An NMF-Based Approach for Hyperspectral Unmixing Using a New Multiplicative-tuning Linear Mixing Model to Address Spectral Variability

Fatima Zohra Benhalouche [1, 2, 3], *Moussa Sofiane Karoui* [1, 2, 3] and *Yannick Deville* [2]

[1] Centre des Techniques Spatiales, Arzew, Algeria
[2] IRAP, Université de Toulouse, UPS, CNRS, OMP, CNES, Toulouse, France
[3] LSI, Université des Sciences et de la Technologie d'Oran Mohamed Boudiaf, Oran, Algeria

{Fatima.Benhalouche, Sofiane.Karoui, Yannick.Deville}@irap.omp.eu

ABSTRACT

In this work, a new approach is presented for unmixing remote sensing hyperspectral data. This approach considers a linear mixing model that is introduced in these investigations to handle the spectral variability phenomenon, which is usually observed in the considered data and which is here modeled in a multiplicative form. The proposed algorithm, which is based on a pixel-by-pixel nonnegative matrix factorization method, uses multiplicative update rules for minimizing a cost function that takes into account the introduced linear mixing model. Tests, by means of realistic synthetic data, are conducted to evaluate the performance of the proposed approach, and the obtained results are compared to those of methods from the literature. These test results show that the proposed approach outperforms all other tested methods.

Index Terms— Hyperspectral imaging, linear spectral unmixing, spectral variability, nonnegative matrix factorization, multiplicative update rules

1. INTRODUCTION

Remote sensing hyperspectral sensors, with high spectral resolution, collect information in the visible and infrared wavelength domains, therefore allowing accurate classification of materials present in imaged scenes. But these sensors are characterized by a low spatial resolution, which results in mixed pixels. Consequently, each observed pixel-spectrum of such data is usually a mixture of contributions from a number of pure-material-spectra (also called endmember-spectra), which are present in the considered data. This mixture is commonly assumed to be linear [1] and unsupervised linear spectral unmixing (LSU) approaches are considered to linearly and blindly unmix all observed pixel-spectra in order to retrieve a collection of unknown endmember-spectra, and a set of associated unknown abundance coefficients with the usual sum-to-one constraint.

Classical LSU methods, which can be viewed as typical linear blind source separation (BSS) approaches [2], [3] consider that each endmember is represented by the same spectrum in all image pixels. However, in order to describe other phenomena that occur in some practical applications [4], [5], more complex models are required, wherein each endmember is often represented by different, but close, spectra in all pixels of the considered image. In that case, the concept of endmembers is replaced by classes of endmembers. This phenomenon, which principally occurs due to varying illumination and atmospheric conditions or material composition [6], is known as spectral/intra-class variability. Not addressing this issue might bring in errors that spread throughout the analysis process of the considered remote sensing data. Therefore, very recently, in [4], [5], authors developed blind unmixing methods, based on nonnegative matrix factorization (NMF) [7], which address this spectral variability issue. These methods, called Unconstrained Pixel-by-pixel NMF (UP-NMF) and Inertia-constrained Pixel-by-pixel NMF (IP-NMF), derive a different estimated spectrum in each pixel for each class of endmembers.

In this article, a new linear mixing model, involving spectral variability modeled in a multiplicative form, is firstly introduced. Then, an algorithm, based on pixel-by-pixel NMF that consists in factorizing a nonnegative matrix into a product of other nonnegative matrices [7], is proposed for unmixing the considered data. This proposed algorithm minimizes a cost function, based on the introduced mixing model, by means of multiplicative update rules.

The remainder of this paper is organized as follows. The introduced data mixing model is described in Section 2. The proposed algorithm is given in Section 3. In Section 4, the experiments, by means of synthetic data, are described. In that section, the results obtained with the proposed approach are compared with those provided by methods from the literature. Finally, a conclusion is given in Section 5.

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

2. PROPOSED DATA MIXING MODEL

In this section, the proposed mathematical data mixing model is described. As mentioned above, most LSU approaches consider that each endmember is represented by the same spectrum in all image pixels. In that case, each observed nonnegative reflectance spectrum $x_i \in \mathbb{R}_+^{L\times 1}$, associated with pixel i, is considered as

$$x_i = \sum_{j=1}^{M} r_j\, c_{ij}\ ,\ i = 1\ldots P, \tag{1}$$

where $r_j \in \mathbb{R}_+^{L\times 1}$ corresponds to the nonnegative reflectance spectrum of the j^{th} endmember. Each sample of this spectrum is lower than or equal to one. The coefficient c_{ij} is the nonnegative abundance fraction of the j^{th} endmember in the considered pixel i. M and P respectively correspond to the number of endmembers and the number of image pixels. L is the number of spectral bands. The nonnegative abundance fractions c_{ij} also obey the well-known abundance sum-to-one constraint [1].

In the present investigation, the spectral variability phenomenon is considered by introducing a new data mixing model. This one considers a different spectrum in each pixel for each endmember. Therefore, the model (1) becomes [5]

$$x_i = \sum_{j=1}^{M} r_{ij}\, c_{ij}\ ,\ i = 1\ldots P, \tag{2}$$

where $r_{ij} \in \mathbb{R}_+^{L\times 1}$ corresponds to the version associated with pixel i for the j^{th} class of endmember nonnegative reflectance spectra. Each sample of this vector is also lower than or equal to one. To address the spectral variability phenomenon, the model of r_{ij} introduced in this paper is defined by

$$r_{ij} = a_{ij} \odot e_{1j}, \tag{3}$$

where the operator $\odot$ denotes the element-wise multiplication, $a_{ij} \in \mathbb{R}_+^{L\times 1}$ corresponds to the vector of nonnegative variability coefficients for the version of the j^{th} class of endmembers associated with pixel i, and $e_{1j} \in \mathbb{R}_+^{L\times 1}$ is the nonnegative reflectance spectrum of this class of endmembers associated with the first pixel. In the proposed model, endmember spectra contained in the first pixel are considered as a reference for the considered spectral variability phenomenon. The scale coefficients in the vector a_{ij} allow tuning the spectra r_{ij} in all pixels with respect to the spectra e_{1j} in the first pixel. These vectors a_{ij} are here constrained as

$$0^{L\times 1} < \alpha^{L\times 1} \leq a_{ij} \leq \beta^{L\times 1}, \tag{4}$$

where $\alpha^{L\times 1}$ and $\beta^{L\times 1}$ are fixed coefficient vectors controlling the minimum and the maximum variability degrees. Eq. (3) clearly shows that the spectral variability is modeled in a multiplicative form. The model (2) then becomes

$$x_i = \sum_{j=1}^{M} (a_{ij} \odot e_{1j})\, c_{ij}\ ,\ i = 1\ldots P. \tag{5}$$

This last data mixing model can be written more compactly, for all pixels, in matrix form

$$X = (A \odot E)\, C, \tag{6}$$

where $X \in \mathbb{R}_+^{L\times P}$ is the observed data matrix in which each row corresponds to one spectral band of the considered image, and each column corresponds to one observed pixel spectrum. The matrix $A \in \mathbb{R}_+^{L\times (MP)}$ contains all nonnegative variability coefficients. This matrix can be written as

$$\begin{aligned} A &= [A_1|A_2| \quad \cdots \quad |A_P], \\ &\text{with } A_t \in \mathbb{R}_+^{L\times M}\ (t = 1\ldots P), \end{aligned} \tag{7}$$

$$\text{and } A_1 = 1^{L\times M}. \tag{8}$$

The matrix $E \in \mathbb{R}_+^{L\times (MP)}$ contains the M-endmember-relative-spectra submatrix $E_1 \in \mathbb{R}_+^{L\times M}$, associated with the first pixel, and replicated P times. Therefore, this matrix can be written as

$$E = \underbrace{[E_1|\, E_1| \quad \cdots \quad |E_1]}_{P}. \tag{9}$$

The matrix $C \in \mathbb{R}_+^{(MP)\times P}$ is block-diagonal and contains abundance fractions. This matrix corresponds to the transpose of that described in [5].

3. PROPOSED UNMIXING ALGORITHM

The proposed unmixing algorithm is a pixel-by-pixel NMF one. It aims at modeling the proposed data mixing function defined by (6). The involved variables are three matrices: $\tilde{A}$, $\tilde{E}$ and $\tilde{C}$ that respectively aim at estimating A, E and C. The proposed method is inspired from the standard NMF methods, which use multiplicative update rules to achieve the unmixing process. The proposed algorithm consists in minimizing the following cost function:

$$J = \frac{1}{2}\left\| X - \left(\tilde{A} \odot \tilde{E}\right) \tilde{C} \right\|_F^2, \tag{10}$$

where $\|.\|_F$ represents the Frobenius norm.

This cost function is optimized with iterative gradient-based update rules. Therefore, and in order to simply allow obtaining the gradient expressions, the cost function J is expressed as

$$J = \frac{1}{2}\,\mathrm{Tr}\,\Big(XX^{\mathrm{T}} - X\tilde{C}^{\mathrm{T}}\big(\tilde{A}\odot\tilde{E}\big)^{\mathrm{T}} - \big(\tilde{A}\odot\tilde{E}\big)\,\tilde{C}X^{\mathrm{T}} + \\ + \big(\tilde{A}\odot\tilde{E}\big)\,\tilde{C}\tilde{C}^{\mathrm{T}}\big(\tilde{A}\odot\tilde{E}\big)^{\mathrm{T}}\Big), \tag{11}$$

where Tr(.) and $(.)^{\mathrm{T}}$, respectively, represent the matrix trace and the matrix transpose. Using the properties provided in [8], the gradient expressions of the considered cost function are as follows (in matrix form), when disregarding the structure (9) of $\tilde{E}$ at this stage (it is taken into account further in this paper: see the discussion after (24))

$$\frac{\partial J}{\partial \tilde{E}} = -\big(X\tilde{C}^{\mathrm{T}} - (\tilde{A}\odot\tilde{E})\tilde{C}\tilde{C}^{\mathrm{T}}\big)\odot\tilde{A}, \tag{12}$$

$$\frac{\partial J}{\partial \tilde{A}} = -\big(X\tilde{C}^{\mathrm{T}} - (\tilde{A}\odot\tilde{E})\tilde{C}\tilde{C}^{\mathrm{T}}\big)\odot\tilde{E}, \tag{13}$$

$$\frac{\partial J}{\partial \tilde{C}} = -\big(\tilde{A}\odot\tilde{E}\big)^{\mathrm{T}}\big(X - (\tilde{A}\odot\tilde{E})\tilde{C}\big). \tag{14}$$

Taking into account that the proposed algorithm is a gradient-based one, the following iterative update rule is used

$$\theta \leftarrow \theta - \varphi_\theta \odot \frac{\partial J}{\partial \theta}, \tag{15}$$

where θ corresponds to one of the three considered matrices $\tilde{A}$ or $\tilde{E}$ or $\tilde{C}$, and φ_θ is a learning rate in matrix form. This update rule is not sufficient since it does not ensure nonnegativity. In order to satisfy this constraint, an iterative and multiplicative rule is designed from the above one as follows. From (12)-(14), it is clear that the derivative $\frac{\partial J}{\partial \theta}$ of J with respect to θ can be written as the difference of two nonnegative functions such that $\frac{\partial J}{\partial \theta} = \frac{\partial J}{\partial \theta}^{+} - \frac{\partial J}{\partial \theta}^{-}$. In the gradient expressions (12)-(14), the nonnegative function $\frac{\partial J}{\partial \theta}^{+}$ corresponds to the terms preceded by a plus sign, while $\frac{\partial J}{\partial \theta}^{-}$ corresponds to the terms preceded by a minus sign. The nonnegativity of θ throughout its adaptation can be satisfied by initializing θ with a nonnegative value and after that choosing the value of the learning rate matrix φ_θ according to

$$\varphi_\theta = \theta \oslash \frac{\partial J}{\partial \theta}^{+}, \tag{16}$$

where the operator $\oslash$ denotes the element-wise division. Consequently, the update rule (15) becomes

$$\theta \leftarrow \theta \odot \frac{\partial J}{\partial \theta}^{-} \oslash \frac{\partial J}{\partial \theta}^{+}. \tag{17}$$

The final proposed iterative and multiplicative update rules for the considered matrices read

$$\tilde{E} \leftarrow \tilde{E} \odot \big(X\tilde{C}^{\mathrm{T}}\big) \oslash \Big(\big(\tilde{A}\odot\tilde{E}\big)\tilde{C}\tilde{C}^{\mathrm{T}} + \varepsilon\Big), \tag{18}$$

$$\tilde{A} \leftarrow \tilde{A} \odot \big(X\tilde{C}^{\mathrm{T}}\big) \oslash \Big(\big(\tilde{A}\odot\tilde{E}\big)\tilde{C}\tilde{C}^{\mathrm{T}} + \varepsilon\Big), \tag{19}$$

$$\tilde{C} \leftarrow \tilde{C} \odot \Big(\big(\tilde{A}\odot\tilde{E}\big)^{\mathrm{T}}X\Big) \oslash \Big(\big(\tilde{A}\odot\tilde{E}\big)^{\mathrm{T}}(\tilde{A}\odot\tilde{\mathrm{E}})\tilde{C} + \varepsilon\Big), \tag{20}$$

where ε is a very small and positive value that is added to the denominator of each update rule to avoid division by zero.

Also, since r_{ij} defined by (3) should be lower than or equal to one and due to (4), the following constraints are also considered

$$\tilde{A}\odot\tilde{E} \leq 1^{L\times(MP)}, \tag{21}$$

$$\alpha^{L\times(MP)} \leq \tilde{A} \leq \beta^{L\times(MP)}. \tag{22}$$

Therefore, in the proposed algorithm the following constraints are also considered

$$\tilde{A} \leftarrow \max\{\tilde{A}, \alpha^{L\times(MP)}\}, \tag{23}$$

$$\tilde{A} \leftarrow \min\{\tilde{A}, \beta^{L\times(MP)}, 1^{L\times(MP)} \oslash (\tilde{E} + \varepsilon)\}. \tag{24}$$

In addition, and as mentioned in (8) and (9), the following additional constraints are taken into account. Each element of the block matrix $\tilde{A}_1$, of the updated matrix $\tilde{A}$, is constrained to take the one value, and each of the P-1 last M-column submatrices of $\tilde{E}$ is constrained to be equal to its first updated submatrix $\tilde{E}_1$.

Finally, the abundance sum-to-one constraint is fulfilled by using the technique described in [9].

The designed algorithm, as standard NMF techniques, does not guarantee that it provides a unique solution and its convergence point depends on its initialization. To keep away from random initialization from the viewpoint of the designed algorithm, and as the initialization stage, each initial spectral variability coefficient is fixed to one. Also, the first block matrix $\tilde{E}_1$ of $\tilde{E}$ is calculated by means of the vertex component analysis (VCA) method [1]. This first initial submatrix is replicated P-1 times to obtain the initial value of the matrix $\tilde{E}$. Besides, each initial non-zero value of the matrix $\tilde{C}$ is set to $1/M$.

A predefined maximum number of iterations is used as the stopping criterion of the proposed algorithm.

4. EXPERIMENTAL RESULTS

Experiments, based on realistic synthetic data, are carried out to evaluate the performance of the proposed algorithm. The obtained results are also compared to those of three methods from the literature: standard (i.e. without taking into account the spectral variability phenomenon) VCA [1], UPNMF and IPNMF [4], [5] that consider this phenomenon.

4.1. Tested data

Four sets of urban environment (photovoltaic panels, tiles, grass and trees) spectra with spectral variability, obtained by ground measurements, from 0.4 to 2.5 μm, using a spectrometer, with 214 wavelengths, are used to linearly create two 100-pixel realistic synthetic hyperspectral images. For each pixel of these two images, a spectrum, for each of the endmember classes, is randomly chosen and considered in the creation process.
In the first generated image, only three endmember classes (photovoltaic panels, tiles and grass) are considered, while in the second generated image, all four endmember classes are considered.
The abundance fractions used in the mixtures are created from a land cover classification, by averaging pixel classification values on a non-overlapping sliding window.

4.2. Performance evaluation criteria

The spectral angle mapper (SAM), the normalized mean square error (NMSE) and the spectral information divergence (SID) are used to evaluate the performance of the tested methods in terms of estimated spectra. A smaller value of these criteria indicates a better unmixing process.
For the designed algorithm and the UPNMF and IPNMF approaches, these criteria are used as follows. Each actual spectrum for an endmember-class is compared with all estimated spectra (i.e. in each pixel) for the same endmember-class, and the minimum criterion value is retained. Then, the mean of these minima, over all endmember classes, is the final considered criterion value.
For the VCA algorithm, for each endmember-class, the only estimated spectrum is compared with all actual spectra of the considered endmember-class, and the mean criterion value is retained. Then, the mean of these means, over all endmember classes, is the final considered criterion value.

4.3. Results and discussion

Hereafter, the mean results provided by the tested methods are given for the two ten-times generated images.
All tested methods are initialized with the appropriate same matrix values.
For the proposed algorithm, all coefficients of the matrix α (resp. β) defined in (4), which controls the minimum (resp. maximum) variability degree, are set to 0.5 (resp. 1.5).
The maximum number of iterations used in each tested algorithm is set to 100.
The next tables show the values of the considered performance criteria for the two generated images.

Table I. Values of the considered performance criteria for the first generated image.

	Proposed	VCA	UPMNF	IPMNF
SAM (°)	**5.62**	10.41	6.17	6.18
NMSE (%)	**16.55**	30.29	19.15	19.24
SID	**1.78**	3.27	2.18	2.20

Table II. Values of the considered performance criteria for the second generated image.

	Proposed	VCA	UPMNF	IPMNF
SAM (°)	**4.53**	10.93	6.42	6.44
NMSE (%)	**12.63**	23.56	19.18	19.25
SID	**1.12**	5.08	2.20	2.21

Globally, the above tables prove that the proposed hyperspectral unmixing algorithm, with spectral variability, achieves much better performance than the tested methods from the literature. Indeed, from these tables, it is clear that the proposed method provides the best SAM, NMSE and SID criteria values.

5. CONCLUSION

In this paper, a new hyperspectral unmixing method, based on a new data mixing model, is designed. The proposed algorithm, which is based on nonnegative matrix factorization, considers the spectral variability phenomenon modeled in a multiplicative form. This designed algorithm uses iterative and multiplicative update rules.
Compared with methods from the literature, and according to the obtained results, the proposed approach proves to be very attractive for unmixing hyperspectral remote sensing data with spectral variability.
Future extensions of this work will especially consist in performing further tests by applying the proposed approach to real data, and finding a way to better constrain the spectral variability coefficients for better spectra estimation.

6. REFERENCES

[1] J.M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral Unmixing Overview: Geometrical, Statistical, and Sparse Regression-Based Approaches," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5(2), pp. 354-379, 2012.

[2] Y. Deville, "Blind Source Separation and Blind Mixture Identification Methods," *Wiley Encyclopedia of Electrical and Electronics Engineering*, pp. 1-33. 2016.

[3] P. Comon and C. Jutten, *Handbook of blind source separation: Independent component analysis and applications*. Academic Press, Oxford, UK, 2010.

[4] Y. Deville, C. Revel, V. Achard, and X. Briottet, "Application and extension of PCA concepts to blind unmixing of hyperspectral data with intra-class variability," *in "Advances in Principal Component Analysis - Research and Development"*, pp. 225-252, G. R. Naik (ed), Springer, Singapore, 2018.

[5] C. Revel, Y. Deville, V. Achard, X. Briottet, and C. Weber, "Inertia-constrained pixel-by-pixel nonnegative matrix factorisation: a hyperspectral unmixing method dealing with intra-class variability," *Remote Sensing*, vol. 10(11), 1706, 2018. https://doi.org/10.3390/rs10111706

[6] A. Zare and K.C. Ho, "Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing," *IEEE Signal Processing Magazine*, vol. 31(1), pp. 95-104, 2014.

[7] A. Cichocki, R. Zdunek, A.H. Phan, and S.-I. Amari, *Nonnegative matrix and tensor factorizations. Applications to exploratory multi-way data analysis and blind source separation*, Wiley, Chichester, UK, 2009.

[8] K.B. Petersen and M.S. Pedersen, *The matrix Cookbook*, Technical Univ. Denmark, 2008 [Online]. Available: URL:http://matrixcook-book.com/

[9] D.C. Heinz, 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*, vol. 39(3), pp. 529-545, 2001.