal{C}(\mathbf{s}^t) \tag{9}$$

Following [9], we define the function G as

$$G(\mathbf{s}, \mathbf{s}^t) = \mathcal{C}(\mathbf{s}^t) + (\mathbf{s} - \mathbf{s}^t)(\nabla\mathcal{C}(\mathbf{s}^t))^T + \\ + \frac{1}{2}(\mathbf{s} - \mathbf{s}^t)\mathbf{K}(\mathbf{s}^t)(\mathbf{s} - \mathbf{s}^t)^T \tag{10}$$

where the diagonal matrix $\mathbf{K}(\mathbf{s}^t)$ is defined as

$$\mathbf{K}(\mathbf{s}^t) = \mathrm{diag}\left(\left(\mathbf{A}^T\mathbf{A}\mathbf{s}^t + \frac{\lambda}{2}(\mathbf{s}^t)^{-\frac{1}{2}}\right)./\mathbf{s}^t\right) \tag{11}$$

Here, $\mathrm{diag}(\mathbf{s})$ denotes the diagonalization of vector $\mathbf{s}$. Because $G(\mathbf{s}, \mathbf{s}) = \mathcal{C}(\mathbf{s})$, the Taylor expansion of $\mathcal{C}(\mathbf{s})$ is

$$\mathcal{C}(\mathbf{s}) = \mathcal{C}(\mathbf{s}^t) + (\mathbf{s} - \mathbf{s}^t)(\nabla\mathcal{C}(\mathbf{s}^t))^T + \\ + \frac{1}{2}(\mathbf{s} - \mathbf{s}^t)\left(\mathbf{A}^T\mathbf{A} - \frac{\lambda}{4}\mathrm{diag}\left((\mathbf{s}^t)^{-\frac{3}{2}}\right)\right)(\mathbf{s} - \mathbf{s}^t)^T + \\ + \mathcal{O}\left(\nabla^{(n\geq 3)}\mathcal{C}(\mathbf{s}^t)\right) \tag{12}$$

where the function $\mathcal{O}$ denotes the Lagrange remainder term, which can be omitted in the following derivation. Therefore, the constraint $G(\mathbf{s}, \mathbf{s}^t) \geq \mathcal{C}(\mathbf{s})$ is satisfied if

$$(\mathbf{s} - \mathbf{s}^t)\Bigg(\mathbf{K}(\mathbf{s}^t) - \mathbf{A}^T\mathbf{A} + \\ + \frac{\lambda}{4}\mathrm{diag}\big(\mathbf{s}^t\big)^{-\frac{3}{2}}\Bigg)(\mathbf{s} - \mathbf{s}^t)^T \geq 0 \Rightarrow$$

$$(\mathbf{s} - \mathbf{s}^t)\Bigg(\mathbf{K}'(\mathbf{s}^t) + \frac{\lambda}{2}\mathrm{diag}\big(\mathbf{s}^t\big)^{-\frac{1}{2}} + \\ + \frac{\lambda}{4}\mathrm{diag}\big(\mathbf{s}^t\big)^{-\frac{3}{2}}\Bigg)(\mathbf{s} - \mathbf{s}^t)^T \geq 0 \tag{13}$$

where $\mathbf{K}'(\mathbf{s}^t)$ is defined as

$$\mathbf{K}'(\mathbf{s}^t) = \mathrm{diag}\left(\mathbf{A}^T\mathbf{A}\mathbf{s}^t./\mathbf{s}^t\right) - \mathbf{A}^T\mathbf{A} \tag{14}$$

Lee and Seung has proved the positive semidefiniteness of $\mathbf{K}'(\mathbf{s}^t)$. Due to the nonnegativity of $\mathbf{s}$, the other two terms in (13) are obviously nonnegative. Because the sum of two positive semidefinite matrices is also positive semidefinite. Equation (13) holds by replacing $G(\mathbf{s}, \mathbf{s}^t)$ in (8) by (10), which results in the update rule

$$\begin{aligned}\mathbf{s}^{(t+1)} &= \mathbf{s}^t - \nabla\mathcal{C}(\mathbf{s}^t)\mathbf{K}^{-1}(\mathbf{s}^t) \\ &= \mathbf{s}^t - \left(\mathbf{A}^T\mathbf{A}\mathbf{s}^t - \mathbf{A}^T\mathbf{x} + \frac{\lambda}{2}\big(\mathbf{s}^t\big)^{-\frac{1}{2}}\right) \\ &\quad .*\mathbf{s}^t./\left(\mathbf{A}^T\mathbf{A}\mathbf{s}^t + \frac{\lambda}{2}\big(\mathbf{s}^t\big)^{-\frac{1}{2}}\right) \\ &= \mathbf{s}^t.*\mathbf{A}^T\mathbf{x} \\ &\quad ./\left(\mathbf{A}^T\mathbf{A}\mathbf{s}^t + \frac{\lambda}{2}\big(\mathbf{s}^t\big)^{-\frac{1}{2}}\right)\end{aligned} \tag{15}$$

Note that the update rule above is the matrix form of Equation (6).

2.2. Implementation Issues

As long as the initial values of $\mathbf{A}$ and $\mathbf{S}$ are all chosen strictly positive, the update rules guarantee that the elements of the two matrices remain nonnegative. In order to satisfy the full additivity constraint of end-member abundances we employ the method in [8]. The data matrix $\mathbf{X}$ and the signature matrix $\mathbf{A}$ are augmented by a row of constants defined by

$$\mathbf{X}_f = \begin{bmatrix} \mathbf{X} \\ \delta\mathbf{1}_N^T \end{bmatrix} \quad \mathbf{A}_f = \begin{bmatrix} \mathbf{A} \\ \delta\mathbf{1}_K^T \end{bmatrix} \tag{16}$$

where δ controls the impact of the full additivity of the abundances. The larger the δ, the closer the columns of $\mathbf{S}$ to the full additivity constraint. In each iteration, these two matrices are taken as the input of the update rule of $\mathbf{S}$ given in Equation (6) as an alternative to $\mathbf{X}$ and $\mathbf{A}$.

Note that the initialization of the signature matrix $\mathbf{A}$ can be computed by applying the end-member extraction method [21] or using directly chosen data [19]. For $L_{1/2}$-NMF, $\mathbf{A}$ and $\mathbf{S}$ are both initialized uniformly via randomly selecting a value in the interval [0, 1]. Also, note that the estimation of the number of end-members present in the scene is crucial in the unmixing process. Here we resort to the HySime algorithm [7], which is one of the most reliable estimators for signal subspace dimensionality. We have also adopted two stopping criteria for our iterative optimisation. The first of these is the maximum iteration number, which is set to be 3000 due to the slow convergence of the multiplicative update [17]. The second one is the gradient difference of the cost function $\mathcal{C}$ between the current iteration and the starting value, i.e.

$$\|\nabla\mathcal{C}(\mathbf{A}^i, \mathbf{S}^i)\|_2^2 \leq \epsilon\|\nabla\mathcal{C}(\mathbf{A}^1, \mathbf{S}^1)\|_2^2$$

where ϵ is a small value, which is set to 10^{-3}. Once either of these criteria is met, the optimisation ends. The value of parameter λ is dependent on the sparsity of the material abundances that cannot be obtained from the prior information. Following [11], we use a rough estimator for λ based

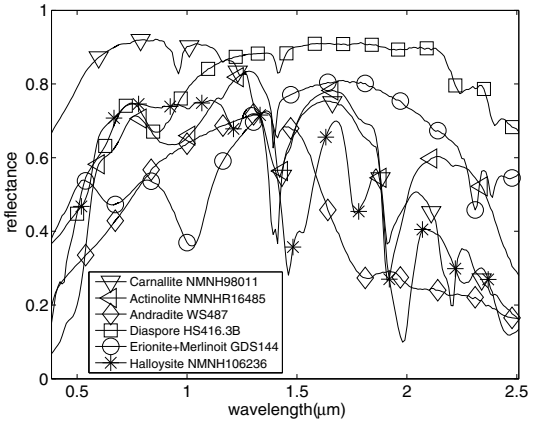


Figure 1. Selected spectral signatures from USGS.

on the sparseness criteria in [10] given by

$$\lambda = \frac{1}{\sqrt{L}} \sum_{l} \frac{\sqrt{N} - \|\mathbf{x}_l\|_1 / \|\mathbf{x}_l\|_2}{\sqrt{N} - 1} \tag{17}$$

where $\mathbf{x}_l$ denotes the lth image band in hyperspectral imagery.

3. Experiments on Synthetic Data

In this section, the proposed $L_{1/2}$-NMF algorithm is compared against three alternatives. These are the NMF algorithm with the full additivity constraint, the L_1 constraint in [9] and L_2 constraint in [25]. From now on, we denote these NMF, L_1-NMF and L_2-NMF. It is worth mentioning that the full additivity constraint is also added to the L_1-NMF and L_2-NMF. Moreover, to evaluate the performance of the algorithms, the spectral angle distance (SAD) is used to compare the similarity of the kth true end-member signature $\mathbf{A}_k$ and its estimate $\widehat{\mathbf{A}}_k$, which is defined as

$$\mathrm{SAD}_k = \arccos\left(\frac{\mathbf{A}_k^T \widehat{\mathbf{A}}_k}{\|\mathbf{A}_k\| \|\widehat{\mathbf{A}}_k\|}\right) \tag{18}$$

We have also used the root mean square error (RMSE) to evaluate the similarity of true versus estimated abundances, which is defined as

$$\mathrm{RMSE}_k = \left(\frac{1}{N} \sum_{n} (\mathbf{S}_{kn} - \widehat{\mathbf{S}}_{kn})^2\right)^{\frac{1}{2}} \tag{19}$$

The mean error values in Equations (18) and (19), denoted as $\overline{\mathrm{SAD}}$ and $\overline{\mathrm{RMSE}}$ are hence computed to compare the performance of the algorithms under consideration.

For our synthetic data experiments, we have chosen six spectral signatures from the United States Geological Survey (USGS) digital spectral library [6]. These are shown in Figure 1. The creation of abundances takes three steps: firstly, the scene, with a size of 49×49 pixels, is divided into 7×7 regions. Each region is initialized with the same type of ground cover, randomly selected as one of the end-member classes. Secondly, a simple 8×8 spatial low pass filter is used to generate mixed pixels. Thirdly, to further remove pure pixels, all the pixels whose abundance is larger than a threshold θ ($0 < \theta \leq 1$) are replaced with a mixture made up of only two end-members, where θ controls the abundance level of the end-members in the data. The added noise item in Equation (1) is zero-mean white Gaussian and the SNR is defined as

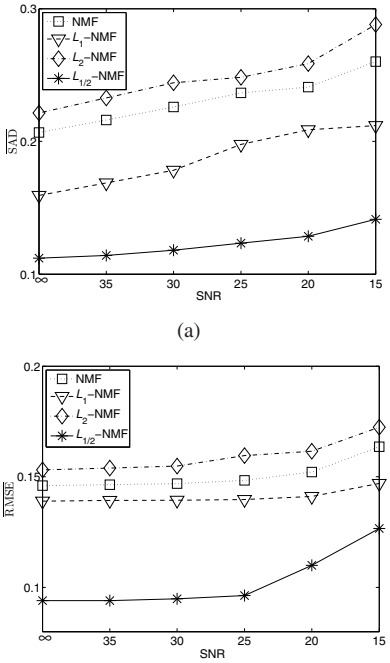
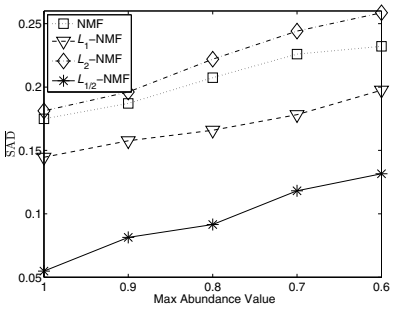


Figure 2. Performance comparison at various noise levels. (a) $\overline{\mathrm{SAD}}$ and (b) $\overline{\mathrm{RMSE}}$.

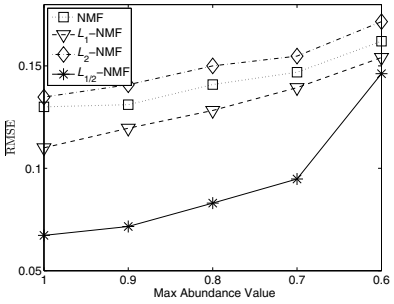
$$\mathrm{SNR} = 10 \log_{10} \frac{E[(\mathbf{As})^T (\mathbf{As})]}{E[\mathbf{e}^T \mathbf{e}]} \tag{20}$$

where $E[\cdot]$ denotes the expectation operator. Two experiments are conducted to demonstrate the efficiency of the proposed $L_{1/2}$-NMF algorithm. Firstly, the four algorithms are evaluated with respect to the SNR. Then the performance for each of the alternatives is measured as a function of the parameter θ.

Figure 2 shows the performance results as function of the SNR for the interval $(\infty, \ldots, 15)$ in steps of 5 dBs. The parameter θ is set to 0.7. As expected, the decrease of SNR incurs in a reduction of the performance for all the four algorithms. From the figure, we can see that L_2-NMF displays the worst results in the two measurement metrics. The performance of L_1-NMF is slightly better than that of NMF. Meanwhile, $L_{1/2}$-NMF not only shows the best per-



(a)



(b)

Figure 3. The performance of the algorithms as a function of signature variability index (η). (a) $\overline{\mathrm{SAD}}$ and (b) $\overline{\mathrm{RMSE}}$.

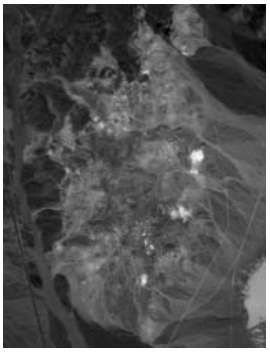


Figure 4. Scene of the AVIRIS Cuprite Nevada dataset (band 80).

formance, but also is robust to noise corruption.

Figure 3 illustrates the unmixing results with different purity levels θ $(1, 0.9, \ldots, 0.6)$. Here, if $\theta = 1$, there are "pure" pixels for each end-member. The SNR is set to 30 dB. From the figure, we can see that the performance improves as θ increases. This is expected, since the end-member abundances become sparser with the growth of θ, making the added sparseness constraint more effective. As with the first experiment, L_2-NMF shows the worst performance followed by L_1-NMF and NMF with $L_{1/2}$-NMF producing the best results.

Substance	SAD
Alunite	0.1318
Andradite	0.0554
Buddingtonite	0.1690
Dumotierite	0.0817
Kaolinite	0.1441
Montmorillonite	0.1174
Muscovite	0.1124
Nontronite	0.0827
Pyrope	0.0516
Sphene	0.0663

Table 1. SAD between extracted end-members and library spectra recovered by $L_{1/2}$-NMF.

4. Experiments on Real Data

Now we turn our attention to real-world data. The real hyperspectral data used here is acquired by the AVIRIS sensor over Cuprite, which contains abundant minerals in southern Nevada [28]. In recent years, the Cuprite dataset has been widely used for hyperspectral unmixing research [20, 19] and, hence, our results can be readily compared with other experiments reported elsewhere in the literature. Figure 4 displays the 80th band, as a subimage (250×190 pixels) of the original data. For our experiments we have removed low SNR and water-vapor absorption bands (1-2, 104-113, 148-167 and 221-224), which yields 188 bands out of the original 224.

According to [20], there are 14 types of minerals present in the scene. Its worth mentioning that variants of the same mineral with slightly different spectra are not considered as dissimilar end-members and, hence, the number of end-member is set to 10, i.e., $K = 10$. In Figure 5, we compare the estimated $L_{1/2}$-NMF end-member signatures with the USGS library spectra. Clearly, the extracted signatures are close to the USGS library spectra. Table 1 quantifies the spectral similarity using the SAD criterion.

5. Conclusion

In this paper, we have extended unmixing methods based upon nonnegative matrix factorization by incorporating the $L_{1/2}$ sparseness constraint over the end-member abundances for hyperspectral data. In contrast with previous approaches which used the L_p, $(p \geq 1)$ regularizer our method based upon the $L_{1/2}$ regularizer produces sparser unmixing results, making the extracted end-member spectra and abundance maps more accurate. We have illustrated the utility of our unmixing method on synthetic and real-world data and compared our method to a number of alternatives. In our experiments, our $L_{1/2}$-NMF algorithm exhibited better performance, specially in the presence of noise corruption and low end-member purity levels.

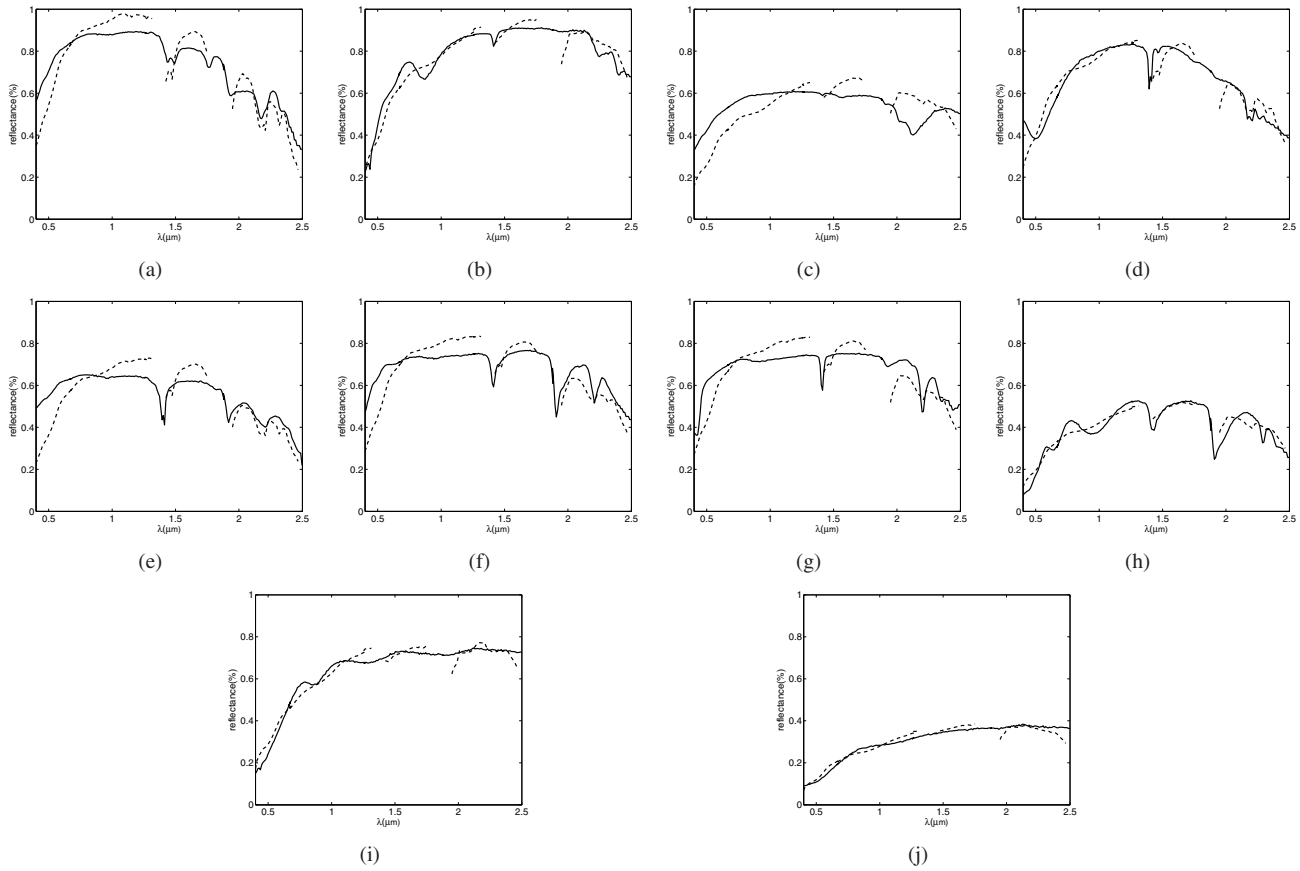


Figure 5. Comparison of the USGS library spectra (solid line) with the signatures extracted by $L_{1/2}$-NMF (dotted line). a) Alunite, b) Andradite, c) Buddingtonite, d) Dumotierite, e) Kaolinite, f) Montmorillonite, g) Muscovite, h) Nontronite, i) Pyrope, j) Sphene. Discontinuity in $L_{1/2}$-NMF results is due to the band removal.

References

[1] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons. Algorithms and applications for approximate nonnegative matrix factorization. *Computational Statistics & Data Analysis*, 52(1):155–173, 2007.

[2] T.-H. Chan, C.-Y. Chi, Y.-M. Huang, and W.-K. Ma. A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 47(11):4418–4432, 2009.

[3] C.-I. Chang. *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. Kluwer Academic/Plenum Publishers, New York, 2003.

[4] J. Chen, X. Jia, W. Yang, and B. Matsushita. Generalization of subpixel analysis for hyperspectral data with flexibility in spectral similarity measures. *IEEE Transactions on Geoscience and Remote Sensing*, 47(7):2165–2171, 2009.

[5] A. Cichocki, R. Zdunek, and S.-i. Amari. Nonnegative matrix and tensor factorization. *IEEE Signal Processing Magazine*, 25(1):142–145, 2008.

[6] 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 microns. U.S. Geol. Surv. Open File Rep. 93-592, 1993.

[7] J. M. B. Dias and J. M. P. Nascimento. Hyperspectral subspace identification. *IEEE Transactions on Geoscience and Remote Sensing*, 46(8):2435–2445, 2008.

[8] D. C. Heinz and C.-I. Chang. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 39(3):529–545, Mar. 2001.

[9] P. O. Hoyer. Non-negative sparse coding. In *Neural Networks for Signal Processing XII (Proc. IEEE Workshop on Neural Networks for Signal Processing)*, pages 557–565, Martigny, Switzerland, 2002.

[10] P. O. Hoyer. Non-negative matrix factorization with sparseness constraints. *The Journal of Machine Learning Research*, 5:1457–1469, November 2004.

[11] A. Huck, M. Guillaume, and J. Blanc-Talon. Minimum dispersion constrained nonnegative matrix factorization to unmix hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 48(6):2590–2602, 2010.

[12] S. Jia and Y. Qian. Constrained nonnegative matrix factorization for hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 47(1):161–173, 2009.

[13] J. Karvanen and A. Cichocki. Measuring sparseness of noisy signals. In *4th Int. Symp. On Independent Component Analysis And Blind Signal Separation (ica2003)*, pages 125–130, 2003.

[14] N. Keshava. A survey of spectral unmixing algorithms. *Lincoln Laboratory Journal*, 14(1):55–78, 2003.

[15] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999.

[16] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems*, pages 556–562. MIT Press, 2001.

[17] C. J. Lin. Projected gradient methods for non-negative factorization. *Neural Computation*, 19(10):2756–2779, Oct. 2007.

[18] B. Liu, S. Chen, M. Qian, and C. Zhang. Sparse norm-regularized reconstructive coefficients learning. In *Proceedings of the Ninth IEEE International Conference on Data Mining*, pages 854–859, 2009.

[19] L. D. Miao and H. R. Qi. Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization. *IEEE Transactions on Geoscience and Remote Sensing*, 45(3):765–777, Mar. 2007.

[20] J. M. P. Nascimento and J. M. B. Dias. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(4):898–910, Apr. 2005.

[21] J. M. P. Nascimento and J. M. B. Dias. Hyperspectral unmixing algorithm via dependent component analysis. In *Proc. IEEE Int. Geoscience and Remote Sensing Symp. IGARSS 2007*, pages 4033–4036, 2007.

[22] B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM Journal on Computing*, 24(2):227–234, 1995.

[23] A. P.-Montano, J. M. Carazo, K. Kochi, D. Lehmann, and R. D. P.-Marqui. Nonsmooth nonnegative matrix factorization (nsNMF). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(3):403–415, March 2006.

[24] P. Paatero and U. Tapper. Positive matrix factorization:a non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5:111–126, 1994.

[25] V. P. Pauca, J. Piper, and R. J. Plemmons. Nonnegative matrix factorization for spectral data analysis. *Linear Algebra and its Applications*, 416(1):29–47, Jul. 2006.

[26] J. Piper, V. P. Pauca, R. Plemmons, and M. Giffin. Object characterization from spectral data using nonnegative factorization and information theory. In *Proc. 2004 AMOS Technical Conference*, 2004.

[27] G. A. Shaw and H. Burke. Spectral imaging for remote sensing. *Lincoln Laboratory Journal*, 14(1):3–28, 2003.

[28] G. A. Swayze, R. L. Clark, S. Sutley, and A. J. Gallagher. Ground-truthing AVIRIS mineral mapping at cuprite, nevada. In *Summaries of the 3rd Annual JPL Airborne Geosciences Workshop*, volume 1, pages 47–49, 1992.

[29] J. Wang and C.-I. Chang. Applications of independent component analysis in endmember extraction and abundance quantification for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 44(9):2601–2616, Sep. 2006.

[30] M. E. Winter. N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data. In *Proc. SPIE Conf. Imaging Spectrometry V*, pages 266–275, 1999.

[31] Z. Xu, H. Zhang, Y. Wang, and Y. L. X.Y. Chang. $L_{1/2}$ regularizer. *Science in China, series F*, 53:1159–1169, 2010.

[32] A. Zare and P. Gader. Hyperspectral band selection and endmember detection using sparsity promoting priors. *IEEE Geoscience and Remote Sensing Letters*, 5(2):256–260, 2008.

[33] A. Zare and P. Gader. Pce: Piecewise convex endmember detection. *IEEE Transactions on Geoscience and Remote Sensing*, 48(6):2620–2632, 2010.