UNMIXING HYPERSPECTRAL IMAGES USING
A NORMAL COMPOSITIONAL MODEL AND MCMC METHODS

O. Eches[(1)], *N. Dobigeon*[(1,2)], *C. Mailhes*[(1)] *and J.-Y. Tourneret*[(1)]

[(1)] University of Toulouse, IRIT/INP-ENSEEIHT, 2 rue Camichel, 31071 Toulouse cedex 7, France
[(2)] University of Michigan, Department of EECS, Ann Arbor, MI 48109-2122, USA

{olivier.eches,jean-yves.tourneret,corinne.mailhes}@enseeiht.fr, dobigeon@umich.edu

ABSTRACT

This paper studies a new unmixing algorithm for hyperspectral images. Each pixel of the image is modeled as a linear combination of *endmembers* which are supposed to be random in order to model uncertainties regarding their knowledge. More precisely, endmembers are modeled as Gaussian vectors with known means (resulting from an endmember extraction algorithm such as the famous N-FINDR or VCA algorithm). This paper proposes to estimate the mixture coefficients (referred to as *abundances*) using a Bayesian algorithm. Suitable priors are assigned to the abundances in order to satisfy positivity and additivity constraints whereas a conjugate prior is chosen for the variance. The computational complexity of the resulting Bayesian estimators is alleviated by constructing an hybrid Gibbs algorithm to generate abundance and variance samples distributed according to the posterior distribution of the unknown parameters. The associated hyperparameter is also generated. The performance of the proposed methodology is evaluated thanks to simulation results conducted on synthetic and real images.

Index Terms— Bayesian inference, Monte Carlo methods, spectral unmixing, hyperspectral images, normal compositional model.

1. INTRODUCTION

The spectral unmixing problem has received considerable attention in the signal and image processing literature (see for instance [1] and references therein). Most unmixing procedures for hyperspectral images assume that the image pixels are linear combinations of a given number of pure materials with corresponding fractions referred to as abundances. More precisely, according to the linear mixing model (LMM) presented in [1], the L-spectrum $\boldsymbol{y} = [y_1, \ldots, y_L]^T$ of a mixed pixel is assumed to be a linear combination of R spectra $\boldsymbol{m}_r$ corrupted by additive white Gaussian noise:

$$\boldsymbol{y} = \sum_{r=1}^{R} \boldsymbol{m}_r \alpha_r + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{m}_r = [m_{r,1}, \ldots, m_{r,L}]^T$ denotes the spectrum of the rth material, α_r is the fraction of the rth material in the pixel, R is the number of pure materials (or endmembers) present in all the observed scene and L is the number of available spectral bands for the image. Supervised algorithms assume that the R endmembers spectra $\boldsymbol{m}_r$ are known, e.g. extracted from a spectral library. In practical applications, they can be obtained by an endmember extraction procedure such as the well-known N-finder (N-FINDR) algorithm developed by Winter [2] or the Vertex Component Analysis (VCA) presented

by Nascimento [3]. Due to physical considerations, the abundances satisfy the following positivity and additivity constraints:

$$\begin{cases} \alpha_r \geq 0, \ \forall r = 1, \ldots, R, \\ \sum_{r=1}^{R} \alpha_r = 1. \end{cases} \tag{2}$$

However, this linear model has some limitations when applied to real images [1]. In particular, the ratio between the intra-class variance (within endmember classes) and the inter-class variance (between endmembers) allows one to question the validity of the deterministic spectrum assumption [4]. Moreover, the endmember extraction procedures based on the LMM can be inefficient when the image does not contain enough pure pixels. This problem is illustrated in Fig. 1 which shows 1) the dual-band projections (on the two most discriminant axes resulting from a principal component analysis (PCA)) of $R = 3$ endmembers (vegetation, water and soil) (red stars corresponding to the vertices of the red triangle), 2) the dual-band domain containing all linear combinations of the $R = 3$ endmembers (i.e. the red triangle), and 3) the dual-band simplex estimated by the N-FINDR algorithm using the black pixels. As there is no pixel close to the vertices of the red triangle, the N-FINDR estimates a much smaller simplex (in blue) than the actual one (in red).

A new LMM referred to as normal compositional model (NCM) was recently proposed in [4] to alleviate the problems mentioned above. The NCM assumes that the pixels of the hyperspectral image are linear combinations of Gaussian endmembers with known means (e.g. resulting from the N-FINDR or the VCA algorithms). This model allows more flexibility regarding the observed pixels and the endmembers. In particular, the endmembers are allowed to be further from the observed pixels which is clearly an interesting property for the problem illustrated in Fig. 1. The NCM assumes that the spectrum of a mixed pixel can be written as follows:

$$\boldsymbol{y} = \sum_{r=1}^{R} \boldsymbol{\mathcal{E}}_r \alpha_r, \tag{3}$$

where the $\boldsymbol{\mathcal{E}}_r$ are independent Gaussian vectors with known means, e.g. extracted from a spectral library or estimated by an appropriate method such as the VCA algorithm. Note that there is no additive noise in (3) since the random nature of the endmembers already models some kind of uncertainty regarding the endmembers. This paper assumes that the covariance matrix of each endmember can be written $\sigma^2 \boldsymbol{I}_L$, where $\boldsymbol{I}_L$ is the $L \times L$ identity matrix and σ^2 is the endmember variance in a spectral band. Note that a more sophisticated model with different variances in the spectral bands could be investigated. However, the simplifying assumption of a common variance in all spectral bands has been considered successfully in many studies [5, 6].

This paper studies a new Bayesian unmixing algorithm derived from the NCM. Appropriate prior distributions are chosen for the NCM abundances to satisfy the positivity and additivity constraints, as in [7]. A vague conjugate inverse Gamma distribution is defined for the endmember variance. The hyperparameter associated to this inverse Gamma distribution is assigned a vague hyperprior resulting in a hierarchical Bayesian model. The posterior distribution of the corresponding unknown model parameter vector is too complex to derive the standard minimum mean square error (MMSE) or maximum *a posteriori* (MAP) estimators. The complexity of the posterior can be handled by the expectation maximization (EM) algorithm [4, 8]. However, this algorithm can have "*serious shortcomings including the convergence to a local maximum of the posterior*" [9, p. 259]. These shortcomings can be bypassed by considering Markov Chain Monte Carlo (MCMC) methods which allow one to generate samples distributed according to the posterior of interest (here the joint posterior of the abundances and the endmember variance). This paper generalizes the hybrid Gibbs sampler of [7] to the NCM.

The paper is organized as follows. Section 2 derives the posterior distribution of the unknown parameter vector resulting from the NCM. Section 3 studies the hybrid Gibbs sampling strategy which is used to generate samples distributed according to this posterior. Simulation results conducted on synthetic and real data are presented in Section 4. Conclusions are reported in Section 5.
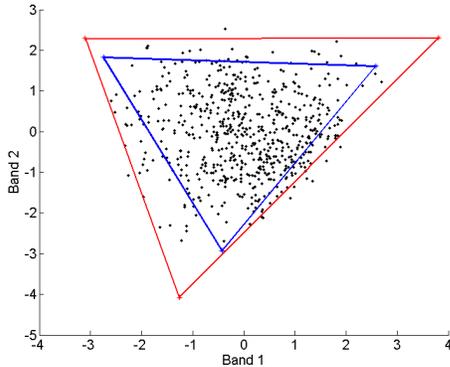

Fig. 1. Scatterplot of dual-band correct (red) and incorrect (blue) results of the N-FINDR algorithm.

2. HIERARCHICAL BAYESIAN MODEL

This section provides more details on the likelihood and the priors inherent to the proposed NCM for the spectral unmixing of hyperspectral images. A particular attention is devoted to defining abundance prior distributions satisfying positivity and additivity constraints.

2.1. Likelihood

The NCM assumes that the endmembers $\mathcal{E}_r$, $r = 1, ..., R$, are independent Gaussian vectors with known means. Moreover, this paper assumes that the endmember components are independent from one band to another, i.e. $\mathcal{E}_r \sim \mathcal{N}(\boldsymbol{m}_r, \sigma^2 \boldsymbol{I}_L)$, where $\boldsymbol{m}_r = [m_{r,1}, \ldots, m_{r,L}]^T$ is the mean vector of $\mathcal{E}_r$, $\sigma^2 \boldsymbol{I}_L$ is its covariance matrix and $\mathcal{N}(.)$ denotes the Gaussian distribution. Using the NCM definition (3) and the independence between the endmembers, the likelihood of $\boldsymbol{y}$ can be written

$$f(\boldsymbol{y} \,|\, \boldsymbol{\alpha}, \sigma^2) = \frac{1}{[2\pi c(\boldsymbol{\alpha})]^{L/2}} \exp\left(-\frac{\|\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\alpha})\|^2}{2c(\boldsymbol{\alpha})}\right), \quad (4)$$

where $\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}^T \boldsymbol{x}}$ is the standard ℓ^2 norm, $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_{R-1}]^T$, and

$$\boldsymbol{\mu}(\boldsymbol{\alpha}) = \sum_{r=1}^R \boldsymbol{m}_r \alpha_r, \quad c(\boldsymbol{\alpha}) = \sigma^2 \sum_{r=1}^R \alpha_r^2. \quad (5)$$

Note that the mean and variance of this Gaussian distribution depend both on the abundance vector $\boldsymbol{\alpha}$ contrary to the classical LMM.

2.2. Parameter priors

2.2.1. Abundance prior

The abundance vector can be rewritten as $\boldsymbol{\alpha}^+ = [\boldsymbol{\alpha}^T, \alpha_R]^T$ where $\alpha_R = 1 - \sum_{r=1}^{R-1} \alpha_r$. Because of the positivity and additivity constraints, the abundance vector $\boldsymbol{\alpha}$ lives in a simplex defined by

$$\mathbb{S} = \left\{ \boldsymbol{\alpha} \,\middle|\, \alpha_r \geq 0, \forall r = 1, \ldots, R-1, \sum_{r=1}^{R-1} \alpha_r \leq 1 \right\}. \quad (6)$$

A uniform distribution on this simplex is chosen as prior for $\boldsymbol{\alpha}$:

$$f(\boldsymbol{\alpha}) \propto \mathbf{1}_{\mathbb{S}}(\boldsymbol{\alpha}), \quad (7)$$

where $\mathbf{1}_{\mathbb{S}}(.)$ is the indicator function defined on the set $\mathbb{S}$. This prior ensures the positivity and additivity constraints are satisfied and reflects the absence of other prior knowledge regarding the abundances. Note that any abundance could be removed from $\boldsymbol{\alpha}^+$ and not only the last one α_R (for symmetry reasons, the sampler proposed in Section 3 removes one abundance from $\boldsymbol{\alpha}^+$ uniformly in $\{1, ..., R\}$. Here, this component is supposed to be α_R to simplify notations).

2.2.2. Endmember variance prior

The prior distribution for the variance σ^2 is a conjugate Inverse Gamma distribution:

$$\sigma^2 | \delta \sim \mathcal{IG}(\nu, \delta), \quad (8)$$

where ν and δ are two adjustable hyperparameters. This paper classically assumes $\nu = 1$ (as in [10]) and estimates δ using a hierarchical Bayesian algorithm. Hierarchical Bayesian algorithms require to define prior distributions for the hyperparameters. This paper assumes that the prior of δ is a non-informative Jeffrey's prior:

$$f(\delta) \propto \frac{1}{\delta} \mathbf{1}_{\mathbb{R}^+}(\delta), \quad (9)$$

where $\propto$ means "proportional to". This prior reflects the lack of knowledge regarding the hyperparameter δ.

2.3. Posterior distribution of θ

The posterior distribution of the unknown parameter vector $\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \sigma^2\}$ can be expressed as:

$$f(\boldsymbol{\theta} \,|\, \boldsymbol{y}) \propto \int f(\boldsymbol{y}|\boldsymbol{\theta}) f(\boldsymbol{\theta} \,|\, \delta) f(\delta) d\delta, \quad (10)$$

where $f(\boldsymbol{y}|\boldsymbol{\theta})$ and $f(\delta)$ have been defined in (4) and (9). Assuming independence between the unknown parameters, the prior distribution of $\boldsymbol{\theta}$ is $f(\boldsymbol{\theta} \,|\, \delta) = f(\boldsymbol{\alpha}) f(\sigma^2 | \nu, \delta)$, yielding:

$$f(\boldsymbol{\theta} \,|\, \boldsymbol{y}) \propto \frac{1}{\sigma^{2R} [c(\boldsymbol{\alpha})]^{L/2}} \exp\left(-\frac{\|\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\alpha})\|^2}{2c(\boldsymbol{\alpha})}\right) \mathbf{1}_{\mathbb{S}}(\boldsymbol{\alpha}). \quad (11)$$

This posterior distribution is too complex to derive the MMSE or MAP estimators of $\boldsymbol{\theta}$. The next section studies a hybrid Gibbs sampler which generates abundances and variances distributed according to the full posterior (11).

3. HYBRID GIBBS SAMPLER

This section studies a hybrid Metropolis-within-Gibbs sampler which generates samples according to the posterior $f(\boldsymbol{\theta}|\boldsymbol{y})$. The sampler iteratively generates $\boldsymbol{\alpha}$ according to $f(\boldsymbol{\alpha}|\boldsymbol{y}, \sigma^2)$, σ^2 according to $f(\sigma^2|\boldsymbol{y}, \boldsymbol{\alpha})$, and δ according to $f(\delta|\sigma^2)$.

3.1. Generation according to $f(\boldsymbol{\alpha}|\boldsymbol{y}, \sigma^2)$

The Bayes' theorem yields:

$$f(\boldsymbol{\alpha}|\boldsymbol{y}, \sigma^2) \propto f(\boldsymbol{y}|\boldsymbol{\theta}) f(\boldsymbol{\alpha}), \qquad (12)$$

which easily leads to:

$$f(\boldsymbol{\alpha}|\boldsymbol{y}, \sigma^2) \propto \frac{1}{[c(\boldsymbol{\alpha})]^{L/2}} \exp\left(-\frac{\|\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\alpha})\|^2}{2c(\boldsymbol{\alpha})}\right) \mathbf{1}_{\mathbb{S}}(\boldsymbol{\alpha}). \quad (13)$$

Note that the conditional distribution of $\boldsymbol{\alpha}$ is defined on the simplex $\mathbb{S}$. As a consequence, the abundance vector $\boldsymbol{\alpha}^+$ satisfies the positivity and sum-to-one constraints. The generation of $\boldsymbol{\alpha}$ according to (13) can be achieved using a Metropolis-within-Gibbs algorithm. We have used the prior distribution (7) as proposal distribution for this algorithm.

3.2. Generation according to $f(\sigma^2|\boldsymbol{y}, \boldsymbol{\alpha}, \delta)$

The conditional distribution of the variance σ^2 can be determined as follows:

$$f(\sigma^2|\boldsymbol{y}, \boldsymbol{\alpha}, \delta) \propto f(\boldsymbol{y}|\boldsymbol{\theta}) f(\sigma^2|\nu, \delta). \qquad (14)$$

Straightforward computations lead to the following result:

$$\sigma^2|\boldsymbol{y}, \boldsymbol{\alpha}, \delta \sim \mathcal{IG}\left(\frac{L}{2} + 1, \frac{\|\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\alpha})\|^2}{2\sum_{r=1}^{R} \alpha_r^2} + \delta\right). \qquad (15)$$

3.3. Generation according to $f(\delta|\sigma^2, \boldsymbol{y}, \boldsymbol{\alpha})$

The conditional distribution of δ is clearly

$$\delta|\sigma^2, \boldsymbol{y}, \boldsymbol{\alpha} \sim \mathcal{G}\left(R + 1, 1 + \frac{R}{\sigma^2}\right), \qquad (16)$$

where $\mathcal{G}(a, b)$ is the Gamma distribution with parameters a and b.

4. SIMULATION RESULTS

The performance of the proposed unmixing algorithm is illustrated with simulation results associated to synthetic and real data. All simulations were obtained with $L = 276$ spectral bands ranging from wavelength $0.4~\mu m$ to $2.5~\mu m$ (from the visible to the near infrared). Classically, the first samples generated by the Gibbs sampler (belonging to the so-called burn in period) are not considered for parameter estimation. The simulation depicted in this section have been obtained with $N_{MC} = 25000$ iterations including $N_{bi} = 5000$ burn-in iterations.

4.1. Synthetic Data

The first experiments consider one synthetic mixture with $R = 2$ endmembers. The means of these endmembers $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$ have been extracted from the spectral libraries distributed with the ENVI package. These spectra correspond to construction concrete and green grass and are depicted in Fig. 2. The endmember variance is $\sigma^2 = 0.01$. The linear mixture considered in this section is defined by $\boldsymbol{\alpha}^+ = [0.3, 0.7]^T$. Figure 3 shows the histograms of the abundances generated by the proposed Gibbs sampler. These histograms

are in good agreement with the actual values of the abundances. Figure 4 shows the estimated posterior distribution of σ^2 which is also in good agreement with the actual endmember variance $\sigma^2 = 0.01$.
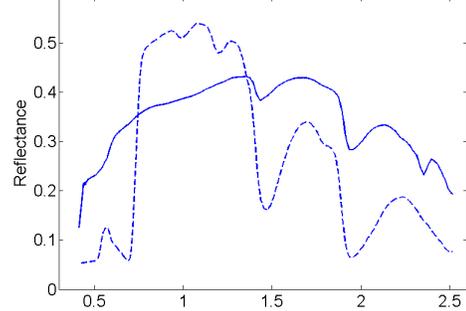

Fig. 2. Endmember spectra: construction concrete(solid line), green grass (dashed line).
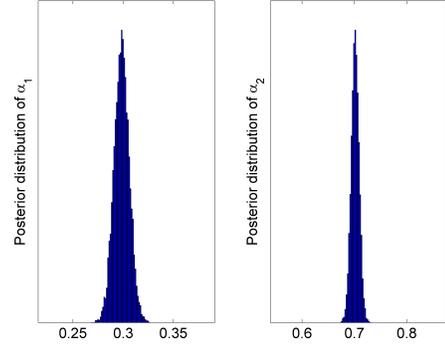

Fig. 3. Estimated posterior distributions of the abundances $[\alpha_1, \ \alpha_2]^T$.
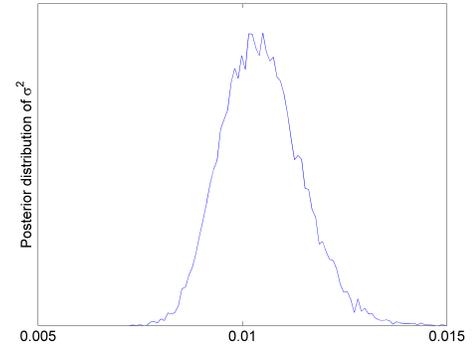

Fig. 4. Estimated posterior distribution of the variance σ^2.

The performance of the proposed Bayesian model has been compared to the LMM and the fully constrained least-squares (FCLS),

respectively studied in [7] and [11]. We built $P = 625$ spectrally impure synthetic pixels (with $R = 6$ endmembers) according to the LMM, with a SNR approximately equal to 21dB. Due to the high number of endmembers, the VCA algorithm were used to estimate the endmembers spectra instead of the N-FINDR algorithm. Once the pixels have been unmixed using the three algorithms, the mean square errors (MSEs) of the abundance vector have then been determined for these algorithms :

$$\text{MSE}^2 = \frac{1}{P} \sum_{p=1}^{P} (\hat{\boldsymbol{\alpha}}_p^+ - \boldsymbol{\alpha}_p^+)^2, \tag{17}$$

where $\hat{\boldsymbol{\alpha}}_p^+$ denotes the MMSE estimate of the abundance vector $\boldsymbol{\alpha}_p^+$ of the pth pixel. Table 1 reports the corresponding results that show that the NCM algorithm performs significantly better than the LMM and FCLS algorithms. The improved performance obtained with the NCM is due to the flexibility of this model when compared to the usual LMM and FCLS algorithms.

Table 1. Mean square errors of abundance vectors.

	LMM		NCM
	Bayesian [7]	FCLS [11]	
MSE^2	8.91×10^{-2}	6.10×10^{-2}	$4.7 \times 5.51 \times 10^{-2}$

4.2. Real Hyperspectral Image

This section considers a real hyperspectral image of size 50×50 depicted in Fig. 5 (top left). This image has been extracted from a larger image acquired in 1997 by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) over Moffett Field in California. The data set has been reduced from the original $L = 224$ bands to $L = 189$ bands by removing water absorption bands. First, the image has been pre-processed by a PCA to reduce the dimensionality of the data and to know the number of endmembers present in the scene as explained in [1]. Then, the N-FINDR algorithm has been applied to this image in order to determine the endmember spectra. The $R = 3$ extracted endmembers correspond to vegetation, water and soil which will be used in place of $\boldsymbol{m}_1$, $\boldsymbol{m}_2$ and $\boldsymbol{m}_3$. The image fraction maps estimated by the proposed algorithm for the $R = 3$ pure materials are depicted in Fig. 5 (top right and bottom). Note that a red (resp. blue) pixel in the map indicates a large (resp. small) value of the abundance coefficient. Thus, the lake area (represented by red pixels in the water fraction map and by blue pixels in the other maps) can be clearly recovered.

5. CONCLUSIONS

A new hierarchical Bayesian unmixing algorithm was derived for hyperspectral images. This algorithm was based on the normal compositional model introduced by Eismann and Stein [4]. The proposed algorithm generated samples distributed according to the joint posterior of the abundances and the variance of the endmembers. These samples were then used to estimate the parameters of interest. The proposed estimation algorithm has several advantages when compared to standard LMM-based algorithms. In particular, it can be easily generalized to more complex models, e.g. considering different endmember variances. Future investigations include the generalization of the proposed strategy to a semi-supervised unmixing algorithm as in [7].
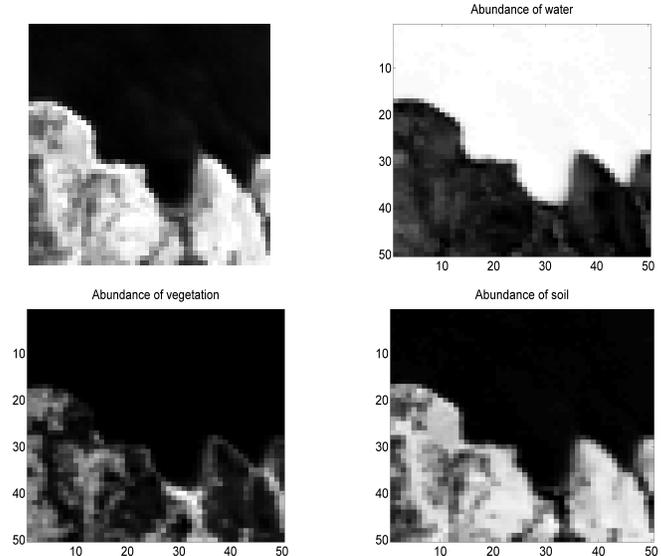

Fig. 5. Top left: region of interest (Moffett field) shown in gray scale (wavelength $\lambda = 0.66$ μm). Top right and bottom: fraction maps estimated by the proposed algorithm.

6. REFERENCES

[1] N. Keshava and J. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, pp. 44–56, Jan. 2002.

[2] M. E. Winter, "Fast autonomous spectral end-member determination in hyperspectral data," in *Proc. 13th Int. Conf. on Applied Geologic Remote Sensing*, vol. 2, Vancouver, April 1999, pp. 337–344.

[3] 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.

[4] M. T. Eismann and D. Stein, "Stochastic mixture modeling," in *Hyperspectral Data Exploitation: Theory and Applications*, C.-I. Chang, Ed. Wiley, 2007, ch. 5.

[5] J. Settle, "On the relationship between spectral unmixing and subspace projection," *IEEE Trans. Geosci. and Remote Sensing*, vol. 34, no. 4, pp. 1045–1046, July 1996.

[6] D. Manolakis, C. Siracusa, and G. Shaw, "Hyperspectral subpixel target detection using the linear mixing model," *IEEE Trans. Geosci. and Remote Sensing*, vol. 39, no. 7, pp. 1392–1409, July 2001.

[7] N. Dobigeon, J.-Y. Tourneret, and C.-I Chang, "Semi-supervised linear spectral using a hierarchical bayesian model for hyperspectral imagery," *IEEE Trans. Signal Processing*, vol. 56, no. 7, pp. 2684–2696, July 2008.

[8] D. Stein, "Application of the normal compositional model to the analysis of hyperspectral imagery," in *Proc. IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, Oct. 2003, pp. 44–51.

[9] J. Diebolt and E. H. S. Ip., "Stochastic EM: method and application," in *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, Eds. London: Chapman & Hall, 1996.

[10] E. Punskaya, C. Andrieu, A. Doucet, and W. Fitzgerald, "Bayesian curve fitting using MCMC with applications to signal segmentation," *IEEE Trans. Signal Processing*, vol. 50, no. 3, pp. 747–758, March 2002.

[11] 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. 39, no. 3, pp. 529–545, March 2001.