A projected gradient-based algorithm to unmixing hyperspectral data

Azar Zandifar, Massoud Babaie-Zadeh, Christian Jutten

A PROJECTED GRADIENT-BASED ALGORITHM TO UNMIX HYPERSPECTRAL DATA

*Azar Zandifar, Massoud Babaie-Zadeh**

Electrical Engineering Department
Sharif University of Technology
Tehran, Iran

Christian Jutten

GIPSA-Lab, Department of Images and Signals
University of Grenoble
38031 Grenoble Cedex, France

ABSTRACT

This paper presents a method to solve hyperspectral unmixing problem based on the well-known linear mixing model. Hyperspectral unmixing is to decompose observed spectrum of a mixed pixel into its constituent spectra and a set of corresponding abundances. We use Nonnegative Matrix Factorization (NMF) to solve the problem in a single step. The proposed method is based on a projected gradient NMF algorithm. Moreover, we modify the NMF algorithm by adding a penalty term to include also the statistical independence of abundances. At the end, the performance of the method is compared to two other algorithms using both real and synthetic data. In these experiments, the algorithm shows interesting performance in spectral unmixing and surpasses the other methods.

Index Terms— Spectral unmixing, linear mixture model (LMM), non-negative matrix factorization (NMF), hyperspectral imagery.

1. INTRODUCTION

Hyperspectral imagery is an imaging technique which acquires information across the electromagnetic spectrum [1]. Due to low spatial resolution, which is predominant in hyperspectral images, each pixel represents a combination of different materials. A simple but sufficiently informative model, known as linear mixture model (LMM), assumes that each pixel is a linear combination of distinct materials, namely endmembers [2].

Different approaches based on LMM have been introduced in the literature to unmix every "mixed pixel" into its constituent endmembers and their corresponding fractions, called abundances [3]. Some of them are a two-step approach for solving the unmixing problem: at the first step, endmembers are extracted using endmember extraction algorithms (EEA), e.g., pixel purity index (PPI) [4], N-Finder [5], and vertex component analysis (VCA) [6]. Then given the extracted endmembers, abundances are determined generally using least square approaches, e.g., full constrained least square (FCLS) [3].

However considering different additional statistical hypothesis for modeling the unmixing problem, some approaches such as independent component analysis (ICA) and nonnegative matrix factorization (NMF) have been proposed to extract endmembers and determine abundances in a single step procedure. The abundances' independence assumption in each pixel has led to usage of ICA as a tool to decompose a mixture into its independent component [7, 8], but this tool didn't show acceptable results in this application [9]. On the other hand, nonnegative nature of abundances and endmembers spectral signatures motivates researchers to use NMF as a tool for spectral unmixing [10, 11, 12].

NMF decomposes a mixed nonnegative matrix into a product of two nonnegative matrices. Unfortunately, NMF decomposition is not unique [13] and also due to nonconvexity of NMF cost function, it is likely that the algorithm gets trapped into local minima. Therefore, in order to escape from these spurious solutions, prior information should be used [13]. Different methods based on constrained nonnegative matrix factorization have been recently proposed to unmix hyperspectral data [10, 11, 12]. In this paper, we are going to constrain nonnegative factorization using abundance independence assumption. More precisely, we are going to propose a projected gradient based NMF algorithm which is enhanced by abundance independence constraint to unmix hyperspectral mixtures. This method somehow covers advantages of both ICA and NMF based approaches. In [14], a related method has been proposed for enforcing NMF to produce independent components for spectrogram dimentionality reduction, but not as constraint for hyperspectral unmixing problem. This paper is organized as follows. The proposed method is presented in Section 2. Then in Section 3 experimental results in both synthetic and real data are stated, and the algorithm is compared to two reference methods of hyperspectral unmixing.

*This work has been partially funded by Iran Telecom Research Center (ITRC). Christian Jutten is with Institut Universitaire de France, too

2. THE PROPOSED APPROACH

To present our proposed algorithm, we first overview the problem statement of hyperspectral unmixing using NMF.

2.1. Hyperspectral unmixing using NMF

Let $\mathbf{X}$ be the $L \times M$ matrix of observed spectra of a hyperspectral image in which each column shows observed spectrum of the corresponding pixel in L bands of imaging. Based on the LMM the hyperspectral unmixing problem can be stated as [3]:

$$\mathbf{X} = \mathbf{AS} + \mathbf{N}, \tag{1}$$

where $\mathbf{A}$ is an $L \times P$ matrix in which each column is a spectral signature of an endmember, and each column of the $P \times M$ matrix $\mathbf{S}$ shows the corresponding abundances of each pixel, and the matrix $\mathbf{N}$ stands for additive observation noise. So, P shows the total number of endmembers while L and M show the number of spectral bands and the number of pixels respectively. The goal of spectral unmixing is to estimate the matrices $\mathbf{A}$ and $\mathbf{S}$, just by observing $\mathbf{X}$, under the following conditions:

- Elements of $\mathbf{A}$ and $\mathbf{S}$ are nonnegative,
- Abundance coefficients of a pixel or each column of matrix $\mathbf{S}$ should sum to one (sum-to-one abundance constraint).

Under the constraints of non-negativity of $\mathbf{A}$ and $\mathbf{S}$, (1) may be regarded as a general NMF problem which tries to factorize the non-negative matrix $\mathbf{X}$ into two nonnegative matrices $\mathbf{A}$ and $\mathbf{S}$, while $\mathbf{N}$ represents the estimation error. The simplest and frequently used similarity measure for NMF problem is [13]:

$$D_F(\mathbf{X}||\mathbf{AS}) \triangleq \|\mathbf{X} - \mathbf{AS}\|_F^2, \tag{2}$$

in which the $\|.\|_F$ represents the Frobenius norm. Then, an alternating projected gradient algorithm updates matrices $\mathbf{A}$ and $\mathbf{S}$ as follows:

$$\mathbf{S} \leftarrow [\mathbf{S} - \mu_{\mathbf{S}} \frac{\partial D_F(\mathbf{A},\mathbf{S})}{\partial \mathbf{S}}]_+ \tag{3}$$

$$\mathbf{A} \leftarrow [\mathbf{A} - \mu_{\mathbf{A}} \frac{\partial D_F(\mathbf{A},\mathbf{S})}{\partial \mathbf{A}}]_+ \tag{4}$$

where $[.]_+$ denotes component wise projection onto the feasible nonnegative subset of real numbers.

The proposed algorithm As mentioned in section 1, some prior information should be considered as additional constraints to prevent the algorithm from getting trapped into local minima.

The assumption of the independence of the abundances of different endmembers in a pixel is the main assumption which supports ICA for spectral unmixing [7, 8]. We here consider this assumption as a prior information to enhance NMF. This assumption can be regarded as independence of the rows of $\mathbf{S}$. It has been shown in [14] that canceling the second order statistics in NMF is sufficient to make the rows of $\mathbf{S}$ independent (see Lemma 1 and 2 of [14]). So, in [14], authors have proposed using Frobenius norm of energy-normalized correlation matrix of $\mathbf{S}$, as a penalty term in NMF problem statement to achieve the independence of the rows of $\mathbf{S}$. This penalty term is defined as follows:

$$J(\mathbf{S}) = \|\mathbf{C_S}\|_F^2, \text{where } \mathbf{C_S} = \mathbf{P_S}^{(-\frac{1}{2})}\mathbf{S}\mathbf{S}^T\mathbf{P_S}^{(-\frac{1}{2})} \tag{5}$$

in which $\mathbf{P_S}$ is the diagonal matrix of sums of squares of the rows of $\mathbf{S}$.

In contrast to the multiplicative method of [14], here we use $J(\mathbf{S})$ in (5) as a penalty term to enhance a gradient projection based NMF algorithm. The cost function (2) and the update rules (3) and (4) are changed as follows:

$$D(\mathbf{X}||\mathbf{AS}) = \|\mathbf{X} - \mathbf{AS}\|_F^2 + \lambda_{\mathbf{S}} J(\mathbf{S}) \tag{6}$$

$$\mathbf{S} \leftarrow [\mathbf{S} - \mu_{\mathbf{S}}(2\mathbf{A}^T(\mathbf{AS} - \mathbf{X}) + \lambda_{\mathbf{S}} \frac{\partial J(\mathbf{S})}{\partial \mathbf{S}})]_+ \tag{7}$$

$$\mathbf{A} \leftarrow [\mathbf{A} - \mu_{\mathbf{A}}(2(\mathbf{AS} - \mathbf{X})\mathbf{S}^T)]_+ \tag{8}$$

As it has been shown in [14], $\frac{\partial J(\mathbf{S})}{\partial \mathbf{S}}$ can be computed as follows:

$$\frac{\partial J(\mathbf{S})}{\partial s_{bc}} = \sum_i \sum_j c_{ij} \frac{\partial c_{ij}}{\partial s_{bc}} \tag{9}$$

$$\frac{\partial c_{ij}}{\partial s_{bc}} = \frac{b_{ij}(\frac{\partial m_{ij}}{\partial s_{bc}}) - m_{ij}(\frac{\partial b_{ij}}{\partial s_{bc}})}{b_{ij}^2} \tag{10}$$

$$\mathbf{M} \triangleq \mathbf{S}\mathbf{S}^T \tag{11}$$

$$g_b \triangleq \|\mathbf{s}_b\| \tag{12}$$

$$\mathbf{B} \triangleq \mathbf{g}\mathbf{g}^T \tag{13}$$

$$\mathbf{U} \triangleq \mathbf{g}(\mathbf{g}_{\text{inv}})^T \tag{14}$$

$$\frac{\partial \mathbf{M}}{\partial s_{bc}} = \mathbf{1}_b \mathbf{s}_c^{*T} + \mathbf{s}_c^* \mathbf{1}_b^T \tag{15}$$

$$\frac{\partial \mathbf{B}}{\partial s_{bc}} = s_{bc}(\mathbf{U}\mathbf{1}_b\mathbf{1}_b^T + \mathbf{1}_b\mathbf{1}_b^T\mathbf{U}) \tag{16}$$

where c_{ij}, s_{ij}, m_{ij} and b_{ij} represent the elements of matrices $\mathbf{C}$, $\mathbf{S}$, $\mathbf{M}$, $\mathbf{B}$ respectively; $\mathbf{1}_b$ indicates a column vector in which all the elements are equal to zero, except the b^{th} element which is equal to one; and $\mathbf{s}_b^T$ represents the b^{th} row of matrix $\mathbf{S}$, while $\mathbf{s}_c^*$ represents the c^{th} column of $\mathbf{S}$, g_b shows the b^{th} element of column vector $\mathbf{g}$, vector $\mathbf{g}_{\text{inv}}$ represents the element-wise inverse of vector $\mathbf{g}$.

Finally, in order to satisfy sum-to-one abundance constraint we modify the update rule (7) as:

$$\mathbf{S} \leftarrow [\mathbf{S} - \mu_{\mathbf{S}}(2\overline{\mathbf{A}}^T(\overline{\mathbf{A}}\mathbf{S} - \overline{\mathbf{X}}) + \lambda_{\mathbf{S}} \frac{\partial J(\mathbf{S})}{\partial \mathbf{S}})]_+ \tag{17}$$

in which $\mathbf{A}$ and $\mathbf{X}$ are defined as follows:

$$\overline{\mathbf{A}} \triangleq \begin{bmatrix} \mathbf{A} \\ \mathbf{1}^T \end{bmatrix}, \overline{\mathbf{X}} \triangleq \begin{bmatrix} \mathbf{X} \\ \mathbf{1}^T \end{bmatrix} \tag{18}$$

where $\mathbf{1}$ represent a vector with all elements equal to one.

The step-sizes $\mu_{\mathbf{A}}$ and $\mu_{\mathbf{S}}$ are adaptively updated using the approach proposed in [11], and $\lambda_{\mathbf{S}}$ is fixed to a value chosen by trial and error. The algorithm uses VCA extracted endmembers as an initial point (this initiation setup is prevalent in similar approaches [11, 10, 15]). The algorithm stops when the maximum change in estimated matrices elements is less than a pre-assumed threshold, and/or algorithm reaches a maximum number of iterations.

In the following, we refer to our algorithm as Projected Gradient NMFICA (PG-NMFICA).

3. EXPERIMENTAL RESULTS

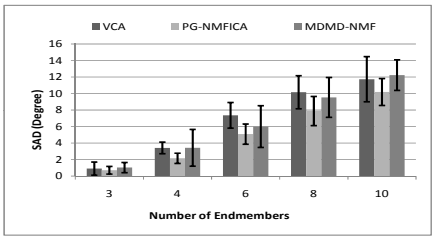
In this section, the performance of the proposed algorithm is studied experimentally using both synthetic and real data. Based on simulation results, the regularization parameter $\lambda_{\mathbf{S}}$ has been empirically set to 5.

3.1. Synthetic data

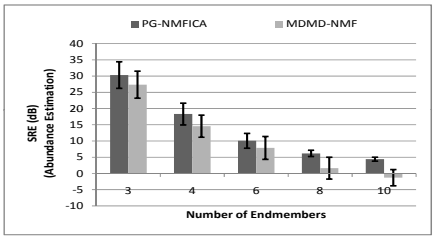
For constructing synthetic data, endmembers have been extracted from United States geological survey (USGS) [16] library. The abundances have been randomly chosen for each pixel to be between 0 and 1, also a normalization have been done to guarantee the sum-to-one property. Then a white, zero centered, Gaussian noise with variance σ^2 has been added to the data to model the observation noise[1]. The additive noise variance is defined according to input data signal to noise ratio in dB (SNR) as follows: $\mathrm{SNR} = 10\log(\frac{\|\mathbf{X}\|_F^2}{LN\sigma^2})$.

We have assumed that the true number of endmembers (P) is known a priori. We have compared the accuracy of our algorithm in estimating $\mathbf{A}$ and $\mathbf{S}$ with minimum spectral dispersion and minimum spatial dispersion NMF (MDMD-NMF) [10] and VCA [6] . The comparison measure for estimating $\mathbf{A}$ was spectral angle distance (SAD) [11], and the comparison measure for estimating $\mathbf{S}$ was signal to reconstruction error ratio (SRE) [17][2].

Figure 1 shows SAD for estimation of $\mathbf{A}$, as well as SRE for estimation of $\mathbf{S}$, averaged across 20 trials, as a function of the number of endmembers (in all the above-mentioned experiments, the number of observed pixels have been fixed to 10000). The results show that the proposed algorithm surpasses both VCA and MDMD-NMF, and has improved the abundance matrix ($\mathbf{S}$) estimation especially when the number of endmembers increases. Figures 2 and 3 shows the above-mentioned values as a function of input SNR and number of observed pixels respectively. The PG-NMFICA algorithm shows better performance in this case, too. As it was expectable, when SNR increases, all methods show better performance (see Fig. 2). Increasing the number of observed pixels will also increase the performance of the methods (see Fig. 3). Furthermore, in our simulations, the run times of both PG-NMFICA and MDMD-NMF were more or less the same (around 200 seconds on our machine with a 2.3GHz Intel quad core CPU and 2GB RAM, by using MATLAB9.2, with 5 Endmembers and 10000 observed pixels).



(a) SAD for spectra estimation.



(b) SRE for $\mathbf{S}$ estimation.

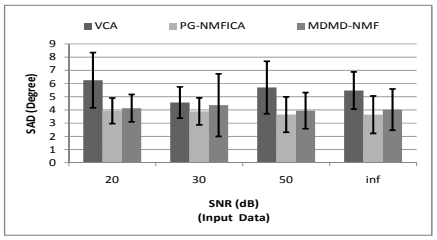
Fig. 1. Comparison between PG-NMFICA, MDMD-NMF and VCA. Spectral angle distance for spectra estimation and signal to reconstruction error for matrix $\mathbf{S}$ estimation as a function of number of endmembers.

3.2. Real data

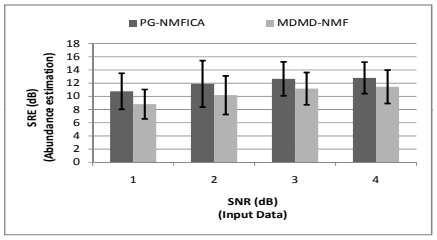
The experiment has been done on "Cuprit" data, which acquired by airborne visible/infrared imaging spectrometer (AVIRIS), and is freely available from [18]. The data has 224 spectral bands with nominal distance of $10\mathrm{nm}$ from $0.4\mu\mathrm{m}$ to $2.5\mu\mathrm{m}$, and spatial resolution is $10\mathrm{m}$. After eliminating the corrupted bands, 188 bands remain. A sub-image of 100×100 pixel was chosen, and results of the proposed algorithm are compared to MDMD-NMF. A list of ground-truth minerals of the cuprit scene is available at [19]. Since the MDMD-NMF algorithm performance decreases when the number of endmembers is set more than 13 [10], we set the number of endmembers equal to 10. The performance detailed results

[1] Although assumption of white zero mean Gaussian noise is naive for positive data, because of simplicity, it has been widely used in the field [6, 3].

[2] The VCA algorithm merely gives an estimation of endmembers [6].

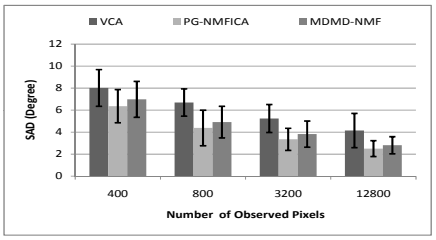


(a) SAD for spectra estimation.

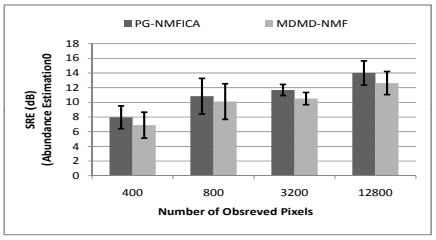


(b) SRE for $\mathbf{S}$ estimation.

Fig. 2. Comparison between PG-NMFICA, MDMD-NMF and VCA. Spectral angle distance for spectra estimation and signal to reconstruction error for matrix $\mathbf{S}$ estimation as a function of number of input data SNR.



(a) SAD for spectra estimation.



(b) SRE for $\mathbf{S}$ estimation.

Fig. 3. Comparison between PG-NMFICA, MDMD-NMF and VCA. Spectral angle distance for spectra estimation and signal to reconstruction error for both matrix $\mathbf{S}$ estimation as a function of number of observed pixels.

of our PG-NMFICA algorithm and the MDMD-NMF has been shown in Table 1. Each extracted endmember is defined by its closest neighbour in SAD sense in USGS spectral library. Accuracy of different methods are compared by the number of those extracted endmembers which correspond to the ground truth minerals [19].

As it has been shown in Table 1, PG-NMFICA algorithm extracts 6 ground truth endmembers, while MDMD-NMF algorithm is able to extract only 3 endmembers. Although the mean SAD for Cuprit endmembers are approximately the same for both methods, the total mean SAD of the proposed algorithm for all endmembers is less than that of MDMD-NMF. These results show that the proposed algorithm gives better results than MDMD-NMF algorithm in analyzing the real data.

4. CONCLUSION

In this paper, we proposed a new algorithm to unmix hyperspectral data to its constituent endmembers and their corresponding abundances. The algorithm was based on a projected gradient NMF method enhanced by the independence assumption of abundances in each pixel of the data.

The performance of the proposed algorithm was compared to similar methods versus the number of endmembers, input data SNR and the number of observed pixels. Our experiments indicate that PG-NMFICA shows better performance than other algorithms. An experiment has also been done on real data and the results of the proposed algorithm has been compared to MDMD-NMF algorithm. The proposed method shows acceptable performance in processing real data too, and overcome the other methods.

Although the proposed method shows better performance than investigated similar methods, unmixing of real data does not seem very successful. This can be regarded as the disability of the LMM to interpret the unmixing problem.

5. REFERENCES

[1] Chein-I Chang, Ed., *Hyperspectral Data Exploitation: Theory and Applications*, John Wiley & Sons, 2007.

[2] D. Manolakis, C. Siracusa, and G. Shaw, "Hyperspectral subpixel target detection using linear mixing model," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 39, no. 7, pp. 1392–1409, June 2001.

[3] N. Keshava and J.F. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 44–57, January 2002.

Table 1. Comparison between PG-NMFICA, MDMD-NMF for real data. Bold face endmembers represent the ground truth endmembers which are extracted using the corresponding algorithm, while the rest of extracted endmembers have been detected wrongly.

(a) PG-NMFICA

Number	USGS Reference	SAD(Degree)
1	**Alunite HS 295.3B**	8.1290
2	Annite WS660	7.8849
3	Hectorite SHCa-1	14.5105
4	Grossular NMNH155371	15.4339
5	**Nontronite NG-1.a**	5.2109
6	**Nontronite SWa-1.b** $< 2\mu$m	3.2245
7	**Kaolin/Smect H89-FR-5 30K**	3.7883
8	Pectolite NMNH94865.b	13.1028
9	**Kaolin-Smect KLF508**	3.0781
10	**Goethite WS219 (limonite)**	11.316
$*$	Mean-SAD(All Endmembers)	8.7593
$*$	**Mean-SAD(Cuprit Endmembers)**	5.7929

(b) MDMD-NMF

Number	USGS Reference	SAD(Degree)
1	**Alunite HS 295.3B**	6.0431
2	Andradite WS487	4.0048
3	Labradorite HS17.3B	4.7525
4	Hydroxyl-Apatite WS425	6.0310
5	Annite WS661	15.8111
6	Sphalerite S26-34	12.0207
7	**Kaolin/Smect H89-FR-5 30K**	2.4818
8	Sodium-Bicarbonate GDS55	29.1539
9	**Montmorillonite+Illi CM37**	3.0781
10	Lepidolite NMNH105538	23.8333
$*$	Mean-SAD(All Endmembers)	10.7210
$*$	**Mean-SAD(Cuprit Endmembers)**	5.6362

[4] J.W. Boardman, F.A. Kruse, and R.O. Green, "Mapping target signature via partial unmixing of aviris data," in *in Summaries of the VI JPL Airborne Earth Science Workshop*, Ottawa (ON, Canada), 1995.

[5] M.E. Winter, "N-finder: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," in *proceedings of SPIE*, Denver (CO, U.S.A), 1999, vol. 3753, pp. 266–275.

[6] J.M.P. Nascimento and J.M. Bioucas-Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, April 2005.

[7] J. Bayliss, J.A. Gualtieri, and R. Cromp, "Analysing hyperspectral data with independent component analysis," in *Proceedings of the SPIE*, Washington (DC, U.S.A), 1997, vol. 3240, pp. 133–143.

[8] J. Wang and Ch. Chang, "Applications of independent component analysis in endmember extraction and abundance quantification for hyperspectral imagery," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 44, no. 9, pp. 2601–2616, Septembre 2006.

[9] J.M.P. Nascimento and J.M.B. Dias, "Does independent component analysis play a role in unmixing hyperspectral data?," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 43, no. 1, pp. 175–187, January 2005.

[10] A. Huck and M. Guillaume, "Robust hyperspectral data unmixing with spatial and spectral regularized NMF," in *WHISPERS*, Reykjavik (Iceland), 2010, pp. 1–4.

[11] A. Huck, M. Guillaume, and J. Blanc-Talon, "Minimum dispersion constrained nonnegative matrix factorization to unmix hyperspectral data," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 48, no. 6, pp. 2590–2602, June 2010.

[12] S. Jia and Y. Qian, "Constrained nonnegative matrix factorization for hyperspectral unmixing," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 47, no. 1, pp. 161–173, January 2009.

[13] A. Cichocki, R. Zdunek, A.H. Phan, and Sh. Amari, Eds., *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*, John Wiley & Sons, 2009.

[14] K.W. Wilson and B. Raj, "Spectrogram dimensionality reduction with independence constraints," in *ICASSP*, Dallas (TX, U.S.A), 2010, pp. 1938–1941.

[15] J. Li and J.M. Bioucas-Dias, "Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data," in *IGARSS*, Boston (MA, U.S.A), 2008, vol. III.

[16] U.S. geological survey, "USGS Digital Spectral Library," `http://speclab.cr.usgs.gov/spectral-lib.html`, 2007.

[17] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 49, no. 6, pp. 2014–1039, June 2011.

[18] Jet propulsion laboratory, "AVIRIS Data," `http://aviris.jpl.nasa.gov/data/free_data.html`, 2007.

[19] U.S. geological survey, "GROUND-TRUTHING AVIRIS MINERAL MAPPING AT CUPRITE, NEVADA," `http://speclab.cr.usgs.gov/PAPERS/cuprite.gr.truth.1992/swayze.1992.html`, 1998.