NMF-BASED HYPERSPECTRAL UNMIXING WITH A SIMPLE EXTENDED LINEAR MIXING MODEL ADDRESSING SPECTRAL VARIABILITY

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

[1] Agence Spatiale Algérienne, Centre des Techniques Spatiales, Arzew, Algeria
[2] Univ Toulouse, CNES, CNRS, IRAP, Toulouse, France

ABSTRACT

Remote sensing hyperspectral data are typically subject to the spectral variability phenomenon. When considering the standard linear mixing model that ignores this issue, the unmixing process of such data can hardly provide accurate estimates of endmember spectra and their corresponding abundance fractions. Recently, more advanced linear mixing models that address spectral variability were developed. Such models, and their more complex associated unmixing algorithms, involve huge adaptive matrices, thus limiting their applicability to large hyperspectral images. In this investigation, a simple extended linear mixing model is first proposed to address the spectral variability phenomenon. This model involves computationally tractable adaptive matrices, which obey the nonnegativity constraint. The proposed associated unmixing algorithm, based on nonnegative matrix factorization, optimizes a penalized cost function with iterative and multiplicative update rules. The conducted experiments prove that the designed unmixing algorithm yields much better results than several tested state-of-the-art methods.

Index Terms— Hyperspectral data, spectral/intra-class variability, linear mixing model, hyperspectral unmixing, nonnegative matrix factorization.

1. INTRODUCTION

The number of hyperspectral sensors has recently considerably increased and they have significant capabilities, which no longer need to be demonstrated, in various applications related to Earth observation. However, many such sensors remain characterized by a lower spatial resolution than that of multispectral ones. Consequently, many mixed pixels are present in hyperspectral data, thus preventing their correct analysis.

The basic linear unmixing process constitutes a conventional tool due to its simplicity, and accordingly various unmixing techniques, based on the standard Linear Mixing Model (LMM), were designed to estimate endmember spectra and their associated abundance fractions composing mixed pixels [1].

The standard LMM does not take into account the spectral variability phenomenon, which is one of the major issues when processing hyperspectral data. Recently, more advanced LMMs and associated unmixing algorithms, which address this phenomenon, were developed [2]. In several such sophisticated LMMs [3]-[5], this phenomenon is considered for spectral variables by using huge varying matrices. These matrices are involved in associated unmixing algorithms based on Nonnegative Matrix Factorization (NMF) [6] and limit the applicability of the latter algorithms to large hyperspectral images. Besides, the same phenomenon is considered for spatial variables in other advanced LMMs [7]-[9]. These models consider either huge varying matrices or matrices that do not obey the nonnegativity constraint, consequently leading to the use of more complex unmixing algorithms, in particular the alternating direction method of multipliers (ADMM) [10] technique, preventing therefore, in both cases, their applicability to large hyperspectral data.

This investigation first introduces a simple extended LMM to deal with the spectral variability phenomenon. The proposed model, wherein the spectral variability phenomenon is considered for spatiospectral variables, has all the flexibility required to address this phenomenon by means of computationally tractable varying matrices that obey the nonnegativity constraint. Indeed, as compared with the standard LMM, the introduced model uses two additional nonnegative varying matrices, of the same size as the observed image matrix, in order to take into account possible spectral variations in an observed scene. As a second step of this work, an NMF-based unmixing algorithm is designed for the introduced mixing model. It optimizes a penalized cost function by means of iterative and multiplicative update rules.

This paper is structured as follows. In Section 2, the introduced mixing model is presented. Section 3 describes the designed NMF-based unmixing algorithm. Section 4 involves experiments performed with synthetic and real hyperspectral data. Finally, Section 5 concludes this investigation.

2. PROPOSED SIMPLE EXTENDED LMM

The proposed simple extended model adds, to the standard LMM, two nonnegative matrices that allow variations around that standard LMM, which may represent spectral variability in an observed scene. Thus, a hyperspectral image $X \in \mathbb{R}_+^{K \times P}$

is modeled as:

$$X = EA + U - L. \quad (1)$$

Each column (respectively row) of X is associated with one pixel spectrum (respectively spectral band). Thus, P (respectively K) corresponds to the number of pixels (respectively spectral bands). Each column of $E \in \mathbb{R}_+^{K\times M}$ is associated with one endmember spectrum (M is the number of endmembers), and each row of $A \in \mathbb{R}_+^{M\times P}$ represents the abundance fractions of one endmember in all pixels. The two additional matrices $U \in \mathbb{R}_+^{K\times P}$ and $L \in \mathbb{R}_+^{K\times P}$, which have the same size as the image matrix X, have all the flexibility required for dealing with spectral variability. Defined in this way, upper and lower spectral variations (respectively above and below a standard contribution given by the standard model with no variations, i.e. EA) are respectively contained in the matrices U and L, which are hereafter forced to obey the nonnegativity constraint. On the other hand, if the manipulated image contains no spectral variability but does contain noise, U and L will take this noise into account.

It should here be mentioned that sophisticated LMMs [3]-[5], [7]-[9] address spectral variability by means of huge varying matrices, some of which can reach a size of $(PM) \times P$, and that do not necessarily meet the nonnegativity constraint, leading to the need for more computational resources and/or more complex unmixing algorithms. In contrast, the introduced model, which also deals with spectral variability, only involves nonnegative varying matrices whose size is sufficiently small for them to be computationally tractable: the largest ones have the same size as the observed image matrix X, i.e. a size of $K \times P$. Such involved nonnegative varying matrices thus lead to the development of a less complex NMF-based unmixing algorithm described hereafter.

3. PROPOSED NMF-BASED HYPERSPECTRAL UNMIXING ALGORITHM

The proposed algorithm employs an iterative and multiplicative NMF [6] technique. It optimizes the variables $\tilde{E}$, $\tilde{A}$, $\tilde{U}$ and $\tilde{L}$, which respectively estimate E, A, U and L, by minimizing the following penalized cost function:

$$J = \frac{1}{2}\left(\left\|X - \tilde{E}\tilde{A} - \tilde{U} + \tilde{L}\right\|_{\mathrm{F}}^2 + \left\|\tilde{U}\right\|_{\mathrm{F}}^2 + \left\|\tilde{L}\right\|_{\mathrm{F}}^2\right), \quad (2)$$

where $\|.\|_{\mathrm{F}}$ represents the Frobenius norm. The first term optimizes the difference between the observed data matrix and that estimated by the introduced model (1). The last two penalization terms (which might be weighted in a modified version of this approach to reduce or increase their contributions) allow constraining, in all observed pixels, upper and lower spectral variability coefficients to be low.

The optimization is performed by means of gradient-based update rules. The gradient expressions of J, in matrix form, are as follows:

$$\frac{\partial J}{\partial \tilde{E}} = \tilde{E}\tilde{A}\tilde{A}^{\mathrm{T}} + \tilde{U}\tilde{A}^{\mathrm{T}} - X\tilde{A}^{\mathrm{T}} - \tilde{L}\tilde{A}^{\mathrm{T}}, \quad (3)$$

$$\frac{\partial J}{\partial \tilde{A}} = \tilde{E}^{\mathrm{T}}\tilde{E}\tilde{A} + \tilde{E}^{\mathrm{T}}\tilde{U} - \tilde{E}^{\mathrm{T}}X - \tilde{E}^{\mathrm{T}}\tilde{L}, \quad (4)$$

$$\frac{\partial J}{\partial \tilde{U}} = \tilde{E}\tilde{A} + 2\tilde{U} - X - \tilde{L}, \quad (5)$$

$$\frac{\partial J}{\partial \tilde{L}} = 2\tilde{L} + X - \tilde{U} - \tilde{E}\tilde{A}, \quad (6)$$

where $(.)^{\mathrm{T}}$ denotes the matrix transpose. Using the classical process described in [11], the following iterative and multiplicative update rules, for the involved adaptive matrices, are designed. Other rules are described further in this paper to satisfy some physical constraints.

To optimize the varying matrix $\tilde{E}$, the following update rule is designed:

$$\tilde{E} \leftarrow \tilde{E} \odot (X\tilde{A}^{\mathrm{T}} + \tilde{L}\tilde{A}^{\mathrm{T}}) \oslash (\tilde{E}\tilde{A}\tilde{A}^{\mathrm{T}} + \tilde{U}\tilde{A}^{\mathrm{T}} + \varepsilon), \quad (7)$$

where $\odot$ (respectively $\oslash$) represents the element-wise multiplication (respectively division), ε is a very small and positive value added to prevent a possible division by zero.

To adapt the varying matrix $\tilde{A}$, the following update rule is defined:

$$\tilde{A} \leftarrow \tilde{A} \odot (\tilde{E}^{\mathrm{T}}X + \tilde{E}^{\mathrm{T}}\tilde{L}) \oslash (\tilde{E}^{\mathrm{T}}\tilde{E}\tilde{A} + \tilde{E}^{\mathrm{T}}\tilde{U} + \varepsilon). \quad (8)$$

The latter rule is supplemented by the use of the abundance sum-to-one constraint: each element of a column of this varying matrix is divided by the sum of all element values of that column. Thus, for a given $j = 1 \dots P$, the following update is used after (8):

$$\tilde{a}_{ij} \leftarrow \frac{\tilde{a}_{ij}}{\mathrm{sum}(\tilde{a}_{.j})}, \forall\, i = 1 \dots M \text{ for a given } j = 1 \dots P. \quad (9)$$

The update rule for optimizing the varying matrix $\tilde{U}$ is given by:

$$\tilde{U} \leftarrow \tilde{U} \odot (X + \tilde{L}) \oslash (\tilde{E}\tilde{A} + 2\tilde{U} + \varepsilon). \quad (10)$$

Since each element of $(\tilde{E}\tilde{A} + \tilde{U} - \tilde{L})$ has to be lower than or equal to one and higher than or equal to zero, and each element of $\tilde{U}$ is nonnegative, $\tilde{U}$ is then readjusted using the following rules:

$$\begin{aligned} &\tilde{U} \leftarrow \min\left(\tilde{U},\, 1^{K\times P} - \tilde{E}\tilde{A} + \tilde{L}\right) \\ &\text{and } \tilde{U} \leftarrow \max\left(\tilde{U}, \tilde{L} - \tilde{E}\tilde{A}, 0^{K\times P}\right). \end{aligned} \quad (11)$$

Then, the update rule for adapting the varying matrix $\tilde{L}$ is expressed as:

$$\tilde{L} \leftarrow \tilde{L} \odot (\tilde{U} + \tilde{E}\tilde{A}) \oslash (2\tilde{L} + X + \varepsilon). \quad (12)$$

Likewise, since each element of $(\tilde{E}\tilde{A} + \tilde{U} - \tilde{L})$ has to be

lower than or equal to one and higher than or equal to zero, and each element of $\tilde{L}$ is nonnegative, the subsequent rules are used to readjust $\tilde{L}$:

$$\tilde{L} \leftarrow \max\left(\tilde{L}, \tilde{E}\tilde{A} + \tilde{U} - 1^{K\times P}, 0^{K\times P}\right) \quad \text{and } \tilde{L} \leftarrow \min\left(\tilde{L}, \tilde{E}\tilde{A} + \tilde{U}\right). \tag{13}$$

To avoid a random initialization of the adaptive matrices of the designed algorithm, the initial endmember spectra, contained in the columns of $\tilde{E}$, may be obtained by means of the Vertex Component Analysis (VCA) [1] algorithm. Additionally, each initial abundance fraction value, contained in $\tilde{A}$, may be set to $1/M$ or may be estimated by using the Fully Constrained Least Squares (FCLS) [1] technique. Moreover, each initial upper (respectively lower) variation coefficient value, contained in $\tilde{U}$ (respectively $\tilde{L}$), may be set to a predefined value.

Following the above initialization stage of the designed algorithm, the optimization stage consists of updating all the varying matrices using the aforementioned rules (7)-(13) until convergence, which is defined by the following conditions: a predefined threshold reached by the relative change between two successive values of J:

$$\left|\frac{J^{(t)} - J^{(t+1)}}{J^{(t)}}\right| \leq thresh, \tag{14}$$

where t is an iteration, and a predefined maximum number of iterations.

4. EXPERIMENT RESULTS

The conducted experiments, to assess the performances of the designed algorithm (called "NMF-SELMM-Var"), which deals with spectral variability and is inferred from the introduced simple extended LMM, are described here. The results, based on two synthetic and two real data sets and obtained by means of the NMF-SELMM-Var algorithm, are compared to those achieved by several tested state-of-the-art methods.

4.1. Tested data

For the first synthetic data, three sets of pure spectra (vegetation, soil and water), with 198 wavelengths and with spectral variability [2], are used to generate 50×50-pixel synthetic hyperspectral data sets. These spectra are linearly mixed with abundance fractions generated according to a Gaussian random field model [2]. More specifically, each observed pixel spectrum is obtained by linearly mixing the considered pure spectra, after randomly selecting one spectrum for each pure material, with associated abundance fractions. In addition, additive noise, with a signal-to-noise ratio value of 30 decibels, is introduced.

The second synthetic image, with a size of 50×50 pixels, is generated from a 100×100-pixel subset of the real hyperspectral image described in [12]. More precisely, seven groups of pure spectra (grass, tree, soil, residential, commercial, road and parking), with 144 spectral bands and with spectral variability, are created from ground truth samples provided with this image subset. These spectra are linearly mixed, following the same process as above, with abundance fraction maps created from a land cover classification of the considered real image subset, using averaging over a non-overlapping 2×2-pixel sliding window [7]. Here too, white Gaussian noise with a signal-to-noise-ratio value of 30 decibels is considered.

The above-mentioned 100×100-pixel real image subset is also considered, in this investigation, as a first tested real image. The second tested real image that contains four classes (vegetation, metallic roofs, concrete and asphalt) and covers the Robertson stadium of the Houston Campus and its surroundings, with a size of 105×128 pixels and with 144 spectral bands, is a subset also obtained from the real image described in [12].

4.2. Performance evaluation criteria

For the tested synthetic data, the average Spectral Angle Mapper (aSAM), the abundance overall Root Mean Square Error (aRMSE), and the reconstruction RMSE (rRMSE) are used as performance evaluation criteria of the tested algorithms. It should here be noted that the aRMSE criterion can be considered only when the ground truth of abundance fractions is available. For the tested real data, since the ground truth of abundance fractions was not available, only the aSAM and rRMSE criteria were used to evaluate the performances of the tested unmixing algorithms. The computation time is also considered as a performance evaluation criterion of tested algorithms, both on synthetic and real data.

4.3. Experiments description

The NMF-SELMM-Var algorithm is tested with the above-described synthetic and real hyperspectral data. Moreover, the literature methods (used with their default parameters), which are related to the Augmented LMM (ALMM) [8] and Extended LMM (ELMM) [9] models and deal with spectral variability, are tested. Furthermore, the standard VCA [1] technique combined with the FCLS [1] algorithm, which do not deal with spectral variability, are also tested as a baseline approach.

The tested algorithms which address spectral variability are initialized in the same way: the initial endmember spectra are calculated by means of the VCA [1] technique, and each initial abundance fraction value is set to $1/M$. The maximum number of iterations considered in all tested iterative algorithms is set to 100. The threshold value defined in (14) is set to 10^{-6} for the NMF-SELMM-Var algorithm.

Finally, experiments are conducted by using an Intel(R) Core(TM) i5-8250U CPU running at 1.60 GHz and a memory capacity of 20 GB.

4.4. Results and discussion

The obtained performance criterion values are given in the following tables (Tables I–IV). These values are the means and standard deviations obtained over 100 runs in order to reduce the random effect present in the creation of synthetic data, and present in the initialization stage (when considering the VCA technique) of tested algorithms.

Globally, these results prove that the NMF-SELMM-Var algorithm achieves much better accuracies than those obtained by the tested literature unmixing algorithms. For the synthetic data, Tables I–II show the overall superiority of the proposed algorithm, in particular when considering the aSAM and rRMSE criteria. For the real data, Tables III–IV also indicate the overall dominance of the proposed unmixing algorithm. Furthermore, Tables I–IV also prove that the designed algorithm is faster than the tested literature ones which deal with spectral variability. All obtained performances clearly indicate the attractiveness of the NMF-SELMM-Var algorithm for unmixing hyperspectral data, in particular when they have large dimensions, while dealing with the spectral variability phenomenon.

Table I. Mean and standard deviation values of the considered performance criteria for the first synthetic data set.

	NMF-SELMM-Var	ALMM-based	ELMM-based	VCA+FCLS
aSAM (°)	**3.87 ± 0.02**	5.98 ± 0.09	4.31 ± 0.35	8.61 ± 0.07
aRMSE	**$5.21\times10^{-2} \pm 6.01\times10^{-4}$**	$5.27\times10^{-2} \pm 1.10\times10^{-3}$	$0.10 \pm 1.20\times10^{-3}$	$0.41 \pm 6.66\times10^{-4}$
rRMSE	**$4.80\times10^{-3} \pm 4.27\times10^{-5}$**	$0.64 \pm 2.30\times10^{-3}$	$5.90\times10^{-3} \pm 4.32\times10^{-4}$	$0.01 \pm 2.45\times10^{-4}$
Time (s)	7.06 ± 1.08	24.95 ± 2.70	12.59 ± 1.04	**0.43 ± 0.17**

Table II. Mean and standard deviation values of the considered performance criteria for the second synthetic data set.

	NMF-SELMM-Var	ALMM-based	ELMM-based	VCA+FCLS
aSAM (°)	**5.68 ± 0.04**	10.67 ± 0.12	7.70 ± 0.07	19.99 ± 4.41
aRMSE	$0.25 \pm 3.91\times10^{-4}$	$0.40 \pm 7.96\times10^{-4}$	**$0.20 \pm 4.37\times10^{-4}$**	0.68 ± 0.56
rRMSE	**$0.70\times10^{-2} \pm 4.49\times10^{-5}$**	$0.01 \pm 1.52\times10^{-4}$	$0.83\times10^{-2} \pm 3.06\times10^{-5}$	0.99 ± 0.14
Time (s)	10.01 ± 0.85	27.43 ± 2.12	15.65 ± 1.81	**0.66 ± 0.24**

Table III. Mean and standard deviation values of the considered performance criteria for the first real data set.

	NMF-SELMM-Var	ALMM-based	ELMM-based	VCA+FCLS
aSAM (°)	**4.30 ± 0.05**	11.16 ± 1.48	7.94 ± 0.9912	18.37 ± 12.36
rRMSE	**$0.51\times10^{-2} \pm 1.25\times10^{-4}$**	$0.01 \pm 0.30\times10^{-2}$	$0.91\times10^{-2} \pm 0.13\times10^{-2}$	0.92 ± 0.10
Time (s)	30.39 ± 1.49	902.77 ± 52.17	40.10 ± 4.71	**2.26 ± 0.49**

Table IV. Mean and standard deviation values of the considered performance criteria for the second real data set.

	NMF-SELMM-Var	ALMM-based	ELMM-based	VCA+FCLS
aSAM (°)	**2.03 ± 0.13**	3.97 ± 1.61	4.03 ± 0.26	7.80 ± 1.28
rRMSE	**$0.36\times10^{-2} \pm 2.00\times10^{-4}$**	0.77 ± 0.03	$0.48\times10^{-2} \pm 3.40\times10^{-4}$	$0.02 \pm 0.43\times10^{-2}$
Time (s)	30.92 ± 1.37	$1.99\times10^{+3} \pm 125.65$	36.45 ± 0.82	**2.08 ± 0.35**

5. CONCLUSION

In this paper, a simple extended linear mixing model, with its associated unmixing algorithm, which addresses the spectral variability phenomenon, was proposed for hyperspectral unmixing. The designed NMF-based algorithm optimizes a penalized cost function by means of iterative and multiplicative update rules.

The proposed algorithm was tested on synthetic and real hyperspectral data, and its efficiency was evaluated using conventional criteria. The obtained results clearly indicate its high superiority in comparison with tested literature algorithms.

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] R.A. Borsoi, T. Imbiriba, J.C.M. Bermudez, C. Richard, J. Chanussot, L. Drumetz, J.-Y. Tourneret, A. Zare, and C. Jutten, "Spectral Variability in Hyperspectral Data Unmixing: A Comprehensive Review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 9(4), pp. 223–270, 2021.

[3] 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), pp. 1706, 2018.

[4] S.E. Brezini, Y. Deville, M.S. Karoui, F.Z. Benhalouche, and A. Ouamri, "A Penalization-based NMF Approach for Hyperspectral Unmixing Addressing Spectral Variability with an Additively-Tuned Mixing Model," *IEEE International Geoscience and Remote Sensing Symposium* (IGARSS 2021), Brussels, Belgium–Virtual Conference, pp. 3841–3844, July 12-16, 2021.

[5] F.Z. Benhalouche, M.S. Karoui, and Y. Deville, "Gradient-based NMF Methods for Hyperspectral Unmixing Addressing Spectral Variability with a Multiplicative-Tuning Linear Mixing Model," *IEEE International Geoscience and Remote Sensing Symposium* (IGARSS 2021), Brussels, Belgium–Virtual Conference, pp. 3841-3844, July 12–16, 2021.

[6] 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*, John Wiley & Sons, Chichester, UK, 2009.

[7] M.S. Karoui, F.Z. Benhalouche, and Y. Deville, "Hyperspectral Unmixing with a Modified Augmented Linear Mixing Model Addressing Spectral Variability," *IEEE Mediterranean and Middle-East Geoscience and Remote Sensing Symposium* (M2GARSS 2022), Virtual Conference, March 7-9, 2022.

[8] D. Hong, N. Yokoya, J. Chanussot, and X.X. Zhu, "An Augmented Linear Mixing Model to Address Spectral Variability for Hyperspectral Unmixing," *IEEE Transactions on Image Processing*, vol. 28(4), pp. 1923–1938, 2019.

[9] L. Drumetz, M. A. Veganzones, S. Henrot, R. Phlypo, J. Chanussot, and C. Jutten, "Blind Hyperspectral Unmixing Using an Extended Linear Mixing Model to Address Spectral Variability," *IEEE Transactions on Image Processing*, vol. 25(8), pp. 3890–3905, 2016.

[10] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundation and Trends® in Machine Learning*, vol. 3(1), pp. 1–122, Now Publishers Inc., 2011.

[11] M.S. Karoui, F.Z. Benhalouche, Y. Deville, K. Djerriri, X. Briottet, T. Houet, A. Le Bris, and C. Weber, "Partial Linear NMF-Based Unmixing Methods for Detection and Area Estimation of Photovoltaic Panels in Urban Hyperspectral Remote Sensing Data," *Remote Sensing*, vol. 11(18), pp. 2164, 2019.

[12] C. Debes, A. Merentitis, R. Heremans, J. Hahn, N. Frangiadakis, T. van Kasteren, W. Liao, R. Bellens, A. Pizurica, S. Gautama, W. Philips, S. Prasad, Q. Du, and F. Pacifici, "Hyperspectral and LiDAR Data Fusion: Outcome of the 2013 GRSS Data Fusion Contest," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7(6), pp. 2405–2418, 2014.