A MAP-BASED NMF APPROACH TO HYPERSPECTRAL IMAGE UNMIXING USING A LINEAR-QUADRATIC MIXTURE MODEL

Lina JARBOUI[1,2], Shahram HOSSEINI[1], Rima GUIDARA[2], Yannick DEVILLE[1], and Ahmed BEN HAMIDA[2]

[1]Institut de Recherche en Astrophysique et Planétologie (IRAP)
Toulouse University, UPS-OMP, CNRS, Toulouse, France

[2]Advanced Technologies for Medicine and Signals (ATMS)
Sfax University, ENIS, Sfax, Tunisia

ABSTRACT

In this paper, we address the problem of spectral unmixing in urban hyperspectral images using a Maximum A Posteriori (MAP)-based Non-negative Matrix Factorization (NMF) approach. Considering a Linear-Quadratic (LQ) mixing model, we seek to decompose the spectrum observed in each pixel of the image into a set of pure material spectra, as well as their abundance fractions and the mixing coefficients associated with products of these pure material spectra. The main idea of the proposed method is to take into account the available prior information about the unknown parameters for a better estimation of them. To this end, we first derive a MAP-based cost function, then minimize it using a projected gradient algorithm by modifying a recently proposed NMF method adapted to LQ mixtures. Simulation results confirm the relevance of our approach.

Index Terms— Unsupervised spectral unmixing, Linear-Quadratic mixture, Non-negative Matrix Factorization, Maximum A Posteriori estimation, Hyperspectral image

1. INTRODUCTION

In the recent years, the growing use of hyperspectral imaging instruments has drawn the signal and image processing community's attention to the crucial importance of hyperspectral unmixing. Because of the limited spatial resolution of hyperspectral images, the spectrum observed at each pixel of these images is often a mixture of the spectra of pure materials (endmembers) present in that pixel. Most existing spectral unmixing methods assume that the mixing model is linear, i.e. the observed pixel spectra result from linear combinations of endmembers (see [1] and [2] for a review). Nevertheless, in areas of variable topographic relief, the mixture model is most likely to be nonlinear due to the 3D structures that induce multiple scattering of light between various surfaces [3].

Recently, physical studies presented in [4] showed the relevance of the Linear-Quadratic (LQ) mixture model to describe this mixing phenomenon in urban environments. Thus, considering an urban hyperspectral image acquired in N spectral bands and composed of K pixels, it has been shown that the reflectance spectrum observed in a pixel i of such an image may be approximated as an LQ mixture of reflectance spectra of L pure materials (sources) present in the observed scene, as stated in the following equation:

$$\mathbf{x}_i = \sum_{j=1}^{L} a_j(i)\mathbf{s}_j + \sum_{j=1}^{L}\sum_{k=j}^{L} a_{j,k}(i)\mathbf{s}_j \odot \mathbf{s}_k + \mathbf{n}_i, \qquad (1)$$

where $\mathbf{x}_i = [x_{i1}, ..., x_{iN}]^T$ is the spectrum associated with the observed pixel i, $\mathbf{s}_j = [s_{j1}, ..., s_{jN}]^T$ is the spectrum of the pure material j, $a_j(i)$ is the abundance fraction of the pure material j in the pixel i, $a_{j,k}(i)$ is the quadratic mixing coefficient associated with the pure materials j and k in the pixel i, the symbol $\odot$ represents the term by term Hadamard product, and $\mathbf{n}_i$ is an additive noise, assumed to be Gaussian, zero-mean, independent and identically distributed (i.i.d.) and of variance σ_n^2. The numerical studies carried out in [4] showed that the following physical constraints are met:

$$\begin{cases} s_{jl} \geq 0 \\ a_j(i) \geq 0 \\ \sum_{j=1}^{L} a_j(i) = 1 \\ a_{j,k}(i) \in [0, 0.5] \end{cases} \forall \begin{cases} 1 \leq i \leq K \\ 1 \leq j \leq L \\ j \leq k \leq L \\ 1 \leq l \leq N \end{cases} \qquad (2)$$

Only a few hyperspectral unmixing methods adapted to LQ mixtures in the field of remote sensing have been proposed in the literature (see [3] and [5] for a review). One of these methods, which we proposed in [6], is based on the non-negativity of data involved in the mixing process, and proposes an extension of Non-negative Matrix Factorization (NMF) [7, 8] adapted to the LQ model. Like the original linear NMF, this extension to the LQ model suffers from the well-known non-uniqueness of the solution and the possible convergence of the algorithms towards spurious minima. To constrain the optimization problem, and so, to increase the chance of finding the right solution, the available prior information about sources or mixing parameters may be used. In this paper, considering an LQ mixing model, we modify

the above-mentioned LQ-NMF method so as to take into account this prior information, and we propose an iterative projected gradient algorithm to estimate the parameters of the prior distributions together with the spectra and the mixing parameters.

2. NMF APPROACH FOR LQ MIXTURES

By gathering spectra observed in all the pixels, Eq. (1) can be rewritten in the following matrix form

$$\mathbf{X} = \mathbf{AS} + \mathbf{N}, \tag{3}$$

where $\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_K]^T$, $\mathbf{S}$ is a matrix containing the *pseudo-sources*, i.e. the actual sources and their products: $\mathbf{S} = [\mathbf{s}_1 \cdots \mathbf{s}_L \;\; \mathbf{s}_1 \odot \mathbf{s}_1 \;\; \mathbf{s}_1 \odot \mathbf{s}_2 \cdots \mathbf{s}_L \odot \mathbf{s}_L]^T$, $\mathbf{N} = [\mathbf{n}_1 \cdots \mathbf{n}_K]^T$, and $\mathbf{A}$ is a matrix containing the abundance fractions and quadratic coefficients:

$$\mathbf{A} = \begin{pmatrix} a_1(1) \cdots a_L(1) & a_{1,1}(1) & a_{1,2}(1) \cdots a_{L,L}(1) \\ \vdots & & \vdots \\ a_1(K) \cdots a_L(K) & a_{1,1}(K) & a_{1,2}(K) \cdots a_{L,L}(K) \end{pmatrix}.$$

According to Eq. (2), matrices $\mathbf{A}$ and $\mathbf{S}$ in (3) are non-negative, which suggests the use of NMF-based methods to unmix data. In [6], supposing the lack of any additional useful information about the sources and the mixing parameters, we proposed a gradient-based LQ-NMF algorithm to minimize the Frobenius (squared) norm $||\mathbf{X} - \mathbf{AS}||_F^2$ subject to $\mathbf{A} \geq 0$ and $\mathbf{S} \geq 0$, while taking into account the particular structure of $\mathbf{S}$, i.e. the fact that the last rows of $\mathbf{S}$ are the Hadamard products of its first rows. This structure was exploited for calculating the derivatives of the Frobenius norm with respect to the actual sources (see [6] for detailed calculations). In the following, we show how available prior information may be used to improve this method by deriving a Maximum A Posteriori (MAP) estimator.

3. PROPOSED METHOD

3.1. MAP estimation

The MAP estimator for the unknown matrices $\mathbf{A}$ and $\mathbf{S}$ reads

$$\begin{aligned} \{\hat{\mathbf{A}}, \hat{\mathbf{S}}\} &= \underset{\mathbf{A},\mathbf{S}}{\operatorname{argmax}} f(\mathbf{A}, \mathbf{S}|\mathbf{X}) \\ &= \underset{\mathbf{A},\mathbf{S}}{\operatorname{argmax}} f(\mathbf{X}|\mathbf{A}, \mathbf{S}) f(\mathbf{A}) f(\mathbf{S}), \end{aligned} \tag{4}$$

where f stands for Probability Density Function (PDF). Since no prior information about the spectra (except their non-negativity) is available, a non-informative uniform (i.e. constant) prior distribution can be assigned to them, so that

$$\begin{aligned} \{\hat{\mathbf{A}}, \hat{\mathbf{S}}\} &= \underset{\mathbf{A},\mathbf{S}}{\operatorname{argmax}} f(\mathbf{X}|\mathbf{A}, \mathbf{S}) f(\mathbf{A}) \\ &= \underset{\mathbf{A},\mathbf{S}}{\operatorname{argmax}} \left[\log f(\mathbf{X}|\mathbf{A}, \mathbf{S}) + \log f(\mathbf{A})\right]. \end{aligned} \tag{5}$$

The noise components in (3) being spectrally and spatially i.i.d., Gaussian, zero-mean and of variance σ_n^2, the likelihood $f(\mathbf{X}|\mathbf{A}, \mathbf{S})$ in (5) and its logarithm respectively read

$$f(\mathbf{X}|\mathbf{A}, \mathbf{S}) = \prod_{i=1}^{K} \prod_{l=1}^{N} \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(\frac{-(X - AS)_{il}^2}{2\sigma_n^2}\right), \tag{6}$$

$$\log f(\mathbf{X}|\mathbf{A}, \mathbf{S}) = C - \frac{1}{2\sigma_n^2} ||\mathbf{X} - \mathbf{AS}||_F^2, \tag{7}$$

where $(X - AS)_{il} = x_{il} - \sum_{j=1}^{L} a_j(i) s_{jl} - \sum_{j=1}^{L} \sum_{k=j}^{L} a_{j,k}(i) s_{jl} s_{kl}$, and C is a constant.

3.2. Prior information about mixing parameters

According to the sum-to-one constraint mentioned in (2), the abundance fractions in each pixel are dependent. Assuming that the quadratic coefficients are independent from the abundance fractions and from each other, and that the mixing parameters related to a pixel are independent from those related to another pixel, we can write

$$f(A) = \prod_{i=1}^{K} \left[f(a_1(i), \cdots, a_L(i)) \prod_{j=1}^{L} \prod_{k=j}^{L} f(a_{j,k}(i)) \right]. \tag{8}$$

The abundance fractions in each pixel are non-negative and sum to one. Thus, they may be modeled using Dirichlet distributions which satisfy the above properties [9, 10]:

$$f(a_1(i), \cdots, a_L(i)) = \frac{\Gamma(\sum_{j=1}^{L} \theta_j)}{\prod_{j=1}^{L} \Gamma(\theta_j)} \prod_{j=1}^{L} a_j(i)^{\theta_j - 1}, \tag{9}$$

where Γ is the Gamma function and θ_j are the Dirichlet parameters.

Numerical studies in an urban environment carried out in [4] showed that the quadratic coefficients may be modeled by a decreasing PDF: in many pixels the quadratic terms of mixtures are negligible and take small values while they take their highest values (0.4-0.5) only in a few pixels. Thus, we decided to assign a half-normal distribution to the quadratic coefficients as follows:

$$f(a_{j,k}(i)) = \frac{2\vartheta_{j,k}}{\pi} \exp\left(\frac{-(a_{j,k}(i))^2 \vartheta_{j,k}^2}{\pi}\right), \forall a_{j,k}(i) \geq 0 \tag{10}$$

where $\vartheta_{j,k} = \dfrac{\sqrt{\pi}}{\sigma_{j,k}\sqrt{2}}$ is the parameter of this PDF.

3.3. Cost function

By inserting (9) and (10) in (8), then computing its logarithm and using (5) and (7), it can easily be shown that the MAP

estimator for $\mathbf{A}$ and $\mathbf{S}$ may be obtained by minimizing the following cost function:

$$J = \frac{1}{2}||\mathbf{X} - \mathbf{AS}||_F^2 - \eta R \tag{11}$$

where R is a regularization term related to prior information and defined by

$$\begin{aligned} R &= K\log\Gamma(\sum_{j=1}^{L}\theta_j) - K\sum_{j=1}^{L}\log\Gamma(\theta_j) \\ &+ \sum_{j=1}^{L}\left[(\theta_j - 1)\sum_{i=1}^{K}\log\, a_j(i)\right] \\ &+ \sum_{j=1}^{L}\sum_{k=j}^{L}\left[K\log\vartheta_{j,k} - \frac{\vartheta_{j,k}^2}{\pi}\sum_{i=1}^{K}(a_{j,k}(i))^2\right], \end{aligned} \tag{12}$$

and η is the regularization parameter whose value depends on the noise variance σ_n^2. In practice, since σ_n^2 is unknown, the value of η is selected manually.

3.4. Optimization Algorithm

The cost function J, defined in (11), should be minimized using an optimization algorithm. If the prior parameters θ_j and $\vartheta_{j,k}$ are known, we can easily adapt the LQ-NMF projected gradient method described in [6] to this cost function just by taking into account the regularization term R, defined by (12), when computing the gradient with respect to the entries of the mixing matrix $\mathbf{A}$, i.e. the parameters $a_j(i)$ and $a_{j,k}(i)$. In practice, however, the prior parameters are not exactly known, although we may have an idea about their possible variation domains[1]. In this case, these parameters may also be estimated during an iterative projected gradient algorithm. The derivatives of J with respect to the different unknowns are listed below. In these formulas, ψ represents the Digamma function, i.e. the logarithmic derivative of the Gamma function. For computing the derivatives of $\frac{1}{2}||\mathbf{X} - \mathbf{AS}||_F^2$ with respect to the mixing parameters and the sources, we used the results provided in Eq. (4) and (8) of [6].

$$\frac{\partial J}{\partial \theta_\ell} = -\eta[K\psi(\sum_{j=1}^{L}\theta_j) - K\psi(\theta_\ell) + \sum_{i=1}^{K}\log a_\ell(i)] \tag{13}$$

$$\frac{\partial J}{\partial \vartheta_{m,n}} = -\eta[\frac{K}{\vartheta_{m,n}} - \frac{2\vartheta_{m,n}}{\pi}\sum_{i=1}^{K}(a_{m,n}(i))^2] \tag{14}$$

$$\frac{\partial J}{\partial a_m(\ell)} = [(AS - X)S^T]_{m\ell} - \eta\frac{\theta_m - 1}{a_m(\ell)} \tag{15}$$

[1]For example, if we know that the observed scene is composed of highly-mixed pixels, we should choose large values for θ_j. Moreover, since $a_{j,k}(i) \in [0, 0.5]$, $\vartheta_{j,k}$ should take large values (corresponding to small variances).

$$\frac{\partial J}{\partial a_{m,n}(\ell)} = [(AS - X)S^T]_{(mn)\ell} + \eta\frac{2\vartheta_{m,n}^2 a_{m,n}(\ell)}{\pi} \tag{16}$$

$$\begin{aligned} \frac{\partial J}{s_{pn}} &= [A^T(AS - X)]_{pn} \\ &+ \sum_{j=1, j\neq p}^{L} s_{jn}[A^T(AS - X)]_{(jp)n} \\ &+ 2s_{pn}[A^T(AS - X)]_{(pp)n}, \end{aligned} \tag{17}$$

where e.g. (jp) is the index of the column of matrix $\mathbf{A}$ corresponding to coefficients $a_{j,p}(n)$.
In each iteration of the projected gradient algorithm, the following rules are used to update the mixing parameters, source samples and PDF parameters:

$$a_j(i) \leftarrow [a_j(i) - \mu_a\frac{\partial J}{\partial a_j(i)}]_P \tag{18}$$

$$a_{j,k}(i) \leftarrow [a_{j,k}(i) - \mu_a\frac{\partial J}{\partial a_{j,k}(i)}]_P \tag{19}$$

$$s_{ji} \leftarrow [s_{ji} - \mu_s\frac{\partial J}{\partial s_{ji}}]_P \tag{20}$$

$$\theta_j \leftarrow [\theta_j - \mu_\theta\frac{\partial J}{\partial \theta_j}]_P \tag{21}$$

$$\vartheta_{j,k} \leftarrow [\vartheta_{j,k} - \mu_\vartheta\frac{\partial J}{\partial \vartheta_{j,k}}]_P, \tag{22}$$

where μ_a, μ_s, μ_θ, μ_ϑ are small fixed positive learning rates, and $[u]_P$ corresponds to the projection of u on the interval P. In fact, if after the gradient update the estimated value is outside P, it is replaced by the value of the nearest bound of P [11]. These intervals are $[0, 1]$ for $a_j(i)$, $[0, 0.5]$ for $a_{j,k}(i)$, $[0, \infty)$ for s_{ji}. Concerning the parameters θ_j and $\vartheta_{j,k}$, if some information about their variation domains is available, it can be used to choose the projection intervals, otherwise $[0, \infty)$ is used.
At each iteration of the algorithm, (20) is used to update the first rows of matrix $\mathbf{S}$ (corresponding to the actual sources), then the last rows of $\mathbf{S}$ are set to the products of the first rows (see [6] for details). To ensure that the sum of the abundance fractions is equal to 1, we normalize the estimated values by dividing them by their sum:

$$a_j(i) = \frac{a_j(i)}{\sum_{k=1}^{L} a_k(i)}. \tag{23}$$

4. SIMULATION RESULTS

The simulations presented in this section aim at evaluating the gain of performance of our proposed method with respect to the initial LQ-NMF method (*Grd LQ* algorithm presented in

[6]) which does not exploit the prior information. The simulations were done for $L = 2$ or $L = 3$ pure materials using a special case of the LQ model (1), called the *bilinear* model, where the coefficients of the squared terms, $a_{j,j}(i)$, are set to zero. In order to evaluate the performance of the algorithms, we compute the Signal-to-Interference Ratio (SIR) for each source and each abundance fraction, defined by

$$SIR_{s_i} = 10log_{10}\frac{\sum_{n=1}^{N} s_{in}^2}{\sum_{n=1}^{N}(s_{in} - \hat{s}_{in})^2} \quad (24)$$

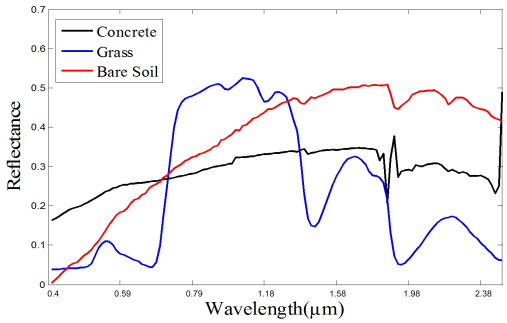
$$SIR_{a_j} = 10log_{10}\frac{\sum_{i=1}^{K} a_j(i)^2}{\sum_{i=1}^{K}(a_j(i) - \hat{a}_j(i))^2} \quad (25)$$

and the $SIR_{a_{j,k}}$ as in (25) for each quadratic coefficient. In these equations, "ˆ" refers to the estimated values.

4.1. Data

We considered the following two cases:
Case 1: The observations are artificial mixtures of two or three synthetic spectra, with $N = 126$ values uniformly distributed over [0,1].
Case 2: The observations are artificial mixtures of two or three real spectra encountered in urban environments, containing 126 wavelengths from 0.4 μm to 2.5 μm (see Fig.1).[2] In both cases, we simulated 100-pixel images. The abundance fractions were generated using Dirichlet distributions with $\boldsymbol{\theta} = [70, 70]$ in the case of two sources and $\boldsymbol{\theta} = [60, 60, 60]$ in the case of three sources. These values correspond to a highly-mixed scene without pure pixels. Quadratic coefficients were generated according to a half-normal distribution with $\vartheta_{j,k} = 8.35$ to obtain values of $a_{j,k}$ in [0,0.5].

Fig. 1. Spectra of used materials



4.2. Results

We used the initial *Grd LQ* algorithm and our proposed method for separating the mixed data. Since the performance of the methods depends on initialization, we performed 100 Monte Carlo simulations for both algorithms. In each simulation, the estimated parameters were initialized as follows: the abundance fractions and sources with random values in [0,1], the quadratic coefficients with random values in [0,0.5], θ_j with large values between 50 and 80, $\vartheta_{j,k}$ with a large value satisfying the constraint $a_{j,k} \in [0, 0.5]$. The gradient learning rates were fixed to 0.0005 for μ_a and μ_s, and 0.01 for μ_θ and μ_ϑ. In our experiments the best results were obtained when the value of the regularization parameter η was between 0.0001 and 0.001.
The mean and standard deviation of SIRs over all Monte Carlo simulations and all sources or all mixing coefficients are shown in Table 1. As can be seen, the proposed algorithm leads to much better results as compared with the original *Grd LQ* algorithm. These results demonstrate the relevance of taking into account the prior information.

Table 1. Results for Case 1: artificial sources, and Case 2: real sources, with $L = 2$ and $L = 3$ sources.

SIR (dB) \ Algorithm		**Grd LQ**	**MAP**
Case 1 (L=2)	Mean/Std SIR_a	11.65/5.74	30.80/2.31
	Mean/Std SIR_s	12.03/4.42	29.58/3.79
Case 2 (L=2)	Mean/Std SIR_a	6.04/1.03	20.89/2.95
	Mean/Std SIR_s	14.73/0.36	20.85/3.12
Case 1 (L=3)	Mean/Std SIR_a	5.38/1.45	28.78/1.07
	Mean/Std SIR_s	11.54/1.70	26.81/2.76
Case 2 (L=3)	Mean/Std SIR_a	2.95/0.89	15.34/2.38
	Mean/Std SIR_s	11.73/0.93	15.37/1.07

5. CONCLUSION

In this paper, we proposed a new method to unmix urban hyperspectral images. For this purpose, we modified a new NMF method adapted to LQ mixtures by using available prior information about the mixing parameters to better estimate them. We defined a regularized criterion derived from the MAP principle, then proposed a projected gradient algorithm to estimate the different unknown parameters of the problem, i.e. the abundance fractions and quadratic coefficients, endmembers, and PDF parameters. The simulation results showed that the proposed algorithm leads to much better results than the initial non-regularized algorithm.
As future work, we will apply our method to real data. In this case, the observed scene is first split in many supposedly stationary sub-images so that the PDF parameters can be supposed to be constant in each sub-image, then our unmixing method is applied to each sub-image. We are also working on other optimization algorithms to get rid of the fixed adaptation step in the gradient algorithm.

[2]These spectra are from the MEMOIRES data base (http://www.onera.fr/dota/memoires).

6. REFERENCES

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

[2] 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 J. of Sel. Topics in Appl. Earth Observ. Remote Sens.,* vol. 5, no. 2, pp. 354-379, April 2012.

[3] N. Dobigeon, J. Y. Tourneret, C. Richard, J. C. M. Bermudez, S. McLaughin, and A. O. Hero, "Nonlinear Unmixing of Hyperspectral Images: models and algorithms," *IEEE Signal processing magazine,* vol. 31, Issue 1, pp. 82-94, January 2014.

[4] I. Meganem, P. Déliot, X. Briottet, Y. Deville, S. Hosseini, "Linear-Quadratic Mixing Model for Reflectances in Urban Environments," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 52, no.1, January 2014.

[5] R. Heylen, M. parente and P. Gader, "A Review of Nonlinear Hyperspectral Unmixing Methods," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 7, Issue 6, pp. 1844-1866, May 2014.

[6] I. Meganem, Y. Deville, S. Hosseini, P. Déliot, X. Briottet, "Linear-Quadratic Blind Source Separation Using NMF to Unmix Urban Hyperspectral Images," *IEEE Transactions on Signal Processing,* vol. 62, no. 7, pp.1822-1833, April 2014.

[7] D.D. Lee and H.S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Letters to Nature,* vol. 401, pp. 788-791, October 1999.

[8] 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," Hoboken, NJ, USA: Wiley, 2009.

[9] J. M. P. Nascimento, and J. M. Bioucas-Dias, "Dependent Component Analysis: A Hyperspectral Unmixing Algorithm," *3rd Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA 2007),* Girona, Spain, June 6-8, 2007.

[10] J. M. P. Nascimento, and J.M. Bioucas-Dias, "Hyperspectral Unmixing Based on Mixtures of Dirichlet Components," *IEEE Transactions On Geoscience And Remote Sensing,* vol. 50, no. 3, pp. 863-878, March 2012.

[11] C.-J. Lin., "Projected gradient methods for non-negative matrix factorization," *Neural Computation,* vol. 19, pp. 2756-2779, 2007.