Robust nonnegative matrix factorization for nonlinear unmixing of hyperspectral images

Nicolas Dobigeon, Cédric Févotte

ROBUST NONNEGATIVE MATRIX FACTORIZATION FOR NONLINEAR UNMIXING OF HYPERSPECTRAL IMAGES

Nicolas Dobigeon

University of Toulouse
CNRS IRIT/INP-ENSEEIHT
Toulouse, France

Cédric Févotte

Laboratoire Lagrange
CNRS, OCA & University of Nice
Parc Valrose, Nice, France

ABSTRACT

This paper introduces a robust linear model to describe hyperspectral data arising from the mixture of several pure spectral signatures. This new model not only generalizes the commonly used linear mixing model but also allows for possible nonlinear effects to be handled, relying on mild assumptions regarding these nonlinearities. Based on this model, a nonlinear unmixing procedure is proposed. The standard nonnegativity and sum-to-one constraints inherent to spectral unmixing are coupled with a group-sparse constraint imposed on the nonlinearity component. The resulting objective function is minimized using a multiplicative algorithm. Simulation results obtained on synthetic and real data show that the proposed strategy competes with state-of-the-art linear and nonlinear unmixing methods.

***Index Terms*—** Hyperspectral imagery, nonlinear unmixing, robust nonnegative matrix factorization, group-sparsity.

1. INTRODUCTION

Spectral unmixing (SU) is an issue of prime interest when analyzing hyperspectral data. SU consists of decomposing P multi-band observations $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_P]^T$ into a collection of K individual spectra $\mathbf{M} = [\mathbf{m}_1, \ldots, \mathbf{m}_K]^T$, called *endmembers*, and estimating their relative proportions (or *abundances*) $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_P]^T$ in each observation [1]. Most of the hyperspectral unmixing algorithms proposed in the signal & image processing and geoscience literatures rely on the commonly admitted linear mixing model (LMM), $\mathbf{Y} \approx \mathbf{MA}$. Indeed, LMM provides a good approximation of the physical process underlying the observations and has resulted in interesting and comprehensive results for numerous applications [2]. However, for several specific applications, LMM may be inaccurate and other nonlinear models need to be advocated. For instance, in remotely sensed images composed of vegetation (e.g., trees), interactions of photons with multiple components of the scene lead to nonlinear effects that can be taken into account using bilinear models. As explained in [3], several bilinear models have been proposed [4–6], and they mainly differ by the constraints imposed on the nonlinearity term. Additionally, to approximate a large range of second-order nonlinearities, Altmann *et al.* [7] introduce a polynomial post-nonlinear model that is able to describe most of the nonlinear effects occurring in the observed scene. A common feature of these models is that they all consist in including a supplementary additive term to the standard LMM, accounting for the nonlinearities. One major drawback of these models, however, is that they require to choose a specific form of nonlinearity, and this can be limiting in practice.

In this paper, a new so-called robust LMM (rLMM) is proposed. Similarly to the nonlinear models detailed above, it is built on the standard LMM and includes a supplementary additive term that accounts for nonlinear effects. However, it does not require to specify an analytical form of the nonlinearity. Instead, nonlinearities are merely treated as *outliers*. Our motivation is that the LMM is a valid model in the majority of pixels and that only a *sparse* number of pixels are affected by nonlinearities. As such, our contribution consists in decomposing the multi-band observations as $\mathbf{Y} \approx \mathbf{MA} + \mathbf{R}$, where $\mathbf{R}$ is a sparse (and nonnegative) residual term accounting for outliers (i.e., nonlinear effects), with sparsity imposed at the group-level (a column of $\mathbf{R}$ is either entirely zero or not). The proposed decomposition relates to *robust nonnegative matrix factorization* (rNMF) as will be explained in more details in the following.

The article is organized as follows. The rLMM is introduced in more details in Section 2. Section 3 describes a multiplicative algorithm for rLMM estimation. Results obtained on synthetic and real data are reported in Section 4. Section 5 concludes.

2. ROBUST LINEAR MIXING MODEL

The proposed rLMM is given by

$$\mathbf{y}_p = \sum_{k=1}^{K} a_{kp}\mathbf{m}_k + \mathbf{r}_p + \mathbf{n}_p, \tag{1}$$

where $\mathbf{y}_p = [y_{1p}, \ldots, y_{Lp}]^T$ denotes the pth pixel spectrum observed in L spectral bands, $\mathbf{m}_k = [m_{1k}, \ldots, m_{Lk}]^T$ denotes the kth endmember spectrum, $\mathbf{a}_p = [a_{1p}, \ldots, a_{Kp}]^T$ denotes the abundances representing the pth pixel, $\mathbf{r}_p = [r_{1p}, \ldots, r_{Lp}]^T$ denotes the outlier term (accounting for nonlinearities) and $\mathbf{n}_p$ denotes residual noise. The matrix formulation of Eq. (1) is given by

$$\mathbf{Y} = \mathbf{MA} + \mathbf{R} + \mathbf{N}. \tag{2}$$

The following extra assumptions are made. The data $\mathbf{y}_p$ is nonnegative by nature, and we take $\mathbf{m}_k$ and $\mathbf{a}_p$ to be nonnegative as well. We also take the abundance coefficients to sum to one, i.e., $\sum_k a_{kp} = 1$, as commonly assumed in most hyperspectral data models. The residual noise $\mathbf{n}_p$ is assumed zero-mean white Gaussian.

Part of this work has been funded by the ESTOMAT PEPS Project supported by CNRS and by the Hypanema ANR Project n°ANR-12-BS03-003.

In this work, we assume the nonlinear component $\mathbf{r}_p$ to be nonnegative, like in the bilinear models of [4–6] and the polynomial model with constructive interferences of [7]. As discussed in the introduction, we expect $\mathbf{r}_p$ to be often zero, i.e., pixels to follow the standard LMM in general. For pixels where the LMM assumption fails, nonlinearities will become "active" and $\mathbf{r}_p$ will become nonzero. This amounts to say that the energy vector

$$\mathbf{e} = \left[\|\mathbf{r}_1\|_2, \ldots, \|\mathbf{r}_P\|_2\right]^T \tag{3}$$

where $\|\mathbf{x}\|_2 = \sqrt{\sum_k x_k^2}$, is sparse. Sparsity can routinely be enforced by ℓ_1 regularisation. As such, our objective is to solve the minimisation problem defined by

$$\min_{\mathbf{M},\mathbf{A},\mathbf{R}} \mathcal{J}(\mathbf{M},\mathbf{A},\mathbf{R}) = \|\mathbf{Y} - \mathbf{M}\mathbf{A} - \mathbf{R}\|_2^2 + \lambda \|\mathbf{R}\|_{2,1} \quad \text{s.t. } \mathbf{M} \geq 0,\ \mathbf{A} \geq 0,\ \mathbf{R} \geq 0 \text{ and } \|\mathbf{a}_p\|_1 = 1, \tag{4}$$

where λ is a nonnegative penalty weight, $\mathbf{A} \geq 0$ denotes nonnegativity of the coefficients of $\mathbf{A}$, $\|\mathbf{x}\|_1 = \sum_k x_k$ and $\|\cdot\|_{2,1}$ is the so-called $\ell_{2,1}$ norm defined by

$$\|\mathbf{R}\|_{2,1} = \|\mathbf{e}\|_1 = \sum_{p=1}^{P} \|\mathbf{r}_p\|_2 . \tag{5}$$

Eq. (4) defines a robust NMF problem. Robust NMF is a nonnegative variant of robust PCA [8] which has appeared in different forms in the literature. In [9], the outlier term $\mathbf{R}$ is nonnegative and penalized by the ℓ_1 norm. In [10] and [11], $\mathbf{R}$ is real-valued and penalized by ℓ_1 and $\ell_{1,2}$ norms, respectively. In [12], there is no residual term $\mathbf{N}$ and the $\ell_{2,1}$ norm $\|\mathbf{Y} - \mathbf{M}\mathbf{A}\|_{2,1}$ is minimized. To the best of our knowledge, the formulation of robust NMF described by Eq. (4), where $\mathbf{R}$ is nonnegative and penalized by the $\ell_{2,1}$ norm (and where the abundances sum to 1), is entirely novel.

3. ALGORITHM

We present a multiplicative algorithm that returns stationary points of Eq. (4). Our algorithm is based on a heuristic commonly used in NMF, see, e.g., [13], and as follows. Let θ be a scalar coefficient of $\mathbf{M}$, $\mathbf{A}$ or $\mathbf{R}$. As it appears, the derivative $\nabla_\theta \mathcal{J}$ of the objective function with respect to θ can always be expressed as the difference of two nonnegative functions such that $\nabla_\theta \mathcal{J} = \nabla_\theta^+ \mathcal{J} - \nabla_\theta^- \mathcal{J}$. The multiplicative algorithm simply writes

$$\theta \leftarrow \theta . \frac{\nabla_\theta^- \mathcal{J}}{\nabla_\theta^+ \mathcal{J}}. \tag{6}$$

It ensures nonnegativity of the parameter updates, provided initialization with a nonnegative value. It produces a descent algorithm in the sense that θ is updated towards left (resp., right) when the gradient is positive (resp., negative). Though based on a simple heuristic, the update Eq. (6) is often an exact majorization-minimization algorithm in disguise, which guarantees the decrease of the objective function at each iteration [14]. We will not give such proofs here and simply apply the heuristic (6). In practice, the resulting algorithm was indeed observed to decrease the objective function at each iteration. The abundance sum-to-one constraint is implemented with a change a variable, following [15].

Using the notation $\hat{\mathbf{Y}} = \mathbf{M}\mathbf{A} + \mathbf{R}$, with coefficients $\hat{y}_{lp}$, the resulting iterative updates are

$$a_{kp} \leftarrow a_{kp} \frac{\sum_l (\hat{y}_{lp} - r_{lp}) \hat{y}_{lp} + m_{lk} y_{lp}}{\sum_l (\hat{y}_{lp} - r_{lp}) y_{lp} + m_{lk} \hat{y}_{lp}} \tag{7}$$

$$a_{kp} \leftarrow \frac{a_{kp}}{\|\mathbf{a}_p\|_1} \tag{8}$$

$$r_{lp} \leftarrow r_{lp} \frac{y_{lp}}{\hat{y}_{lp} + \frac{\lambda}{2} \frac{r_{lp}}{\|\mathbf{r}_p\|_2}} \tag{9}$$

$$m_{lk} \leftarrow m_{lk} \frac{\sum_p a_{kp}\, y_{lp}}{\sum_p a_{kp}\, \hat{y}_{lp}} \tag{10}$$

4. SIMULATION RESULTS

4.1. Synthetic data

First, to evaluate the relevance of the proposed rLMM and the accuracy of the corresponding robust NMF algorithm, some simulations have been conducted on synthetic data. Four 64×64-pixel images composed of $K = 3$ pure spectral components have been generated according to four different linear and nonlinear models. The endmember spectra have been extracted from the spectral library provided with the ENVI software [16] and correspond to micaceous loam, green grass and bare red brick. The first image, denoted as $\mathcal{I}_{\mathrm{LMM}}$, is composed of pixels following the standard linear mixing model

$$\mathbf{y}_p = \sum_{k=1}^{K} a_{kp} \mathbf{m}_k .$$

The three other images are supposed to be mainly composed of pixels following LMM. However, one fourth of each image (i.e., 1024 pixels) consists of pixels coming from nonlinear mixtures of the endmembers. More precisely, in images denoted $\mathcal{I}_{\mathrm{FM}}$ and $\mathcal{I}_{\mathrm{GBM}}$, some pixels are subjected to bilinear interactions between components, according to the Fan bilinear model (FM) [5]

$$\mathbf{y}_p = \sum_{k=1}^{K} a_{kp} \mathbf{m}_k + \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} a_{ip} a_{jp} \mathbf{m}_i \odot \mathbf{m}_j$$

or the generalized bilinear model (GBM) [7]

$$\mathbf{y}_p = \sum_{k=1}^{K} a_{kp} \mathbf{m}_k + \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \gamma_{ijp} a_{ip} a_{jp} \mathbf{m}_i \odot \mathbf{m}_j$$

where $\mathbf{m}_i \odot \mathbf{m}_j$ stands for the termwise (Hadamard) product and γ_{ijp} adjusts the bilinear interaction between the ith and jth endmembers in the pth pixel. One fourth of the last image $\mathcal{I}_{\mathrm{PNLMM}}$ pixels are generated using the polynomial nonlinear mixing model (PNLMM) introduced in [7]

$$\mathbf{y}_p = \mathbf{M}\mathbf{a}_p + b\,(\mathbf{M}\mathbf{a}_p) \odot (\mathbf{M}\mathbf{a}_p) .$$

The abundance coefficients a_{kp} for each model are randomly generated on their admissible set defined by the the nonnegativity and additivity constraints appearing in (4), with the specific scenario that a cutoff has been imposed to remove pure pixels from the observations. The interaction coefficients γ_{ijp} in the GBM have been uniformly drawn over the set $(0, 1)$ and the nonlinear coefficient b in the PNLMM has been arbitrary fixed to 0.3. The four images have

been unmixed using standard algorithms specially designed for the considered models. First, vertex component analysis (VCA) [17] has been used as an endmember extraction algorithm to recover the spectral signatures of the pure components. For comparison, endmembers were also extracted from images using the nonlinear endmember extraction algorithm proposed in [18], denoted as Heylen's algorithm in what follows. Then, in an inversion step, the mixing coefficients have been estimated by algorithms dedicated to the LMM, FM, GBM and PNLMM, respectively. More precisely, we consider FCLS as a linear inversion algorithm [19], the FM-based unmixing technique proposed in [5], the gradient descent algorithm to estimate the GBM parameters [20] and the subgradient-based optimization scheme detailed in [7] dedicated to PNLMM. The performance of the unmixing algorithms are evaluated in terms of global mean square error related to the the endmember spectra

$$\text{GMSE}^2\left(\mathbf{M}\right) = \frac{1}{LK}\sum_{k=1}^{K}\left\|\mathbf{m}_k - \hat{\mathbf{m}}_k\right\|^2$$

and abundance matrix

$$\text{GMSE}^2\left(\mathbf{A}\right) = \frac{1}{KP}\sum_{p=1}^{P}\left\|\mathbf{a}_p - \hat{\mathbf{a}}_p\right\|^2$$

The results in Table 1 show that the proposed method clearly outperforms both VCA and Heylen's algorithm to recover the endmember. In particular, these results demonstrate the ability of the rLMM-based unmixing technique to mitigate several kinds of nonlinear effects while preserving good estimation performance when analyzing only linear mixtures.

	VCA	Heylen's algo.	rLMM
$\mathcal{I}_{\text{LMM}}$	2.13	25.4	**1.92**
$\mathcal{I}_{\text{FM}}$	1.94	14.2	**1.83**
$\mathcal{I}_{\text{GBM}}$	2.10	26.0	**1.78**
$\mathcal{I}_{\text{PNLMM}}$	1.88	28.6	**1.69**

Table 1. Endmember estimation performance in term of $\text{GMSE}^2\left(\mathbf{M}\right)$ ($\times 10^{-3}$).

The performance in term of $\text{GMSE}^2\left(\mathbf{A}\right)$ is reported[1] in Table 2. Similarly, these results demonstrate the flexibility of the rLMM to model observations coming from various scenarios.

	LMM	FM	GBM	PNLMM	rLMM
$\mathcal{I}_{\text{LMM}}$	1.85	9.03	1.84	1.85	**1.69**
$\mathcal{I}_{\text{FM}}$	5.07	13.9	4.60	4.89	**4.56**
$\mathcal{I}_{\text{GBM}}$	4.93	12.5	4.66	4.65	**4.43**
$\mathcal{I}_{\text{PNLMM}}$	1.85	11.6	1.82	1.89	**1.66**

Table 2. Abundance estimation performance in term of $\text{GMSE}^2\left(\mathbf{A}\right)$ ($\times 10^{-6}$).

[1]Note that, for brevity, the inversion methods used for LMM, FM, GBM and PNLMM unmixing are only coupled with the results provided by VCA since the endmember spectra estimated by the Heylen's algorithm were not relevant.

4.2. Real data

As an illustration, the proposed rLMM-unmixing technique has been applied on the real Moffett Field dataset previously used for instance in [6, 21, 22]. The area of interest is a lake shore thus mainly composed of water, vegetation and soil. The endmember spectra and the corresponding abundance maps recovered while using the rLMM are depicted in Fig. 1 (top) and Fig. 1 (bottom) where black (resp. white) pixels correspond to absence (resp. presence) of the associated endmembers. All these results are in good agreement with previous results shown in [6, 21, 22].

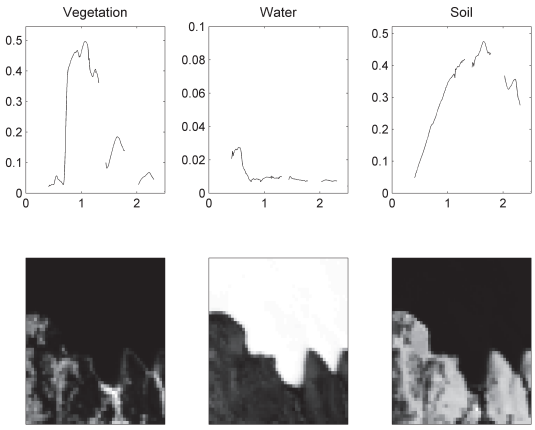

Fig. 1. Top: endmembers estimated by the proposed rNMF-based unmixing algorithm. Bottom: corresponding estimated abundance maps.

However, in addition to this standard description of the data by linearly mixed endmembers, the proposed model also provides information regarding the pixels that can not be explained with the standard LMM. Figure 2 shows the energy $\mathbf{e} = \left[\left\|\mathbf{r}_1\right\|_2, \ldots, \left\|\mathbf{r}_P\right\|\right]^T$ of the residual component (previously introduced in (3)) estimated by the algorithm detailed in Section 3 (a white pixel corresponds to a residual of high energy). This map demonstrates that most of the pixels of this scene can be accurately described using the LMM. However, some few pixels, mainly located in the lake shore, appear at outliers. These pixels probably correspond to areas where some interactions between several endmembers occur (e.g., water/vegetation, water/soil).

5. CONCLUSION

This paper presented a new mixing model to describe hyperspectral data. This model, denoted as rLMM, extends the standard LMM by including a residual term that can capture so-called nonlinear effects. These nonlinear effects are treated as additive and sparsely active outliers. The resulting unmixing problem was formulated as a new form of robust NMF problem, for which we developed a simple and effective multiplicative algorithm. Simulations conducted on synthetic and real data illustrated the effectiveness of rLMM, which outperformed many unmixing methods designed for various linear and nonlinear models.

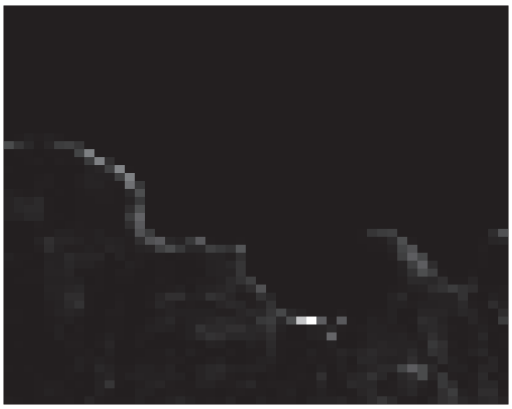

Fig. 2. Energy of the nonlinear components estimated by the proposed algorithm.

6. ACKNOWLEDGEMENTS

C. Févotte acknowledges Pablo Sprechmann, Alex Bronstein & Vincent Y. F. Tan for inspiring discussions about robust NMF.

7. 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 J. Sel. Topics Appl. Earth Observations and Remote Sens.*, vol. 5, no. 2, pp. 354–379, April 2012.

[2] C.-I Chang, *Hyperspectral Data Exploitation: theory and applications*. Hoboken, NJ: John Wiley & Sons, 2007.

[3] Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Bilinear models for nonlinear unmixing of hyperspectral images," in *Proc. IEEE GRSS Workshop Hyperspectral Image SIgnal Process.: Evolution in Remote Sens. (WHISPERS)*, Lisbon, Portugal, June 2011, pp. 1–4.

[4] J. M. Bioucas-Dias and J. M. P. Nascimento, "Nonlinear mixture model for hyperspectral unmixing," in *Proc. SPIE Image and Signal Processing for Remote Sensing XV*, L. Bruzzone, C. Notarnicola, and F. Posa, Eds., vol. 7477, no. 1. SPIE, 2009, p. 74770I.

[5] W. Fan, B. Hu, J. Miller, and M. Li, "Comparative study between a new nonlinear model and common linear model for analysing laboratory simulated-forest hyperspectral data," *Int. J. Remote Sens.*, vol. 30, no. 11, pp. 2951–2962, June 2009.

[6] A. Halimi, Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Nonlinear unmixing of hyperspectral images using a generalized bilinear model," *IEEE Trans. Geosci. and Remote Sensing*, vol. 49, no. 11, pp. 4153–4162, Nov. 2011.

[7] Y. Altmann, A. Halimi, N. Dobigeon, and J.-Y. Tourneret, "Supervised nonlinear spectral unmixing using a post-nonlinear mixing model for hyperspectral imagery," *IEEE Trans. Image Process.*, vol. 21, no. 6, pp. 3017–3025, June 2012.

[8] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of ACM*, vol. 58, no. 1, pp. 1–37, 2009.

[9] P. Sprechmann, A. Bronstein, and G. Sapiro, "Real-time online singing voice separation from monaural recordings using robust low-rank modeling," in *Proc. International Society for Music Information Retrieval Conference (ISMIR)*, Porto, Portugal, Oct. 2012.

[10] L. Zhang, Z. Chen, M. Zheng, and X. He, "Robust nonnegative matrix factorization," *Front. Electr. Electron. Eng. China*, vol. 6, no. 2, pp. 192–200, 2011.

[11] B. Shen, L. Si, R. Ji, and B. Liu, "Robust nonnegative matrix factorization via ℓ_1 norm regularization," *ArXiv preprint arXiv:1204.2311*, 2012.

[12] D. Kong, C. Ding, and H. Huang, "Robust nonnegative matrix factorization using ℓ_{21}-norm," in *Proc. 20th ACM Int. Conf. Information and Knowledge Management*, 2011, pp. 673–682.

[13] T. Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 3, pp. 1066–1074, Mar. 2007.

[14] C. Févotte and J. Idier, "Algorithms for nonnegative matrix factorization with the beta-divergence," *Neural Computation*, vol. 23, no. 9, pp. 2421–2456, Sep. 2011. [Online]. Available: http://arxiv.org/pdf/1010.1763v3

[15] J. Eggert and E. Körner, "Sparse coding and NMF," in *Proc. IEEE International Joint Conference on Neural Networks*, 2004, pp. 2529–2533.

[16] RSI (Research Systems Inc.), *ENVI User's guide Version 4.0*, Boulder, CO 80301 USA, Sept. 2003.

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

[18] R. Heylen, D. Burazerovic, and P. Scheunders, "Non-linear spectral unmixing by geodesic simplex volume maximization," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 534–542, June 2011.

[19] D. C. Heinz and C. -I Chang, "Fully constrained least-squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. and Remote Sensing*, vol. 29, no. 3, pp. 529–545, March 2001.

[20] A. Halimi, Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Unmixing hyperspectral images using the generalized bilinear model," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, Vancouver, Canada, July 2011, pp. 1886–1889.

[21] N. Dobigeon, J.-Y. Tourneret, and C.-I Chang, "Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2684–2695, July 2008.

[22] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A. O. Hero, "Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4355–4368, Nov. 2009.