Bayesian estimation of linear mixtures using the normal compositional model. Application to hyperspectral imagery

Olivier Eches, Nicolas Dobigeon, Corinne Mailhes, Jean-Yves Tourneret

Bayesian Estimation of Linear Mixtures Using the Normal Compositional Model. Application to Hyperspectral Imagery

Olivier Eches, Nicolas Dobigeon, *Member, IEEE*, Corinne Mailhes, *Member, IEEE*, and Jean-Yves Tourneret, *Senior Member, IEEE*

***Abstract*—This paper studies a new Bayesian unmixing algorithm for hyperspectral images. Each pixel of the image is modeled as a linear combination of so-called *endmembers*. These endmembers are supposed to be random in order to model uncertainties regarding their knowledge. More precisely, we model endmembers as Gaussian vectors whose means have been determined using an endmember extraction algorithm such as the famous N-finder (N-FINDR) or Vertex Component Analysis (VCA) algorithms. 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 conjugate priors are chosen for the remaining parameters. A hybrid Gibbs sampler is then constructed to generate abundance and variance samples distributed according to the joint posterior of the abundances and noise variances. The performance of the proposed methodology is evaluated by comparison with other unmixing algorithms on synthetic and real images.**

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

I. 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 mixture of R spectra $\boldsymbol{m}_r$, $r = 1 \ldots R$, corrupted by additive white Gaussian noise

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

The authors are with the University of Toulouse, IRIT/INP-ENSEEIHT/ TéSA, 31071 Toulouse cedex 7, France (e-mail: olivier.eches@enseeiht.fr; nicolas.dobigeon@enseeiht.fr; corinne.mailhes@enseeiht.fr; jean-yves.tourneret@enseeiht.fr).

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 the observed scene, and L is the number of available spectral bands for the image. Supervised algorithms assume that the R endmember 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 sum-to-one constraints:

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

The LMM 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, outlined in [3], is illustrated in Fig. 1. This figure shows 1) the dual-band projections [on the two most discriminant axes identified by a principal component analysis (PCA)] of $R = 3$ endmembers (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 model referred to as normal compositional model (NCM) was recently proposed in [4]. The NCM allows one to alleviate the problems mentioned above by assuming that the pixels of the hyperspectral image are linear combinations of random endmembers (as opposed to deterministic for the LMM) with known means (e.g., resulting from the N-FINDR or 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

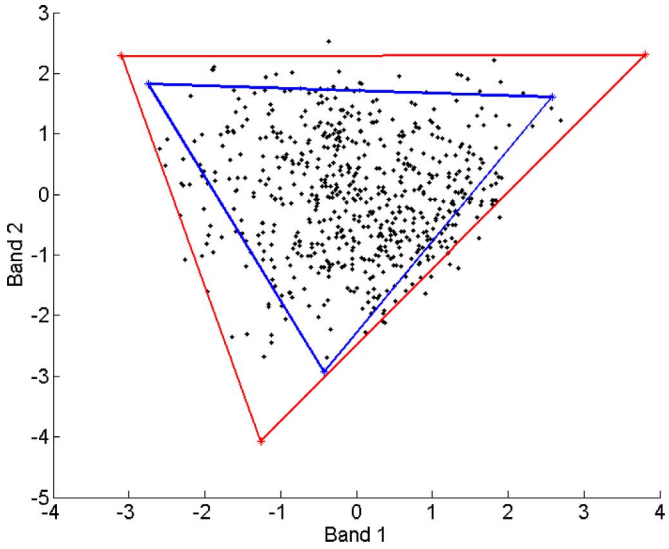

Fig. 1. Scatterplot of dual-band correct (red/light gray) and incorrect (blue/dark gray) results of the N-FINDR algorithm.

illustrated in Fig. 1. The NCM assumes that the spectrum of a mixed pixel can be written as follows:

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

where the $\boldsymbol{\varepsilon}_r$ 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 is proportional to the identity matrix. As a consequence, the endmember variances do not vary from one spectral band to another.[1]

In this paper, a new Bayesian unmixing algorithm is derived for the NCM to estimate the abundance coefficients in (3) under the constraints in (2). Appropriate prior distributions are chosen for the NCM abundances to satisfy the positivity and sum-to-one constraints, as in [7]. A conjugate inverse Gamma distribution is defined for the endmember variance. The hyperparameter of this model can be fixed using appropriate prior information, or estimated jointly with the other unknown parameters. A classical procedure consists of assigning a vague prior to this hyperparameter resulting in a hierarchical Bayesian model [8, p. 392]. The parameters and hyperparameter of this hierarchical Bayesian model can then be estimated by using the full posterior distribution. Unfortunately the joint posterior distribution for the NCM 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], [9]. However, this algorithm can have "*serious shortcomings including the convergence to a local maximum of the posterior*" [10, p. 259]. These shortcomings can be bypassed by considering Markov Chain Monte Carlo (MCMC) methods that 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 developed in [7] and shows that it can be used efficiently for the NCM. Note that other Bayesian algorithms have been also proposed for multispectral and hyperspectral image analysis. In [11], Moussaoui *et al.* have coupled Bayesian blind source separation with independent component analysis to investigate the composition of the Mars surface. This approach, relied on MCMC methods, has allowed them to handle the spectral unmixing problem in an unsupervised framework. In [12], classification and segmentation of hyperspectral images have been addressed using a Bayesian model with a Potts–Markov field to take into account spatial constraints. More recently, Snoussi introduced in [13] an MCMC algorithm to extract the cosmic microwave background power spectrum in astrophysical data.

The paper is organized as follows. Section II derives the posterior distribution of the unknown parameter vector resulting from the proposed Bayesian model. Section III studies the hybrid Gibbs sampling strategy that is used to generate samples distributed according to the NCM posterior. Sections IV and V extend the proposed result to endmembers with different variances. Simulation results conducted on synthetic data are presented in Section VI. In particular, some comparisons between the proposed Bayesian strategies and classical unmixing algorithms are presented in this section. Results obtained with these algorithms on a real image are finally presented in Section VII. Conclusions are reported in Section VIII.

II. Hierarchical Bayesian Model

This section studies 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 sum-to-one constraints.

A. Likelihood

The NCM assumes that the endmember spectra $\boldsymbol{\varepsilon}_r$, $r = 1, \ldots, R$, are independent Gaussian vectors with known mean vectors $\boldsymbol{m}_r = [m_{r,1}, \ldots, m_{r,L}]^T$, $r = 1, \ldots, R$. Moreover, we first assume 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 any spectral band, i.e., $\boldsymbol{\varepsilon}_r | \boldsymbol{m}_r, \sigma^2 \sim \mathcal{N}(\boldsymbol{m}_r, \sigma^2 \boldsymbol{I}_L)$ where $\mathcal{N}(\boldsymbol{m}, \boldsymbol{\Sigma})$ denotes the multivariate Gaussian distribution with mean vector $\boldsymbol{m}$ and covariance matrix $\boldsymbol{\Sigma}$. Using (3) and the *a priori* independence between the endmember spectra, the likelihood of the observed pixel $\boldsymbol{y}$ can be written as

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

[1]Note that more sophisticated models 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].

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

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

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

B. Parameter Priors

1) Abundance Prior: Because of the sum-to-one constraint inherent to the mixing model, the abundance vector can be rewritten as $\boldsymbol{\alpha}^+ = \left[\boldsymbol{\alpha}^T, \alpha_R\right]^T$, where $\alpha_R = 1 - \sum_{r=1}^{R-1} \alpha_r$. Moreover, to satisfy the positivity constraint, the abundance sub-vector $\boldsymbol{\alpha}$ has to live in a simplex defined by

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

A uniform distribution on this simplex is chosen as prior distribution for the partial abundance vector $\boldsymbol{\alpha}$

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

where $\propto$ means "proportional to" and $\mathbf{1}_{\mathbb{S}}(\cdot)$ is the indicator function defined on the set $\mathbb{S}$

$$\mathbf{1}_{\mathbb{S}}(\boldsymbol{\alpha}) = \begin{cases} 1, & \text{if } \boldsymbol{\alpha} \in \mathbb{S}; \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

This prior ensures the positivity and sum-to-one constraints of the abundance coefficients and reflects the absence of other prior knowledge regarding these parameters. Note that any abundance could be removed from $\boldsymbol{\alpha}^+$ and not only the last one α_R. For symmetry reasons, the algorithm proposed in Section III will remove one abundance coefficient from $\boldsymbol{\alpha}^+$ uniformly drawn in $\{1, \ldots, R\}$. Here, this component is supposed to be α_R to simplify notations. Moreover, for sake of conciseness, the notations $\boldsymbol{\mu}(\boldsymbol{\alpha})$ and $c(\boldsymbol{\alpha})$ will be used in the sequel to denote the quantities in (5), where α_R has been replaced by $1 - \sum_{r=1}^{R-1} \alpha_r$.

2) Endmember Variance Prior: The prior distribution for the variance σ^2 is a conjugate inverse Gamma distribution

$$\sigma^2 | \delta \sim \mathcal{IG}(\nu, \delta) \tag{9}$$

where ν and δ are two adjustable hyperparameters (referred to as shape and scale parameters [8, p. 582]. This paper classically assumes $\nu = 1$ (as in [14] or [15]) 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 the non-informative Jeffreys' prior defined by

$$f(\delta) \propto \frac{1}{\delta} \mathbf{1}_{\mathbb{R}^+}(\delta). \tag{10}$$

This prior reflects the lack of knowledge regarding the hyperparameter δ.

C. Posterior Distribution of the Parameters

The joint posterior distribution of the unknown parameter vector $\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \sigma^2\}$ and hyperparameter δ can be derived using the hierarchical structure

$$f(\boldsymbol{\theta}, \delta | \boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\delta) f(\delta) \tag{11}$$

where $f(\boldsymbol{y}|\boldsymbol{\theta})$ and $f(\delta)$ have been defined in (4) and (10), respectively. Assuming independence between the unknown parameters, the prior distribution of $\boldsymbol{\theta}$ is $f(\boldsymbol{\theta}\,|\delta) = f(\boldsymbol{\alpha}) f(\sigma^2|\delta)$, yielding

$$f(\boldsymbol{\theta}, \delta | \boldsymbol{y}) \propto \exp\left(-\frac{\|\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\alpha})\|^2}{2\sigma^2 c(\boldsymbol{\alpha})} - \delta \right) \frac{\mathbf{1}_{\mathbb{S}}(\boldsymbol{\alpha}) \mathbf{1}_{\mathbb{R}^+}(\delta)}{c\left(\boldsymbol{\alpha}^+\right)^{L/2} \sigma^{L+2}}. \tag{12}$$

The posterior distribution (12) is too complex to derive the MMSE or MAP estimators of the unknown parameter vector of interest, i.e., the vector of abundances $\boldsymbol{\alpha}^+$. An interesting alternative is to generate samples distributed according to the posterior and to use the generated samples to approximate the Bayesian estimators [8]. Section III studies a hybrid Gibbs sampler that generates abundances and variances distributed according to the full posterior (12).

III. Hybrid Gibbs Sampler

This section studies a hybrid Metropolis-within-Gibbs sampler that generates samples according to the posterior $f(\boldsymbol{\theta}, \delta|\boldsymbol{y})$. The sampler iteratively generates $\boldsymbol{\alpha}$ according to $f(\boldsymbol{\alpha}|\boldsymbol{y}, \sigma^2)$, σ^2 according to $f(\sigma^2|\boldsymbol{y}, \boldsymbol{\alpha}, \delta)$, and δ according to $f(\delta|\sigma^2)$, as detailed below. The overall hybrid Gibbs sampler algorithm is summarized in Algorithm 1.

Algorithm 1: Hybrid Gibbs sampler for hyperspectral unmixing using the NMC

1) *Initialization:*
   - Sample $\delta^{(0)}$ from the probability density function (pdf) in (10),
   - Sample $\sigma^{2(0)}$ from the pdf in (9),
2) *Iterations*: For $t = 1, 2, \ldots$, do
   - Sample $\boldsymbol{\alpha}^{(t)}$ from the pdf in (14) using Metropolis-within-Gibbs step,
   - Sample $\sigma^{2(t)}$ from the pdf in (16),
   - Sample $\delta^{(t)}$ from the pdf in (17),

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

The Bayes' theorem yields

$$f(\boldsymbol{\alpha}|\boldsymbol{y}, \sigma^2) \propto f(\boldsymbol{y}|\boldsymbol{\theta}) f(\boldsymbol{\alpha}) \tag{13}$$

which easily leads to

$$f(\boldsymbol{\alpha}|\boldsymbol{y},\sigma^2) \propto \frac{1}{\left[\sigma^2 c(\boldsymbol{\alpha})\right]^{L/2}} \exp\left(-\frac{\|\boldsymbol{y}-\boldsymbol{\mu}(\boldsymbol{\alpha})\|^2}{2\sigma^2 c(\boldsymbol{\alpha})}\right) \mathbf{1}_{\mathbb{S}}(\boldsymbol{\alpha}). \tag{14}$$

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 (14) can be achieved using a Metropolis-within-Gibbs algorithm. We have used the uniform prior distribution (7) as proposal distribution for this algorithm.

B. Generation According to $f\left(\sigma^2|\boldsymbol{y},\boldsymbol{\alpha},\delta\right)$

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

$$f\left(\sigma^2|\boldsymbol{y},\boldsymbol{\alpha},\delta\right) \propto f(\boldsymbol{y}|\boldsymbol{\theta}) f\left(\sigma^2|\delta\right). \tag{15}$$

Consequently, $\sigma^2|\boldsymbol{y},\boldsymbol{\alpha},\delta$ is distributed according to the following inverse-Gamma distribution:

$$\sigma^2|\boldsymbol{y},\boldsymbol{\alpha},\delta \sim \mathcal{IG}\left(\frac{L}{2}+1, \frac{\|\boldsymbol{y}-\boldsymbol{\mu}(\boldsymbol{\alpha})\|^2}{2c(\boldsymbol{\alpha})}+\delta\right). \tag{16}$$

C. Generation According to $f\left(\delta|\sigma^2\right)$

The conditional distribution of δ is

$$\delta|\sigma^2 \sim \mathcal{G}\left(1, \frac{1}{\sigma^2}\right) \tag{17}$$

where $\mathcal{G}(a,b)$ is the Gamma distribution with shape parameter a and scale parameter b [8, p. 581].

IV. Extension to Endmember Spectra With Different Variances

In the previous sections, all endmember spectra shared the same variance σ^2. We propose here to extend the previous model to the case where endmembers have different variances. This additional degree of freedom can be particularly interesting when different levels of confidence are given to the mean vectors $\boldsymbol{m}_r$ $(r=1,\ldots,R)$ identified by the N-FINDR or VCA algorithms. Thus, a new vector $\boldsymbol{\sigma} = \left[\sigma_1^2,\ldots,\sigma_R^2\right]^T$ is introduced, where σ_r^2 is the rth endmember variance. This assumption leads to

$$\boldsymbol{\varepsilon}_r|\boldsymbol{m}_r,\sigma_r^2 \sim \mathcal{N}\left(\boldsymbol{m}_r,\sigma_r^2\boldsymbol{I}_L\right). \tag{18}$$

A. Identifiability Issue

1) General Theory: If the prior distributions chosen for σ_r^2 $(r=1,\ldots,R)$ are not sufficiently informative, identifiablity issues may occur. In order to clarify this identifiability problem, assume that R endmembers are involved in the mixture, leading to the following log-likelihood:

$$\log f\left(\boldsymbol{y}|\boldsymbol{\sigma},\boldsymbol{\alpha}\right) = -\frac{L}{2}\log C(\boldsymbol{\sigma}) - \frac{K(\boldsymbol{y},\boldsymbol{\alpha})}{C(\boldsymbol{\alpha},\boldsymbol{\sigma})} \tag{19}$$

where $K(\boldsymbol{y},\boldsymbol{\alpha}) = 1/2\,\|\boldsymbol{y}-\boldsymbol{\mu}(\boldsymbol{\alpha})\|^2$ and $C(\boldsymbol{\alpha},\boldsymbol{\sigma}) = \sum_{r=1}^R \sigma_r^2\alpha_r^2$. Looking for the values of the vector $\boldsymbol{\sigma}$ which maximize the log-likelihood, we equal its R partial derivatives to zero

$$\begin{cases} \frac{\partial \log f(\boldsymbol{y}|\boldsymbol{\sigma})}{\partial \sigma_1^2} = -\frac{L\alpha_1^2}{2C(\boldsymbol{\alpha},\boldsymbol{\sigma})} - \frac{K(\boldsymbol{y},\boldsymbol{\alpha})\alpha_1^2}{[C(\boldsymbol{\alpha},\boldsymbol{\sigma})]^2} = 0 \\ \vdots \\ \frac{\partial \log f(\boldsymbol{y}|\boldsymbol{\sigma})}{\partial \sigma_R^2} = -\frac{L\alpha_R^2}{2C(\boldsymbol{\alpha},\boldsymbol{\sigma})} - \frac{K(\boldsymbol{y},\boldsymbol{\alpha})\alpha_R^2}{[C(\boldsymbol{\alpha},\boldsymbol{\sigma})]^2} = 0 \end{cases} \tag{20}$$

which easily leads to

$$C(\boldsymbol{\alpha},\boldsymbol{\sigma}) = \sum_{r=1}^R \sigma_r^2\alpha_r^2 = \frac{2K}{L}.$$

Consequently, the likelihood $f\left(\boldsymbol{y} \mid \boldsymbol{\sigma},\boldsymbol{\alpha}\right)$ has several maxima located on the hyperplane $\mathcal{H}$ defined by

$$\mathcal{H} = \left\{\boldsymbol{\sigma} = \left[\sigma_1^2,\ldots,\sigma_R^2\right]^T \middle| C(\boldsymbol{\alpha},\boldsymbol{\sigma}) = \frac{2K}{L}\right\} \tag{21}$$

yielding identifiability problems.

However, this problem is alleviated when several pixels with the same characteristics are considered. Assuming the variance vector $\boldsymbol{\sigma}$ is the same for P pixels (with $P>1$), a linear system of P equations is obtained

$$\begin{cases} \sum_{r=1}^R \sigma_r^2\alpha_{r,1}^2 = \frac{2K_1}{L} \\ \vdots \\ \sum_{r=1}^R \sigma_r^2\alpha_{r,P}^2 = \frac{2K_P}{L} \end{cases} \tag{22}$$

where $\alpha_{r,p}$ denotes the abundance of the rth endmember in the pth pixel, $K_p = 1/2\|\boldsymbol{y}_p - \mu\left(\boldsymbol{\alpha}_p\right)\|^2$, $\boldsymbol{\alpha}_p = \left[\alpha_{1,p},\ldots\alpha_{R-1,p}\right]^T$ and $\boldsymbol{y}_p$ is the pth measured spectrum pixel (with $p=1,\ldots,P$). This system can be rewritten as

$$\boldsymbol{\Lambda}\boldsymbol{\sigma} = \frac{2}{L}\boldsymbol{K}$$

with

$$\boldsymbol{\Lambda} = \begin{pmatrix} \alpha_{1,1}^2 & \cdots & \alpha_{R,1}^2 \\ \vdots & \ddots & \vdots \\ \alpha_{1,P}^2 & \cdots & \alpha_{R,P}^2 \end{pmatrix}, \quad \boldsymbol{K} = \begin{pmatrix} K_1 \\ \vdots \\ K_P \end{pmatrix}. \tag{23}$$

Thus, the vector $\boldsymbol{\sigma}$ maximizing the likelihood is unique provided the rank of the matrix $\boldsymbol{\Lambda}$ is equal to R.

2) Examples: We illustrate the identifiability condition when different numbers of pixels are generated from the mixture of $R=2$ endmembers. As an example, a pixel has been generated with $\boldsymbol{\sigma} = [0.006, 0.002]^T$. Fig. 2 shows the corresponding log-likelihood as a function of $\left(\sigma_1^2,\sigma_2^2\right)$ for $P=1$ pixel. This figure clearly shows that the maxima are reached for an infinity of couples $\left(\sigma_1^2,\sigma_2^2\right)$ located on a hyperplane (here a line).

Fig. 3 shows the likelihood as a function of $\boldsymbol{\sigma}$ for $P=2$ pixels. A unique maximum can be observed since the rank of $\boldsymbol{\Lambda}$ equals 2 for this example. The results depicted in Fig. 4 obtained

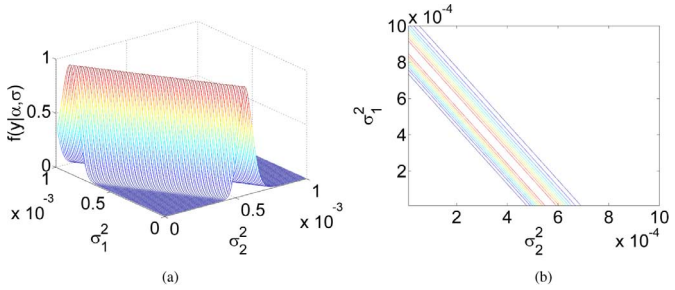

Fig. 2. Likelihood for $P = 1$ pixel as a function of (σ_1^2, σ_2^2). (a) 3-D view. (b) Top view.

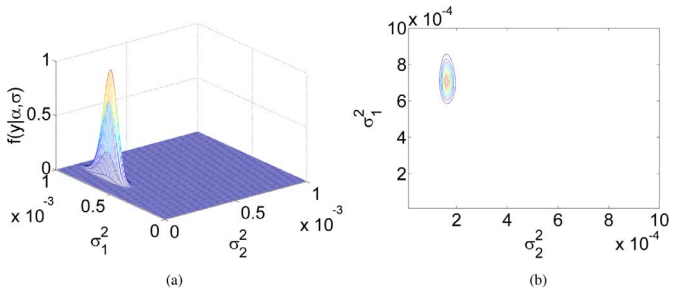

Fig. 3. Likelihood for $P = 2$ pixels as a function of (σ_1^2, σ_2^2). (a) 3-D view. (b) Top view.

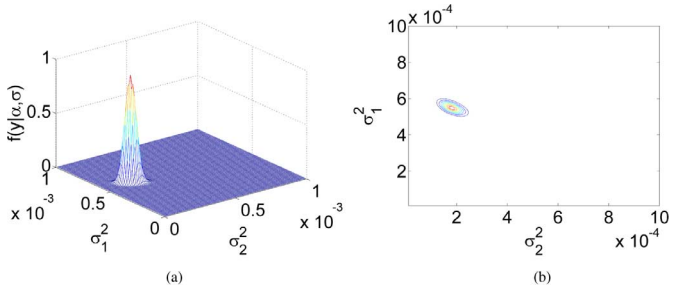

Fig. 4. Likelihood for $P = 9$ pixels as a function of (σ_1^2, σ_2^2). (a) 3-D View. (b) Top view.

for $P = 9$ pixels show that the likelihood is more peaky around the true value of $\boldsymbol{\sigma}$ when more pixels are considered.

B. Hierarchical Bayesian Model

This section derives the hierarchical Bayesian model that can be used to consider different endmember variances σ_r^2 $(r = 1, \ldots, R)$. Motivated by the considerations of the previous paragraph, P pixel spectra are considered

$$\boldsymbol{y}_p = \sum_{r=1}^{R} \boldsymbol{\varepsilon}_r \alpha_{r,p}, \quad p = 1, \ldots, P \tag{24}$$

where $\boldsymbol{\varepsilon}_r | \boldsymbol{m}_r, \sigma_r^2 \sim \mathcal{N}\left(\boldsymbol{m}_r, \sigma_r^2 \boldsymbol{I}_L\right)$, $\boldsymbol{m}_r = [m_{r,1}, \ldots, m_{r,L}]^T$ represents the known mean vector of the endmember vector $\boldsymbol{\varepsilon}_r$, and $\boldsymbol{\sigma} = \left[\sigma_1^2, \ldots, \sigma_R^2\right]^T$ is the unknown variance vector. A standard matrix formulation yields

$$\boldsymbol{Y} = \boldsymbol{E}\boldsymbol{A} \tag{25}$$

where

$$\boldsymbol{Y} = \begin{pmatrix} y_{1,1} & \cdots & y_{1,P} \\ \vdots & \ddots & \vdots \\ y_{L,1} & \cdots & y_{L,P} \end{pmatrix}, \quad \boldsymbol{E} = [\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_R]$$

and

$$\boldsymbol{A} = \begin{pmatrix} \alpha_{1,1} & \cdots & \alpha_{1,P} \\ \vdots & \ddots & \vdots \\ \alpha_{R,1} & \cdots & \alpha_{R,P} \end{pmatrix}. \tag{26}$$

The corresponding likelihood and prior distributions are described next.

1) Likelihood: The likelihood function for the pixel $\#p$ is

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

with

$$\begin{aligned} c(\boldsymbol{\alpha}_p) &= \sum_{r=1}^{R} \sigma_r^2 \alpha_{r,p}^2, \\ \boldsymbol{\mu}(\boldsymbol{\alpha}_p) &= \sum_{r=1}^{R} \boldsymbol{m}_r \alpha_{r,p}. \end{aligned} \tag{27}$$

Assuming the pixel spectra $\boldsymbol{y}_p$ $(p = 1, \ldots, P)$ are *a priori* independent, the joint likelihood for the set of P pixels can be written

$$f(\boldsymbol{Y} | \boldsymbol{A}, \boldsymbol{\sigma}) \propto \prod_{p=1}^{P} \frac{1}{[c(\boldsymbol{\alpha}_p)]^{L/2}} \exp\left[-\sum_{p=1}^{P} \frac{\|\boldsymbol{y}_p - \boldsymbol{\mu}(\boldsymbol{\alpha}_p)\|^2}{2c(\boldsymbol{\alpha}_p)}\right]. \tag{28}$$

2) Prior Distributions: Independent uniform distributions on the simplex defined in (6) are chosen as prior distributions for the partial abundance vectors $\boldsymbol{\alpha}_p$ $(p = 1, \ldots, P)$ yielding

$$f(\boldsymbol{A}) \propto \prod_{p=1}^{P} \mathbf{1}_{\mathbb{S}}(\boldsymbol{\alpha}_p). \tag{29}$$

The prior distributions for the endmember variances are conjugate inverse Gamma distributions with a common hyperparameter δ [as in (9)]. A Jeffreys' prior is assigned to the hyperparameter δ as in (10).

V. MCMC Algorithm for Endmembers With Different Variances

As in the previous case, a hybrid Metropolis-within-Gibbs sampler will be used to generate samples asymptotically distributed according to the joint distribution of the abundance vectors and endmember variances. The sampler iteratively generates $\boldsymbol{\alpha}_p$ according to $f\left(\boldsymbol{\alpha}_p | \boldsymbol{y}_p, \boldsymbol{\sigma}\right)$ for each pixel $p = 1, \ldots, P$, σ_r^2 according to $f\left(\sigma_r^2 | \boldsymbol{\sigma}_{-r}, \boldsymbol{Y}, \boldsymbol{A}, \delta\right)$ for each endmembers $r = 1, \ldots, R$ ($\boldsymbol{\sigma}_{-r}$ denotes the variance vector $\boldsymbol{\sigma}$ whose rth component has been removed), and δ according to $f(\delta | \boldsymbol{\sigma})$.

A. Abundance Generation

The conditional posterior distribution of the abundance vector $\boldsymbol{\alpha}_p$ does not depend on the other pixels and is expressed as

$$f\left(\boldsymbol{\alpha}_p|\boldsymbol{y}_p,\boldsymbol{\sigma}\right) \propto \frac{1}{\left[c(\boldsymbol{\alpha}_p)\right]^{L/2}} \exp\left[-\frac{||\boldsymbol{y}_p-\boldsymbol{\mu}(\boldsymbol{\alpha}_p)||^2}{2c(\boldsymbol{\alpha}_p)}\right] \mathbf{1}_{\mathbb{S}}(\boldsymbol{\alpha}_p). \tag{30}$$

Generating $\boldsymbol{\alpha}_p$ according to this posterior is achieved with a Metropolis-within-Gibbs algorithm similar to the one described in paragraph Section III-A.

B. Variance Generation

The generation according to $f\left(\sigma_r^2|\boldsymbol{\sigma}_{-r},\boldsymbol{Y},\boldsymbol{A}\right)$ is achieved by R Metropolis–Hastings moves. Each Metropolis–Hastings move consists of drawing a variance σ_r^2 according to its conditional distribution

$$f\left(\sigma_r^2|\boldsymbol{\sigma}_{-r},\boldsymbol{Y},\boldsymbol{A},\delta\right) \propto f(\boldsymbol{Y}|\boldsymbol{A},\boldsymbol{\sigma}) f\left(\sigma_r^2|\nu,\delta\right)$$

with $\boldsymbol{\sigma}_{-r} = \left[\sigma_1^2,\ldots,\sigma_{r-1}^2,\sigma_{r+1}^2,\ldots,\sigma_r^2\right]^T$. Introducing $c(\boldsymbol{\alpha}_{-r}) = \sum_{i\neq r}\sigma_i^2\alpha_i^2$ straightforward computations lead to (see the Appendix)

$$f\left(\sigma_r^2|\boldsymbol{\sigma}_{-r},\boldsymbol{Y},\boldsymbol{A},\delta\right) \propto \left(\frac{1}{\sigma_r^2}\right)^{\nu+1} \prod_{p=1}^{P}\left[\sigma_r^2\alpha_{r,p}^2 + c(\boldsymbol{\alpha}_{-r})\right]^{-L/2} \times \exp\left[-\sum_{p=1}^{P}\frac{||\boldsymbol{y}_p-\boldsymbol{\mu}(\boldsymbol{\alpha}_p)||^2}{2\left[\sigma_r^2\alpha_{r,p}^2 + c(\boldsymbol{\alpha}_{-r})\right]} - \frac{\delta}{\sigma_r^2}\right]. \tag{31}$$

Sampling according to (31) is achieved thanks to a Metropolis–Hastings step. The proposal distribution for this algorithm is an inverse Gamma distribution

$$\sigma_r^2 \sim \mathcal{IG}\left(\alpha_\sigma,\beta_\sigma\right) \tag{32}$$

where α_σ and β_σ are adjustable parameters. These parameters have been chosen in order to obtain the mean and the variance of the distribution (16), which improves the acceptance rate of the sampler.

C. Hyperparameter Generation

The conditional distribution of the hyperparameter δ upon $\boldsymbol{\sigma}$ is the following Gamma distribution:

$$\delta|\boldsymbol{\sigma} \sim \mathcal{G}\left(R,\sum_{r=1}^{R}\frac{1}{\sigma_r^2}\right). \tag{33}$$

A detailed step-by-step algorithm is presented in Algorithm 2.

ALGORITHM 2: Spectral unmixing using the NCM with different endmember variances.

1) *Initialization:*
   - Sample the hyperparameter $\delta^{(0)}$ from the pdf in (10),
   - Sample $\boldsymbol{\sigma}^{(0)} = \left[\sigma_1^{2\,(0)},\ldots,\sigma_R^{2\,(0)}\right]$ from the pdf in (9),
   - For each pixel p, sample $\boldsymbol{\alpha}^{(0)}$ according to a uniform distribution on $\mathbb{S}$,
2) *Iterations*: For $t = 1,2,\ldots$, do
   - For $p = 1,\ldots,P$, sample $\boldsymbol{\alpha}_p^{(t)}$ from the pdf in (31) using Metropolis-within-Gibbs,
   - For $r = 1,\ldots,R$, sample $\sigma_r^{2\,(t)}$ from the pdf in (31) using Metropolis-within-Gibbs,
   - Sample $\delta^{(t)}$ from the pdf in (17),

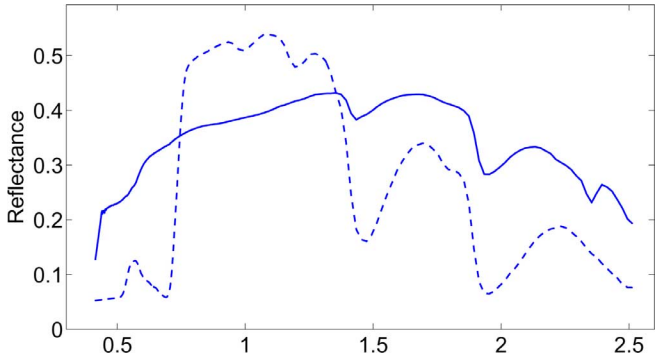

Fig. 5. Endmember spectra: construction concrete (solid line), green grass (dashed line).

VI. Simulation Results on Synthetic Data

This section illustrates the performance of the two proposed unmixing algorithms via simulations on synthetic data. The simulations have been conducted on pixels observed in $L = 276$ spectral bands ranging from wavelength 0.4 μm to 2.5 μm (from the visible to the near infrared).

A. NCM Algorithm With a Single Endmember Variance

The simulation depicted in this section have been obtained for the NCM algorithm introduced in Section III. A synthetic mixture of $R = 2$ endmembers is considered in this experiment. This trivial example has the advantage of having few parameters whose posteriors can be represented more easily. The means of these endmembers $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$ have been extracted from the spectral libraries distributed with the ENVI package [16]. These spectra correspond to construction concrete and green grass and are depicted in Fig. 5. The endmember variance is $\sigma^2 = 0.01$. The linear mixture considered in this section is defined by $\boldsymbol{\alpha}^+ = [0.3,0.7]^T$. Fig. 6 shows the posterior distributions of the abundances generated by the proposed Gibbs sampler with $N_{\mathrm{MC}} = 25000$ iterations including $N_{\mathrm{bi}} = 5000$ burn-in iterations[2]. These distributions are in good agreement with the actual values of the abundances. Fig. 7 shows the estimated posterior distribution of σ^2 that is also in good agreement with the actual endmember variance $\sigma^2 = 0.01$.

The proposed Gibbs algorithm has been also tested for different values of the signal-to-noise ratio (SNR). Fig. 8 shows the abundance MAP estimates of α_r and the corresponding standard deviations as a function of the SNR. Note that the proposed Bayesian algorithm allows one to derive confidence intervals for

[2]Classically, the first samples generated by the Gibbs sampler (belonging to the so-called burn-in period) are not considered for parameter estimation.

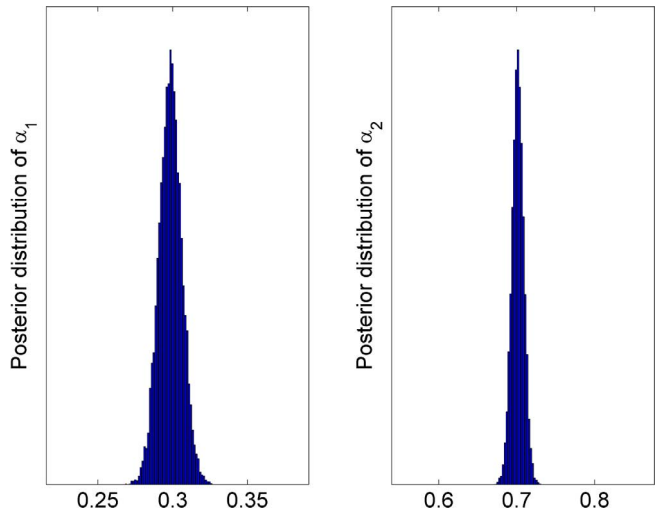

Fig. 6. Estimated posterior distributions of the abundances $[\alpha_1,\ \alpha_2]^T$.

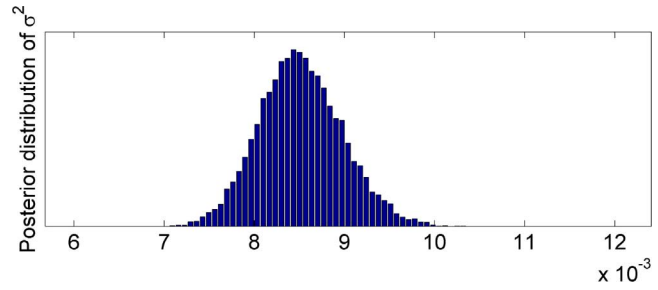

Fig. 7. Estimated posterior distribution of the variance σ^2.

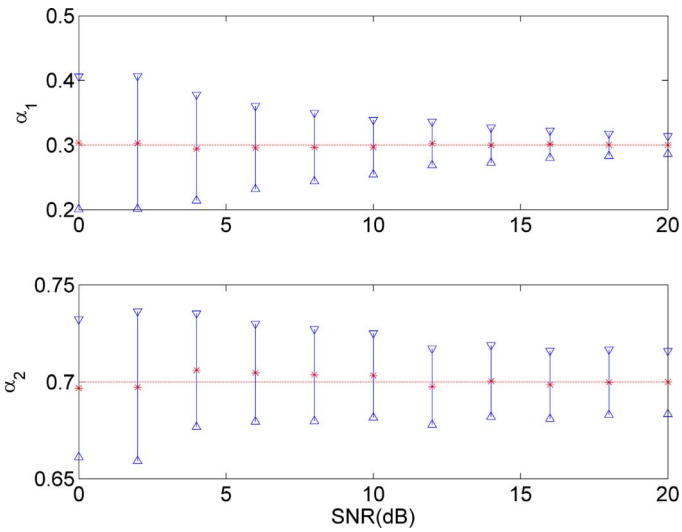

Fig. 8. MAP estimates (cross) and standard deviations (vertical bars) of the components of $\boldsymbol{\alpha}^+$ versus SNR.

the different estimates. These confidence intervals are computed from the samples generated by the Gibbs sampler. Note also that the SNRs of the actual spectrometers like AVIRIS are not below 20 dB when the water absorption bands have been removed [17]. The results in Fig. 8 indicate that the proposed Bayesian algorithm performs satisfactorily for these SNRs. Fig. 8 also shows that the proposed estimates of α_r converge (in the mean square sense) to the actual values of α_r when the SNR level increases.

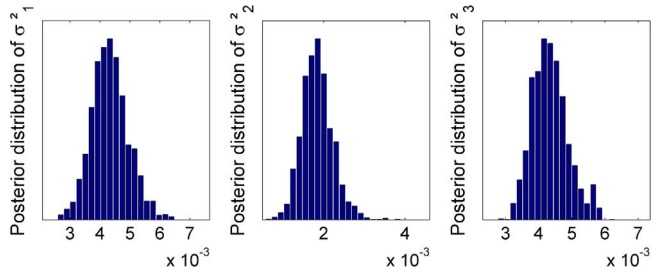

Fig. 9. Estimated posterior distribution of the variances for $P = 3$ pixels.

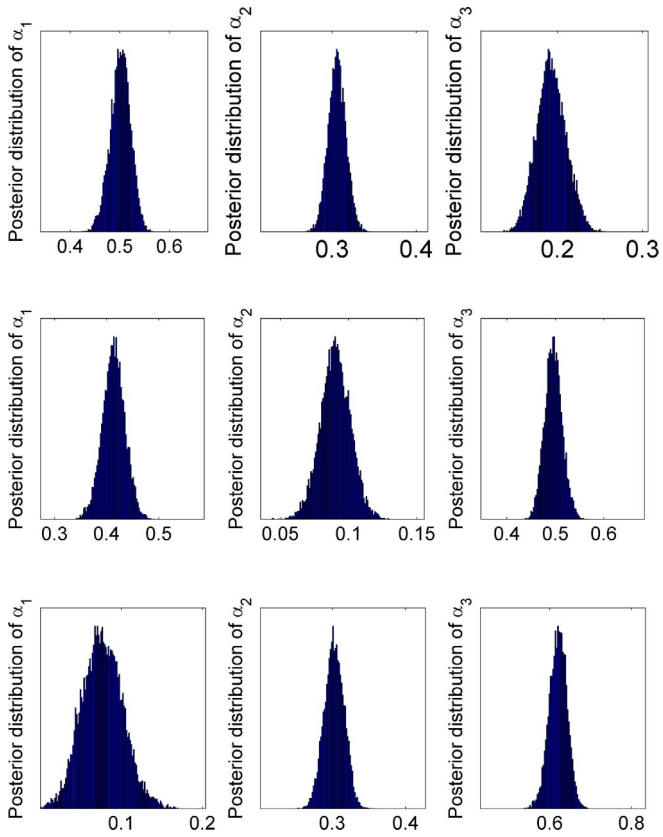

Fig. 10. Estimated posterior distributions of the abundances for each pixel (top: pixel 1, center: pixel 2, bottom: pixel 3).

B. NCM Algorithm With Different Variances

The performance of the algorithm introduced in Section IV is illustrated with simulation results associated to synthetic data. In these simulations, $P = 3$ pixels have been generated by mixing $R = 3$ endmembers according to (24). The actual parameter values are as follows.

- Pixel 1: $\boldsymbol{\alpha}_1^+ = [0.5, 0.3, 0.2]^T$, $\sigma_1^2 = 0.004$.
- Pixel 2: $\boldsymbol{\alpha}_2^+ = [0.4, 0.1, 0.5]^T$, $\sigma_2^2 = 0.002$.
- Pixel 3: $\boldsymbol{\alpha}_3^+ = [0.1, 0.3, 0.6]^T$, $\sigma_2^2 = 0.0035$.

Fig. 9 shows the estimated posterior distributions of the variances σ_r^2 $(r = 1, \ldots, R)$ that are clearly centered around the actual values. The histograms of the abundances generated for each pixel by the proposed hybrid Gibbs sampler are depicted in Fig. 10. These results are in good agreement with the actual values of the abundances.

The performance of the algorithm based on different endmember variances (described in Section IV) is compared to the

TABLE II
GLOBAL MSES OF EACH ABUNDANCE COMPONENT FOR DIFFERENT UNMIXING ALGORITHMS ($\times 10^{-3}$)

	Bayesian NCM	Bayesian LMM	FCLS	MVC-NMF	NN-ICA
MSE_1^2	7.8	13	9.1	**7.7**	18.2
MSE_2^2	**9.6**	10.4	9.9	24.1	41.4
MSE_3^2	**8.5**	23.2	10.2	45.4	45.2
MSE_4^2	**8.2**	15.9	8.8	26.2	45.3
MSE_5^2	**10.2**	14.8	11.5	12.5	46.8
MSE_6^2	**10.8**	11.7	11.5	35.6	44.9

TABLE I
GLOBAL MSE OF THE ABUNDANCE VECTOR FOR THE NCM WITH UNIQUE VARIANCE AND WITH DISTINCT VARIANCES

NCM with single variance	NCM with multiple variances
1.72×10^{-2}	1.54×10^{-2}

algorithm based on a single endmember variance (described in Section III). $P = 9$ synthetic pixels, generated according to the NCM with distinct variances, have been unmixed by the two different algorithms. The mean square errors (MSEs) of the abundance vectors are then computed for these algorithms using 100 Monte Carlo runs. Table I summarizes the corresponding results. Taking into account several variances allows one to improve the estimation performance for this example.

C. Comparison With Other Algorithms

This paragraph presents a comparison between the two algorithms developed in this paper and other strategies previously proposed in the literature. More precisely, we compare the following unmixing strategies:

- the proposed Bayesian NCM algorithm presented in Section II;
- a Bayesian algorithm derived from the LMM [7];
- the fully constrained least-squares (FCLS) method [18];
- the minimum volume constrained nonnegative matrix factorization (MVC-NMF) [19];
- the non-negative independent component analysis (NN-ICA) [20].

The Bayesian NCM and the LMM-based algorithms of [7] and [18] are coupled with the VCA algorithm as an endmember extraction algorithm (EEA). Note that any other standard EEAs (such as N-FINDR and pixel purity index [21]) could have been used in place of VCA. $P = 625$ synthetic pixels are generated according to the LMM with $R = 6$ endmembers, corrupted by an additive Gaussian noise leading to an SNR equal to 20 dB. To evaluate the robustness of the NCM to the absence of pure pixels, the observations close to the endmember means (i.e., such that $\left\|\boldsymbol{y}_p - \boldsymbol{m}_r\right\|^2 / L < \delta, \forall\, p, r$, with $\delta = 6.0 \times 10^{-2}$) have been removed from the synthetic image. The global MSE of the rth estimated abundance is defined as

$$\text{MSE}_r^2 = \frac{1}{P}\sum_{p=1}^{P}(\hat{\alpha}_{r,p} - \alpha_{r,p})^2 \tag{34}$$

where $\hat{\alpha}_{r,p}$ denotes the MMSE estimate of the abundance $\alpha_{r,p}$. Table II shows the global MSEs for the five different unmixing

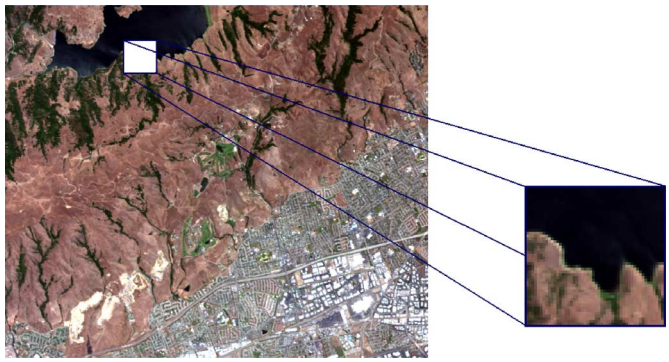

Fig. 11. Real hyperspectral data: Moffett field acquired by AVIRIS in 1997 (left) and the region of interest shown in true colors (right).

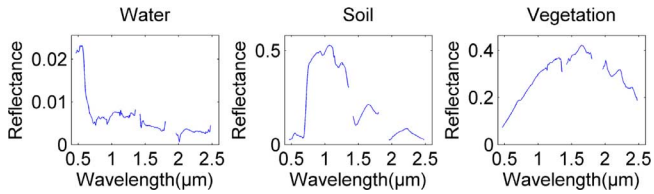

Fig. 12. $R = 3$ endmember spectra obtained by the N-FINDR algorithm.

TABLE III
RECONSTRUCTION ERRORS FOR THE BAYESIAN NCM, THE BAYESIAN LMM AND THE FCLS ALGORITHMS

	NCM	LMM	FCLS
e	1.26	1.32	1.28

strategies mentioned before (Bayesian NCM, Bayesian LMM, FCLS, MVC-NMF and NN-ICA). The proposed Bayesian NCM algorithm performs significantly better than the other unmixing algorithms. The improved performance obtained with the NCM is due to the robustness of this model (when compared to the usual LMM) to the absence of pure pixels in the image.

As a complementary study for this set of pixels, the global reconstruction error defined by

$$e = \sqrt{\frac{1}{P}\sum_{p=1}^{P}\left\|\boldsymbol{y}_p - \boldsymbol{M}\hat{\boldsymbol{\alpha}}_p^+\right\|^2} \tag{35}$$

is reported in Table III for the Bayesian NCM, the Bayesian LMM and the FCLS algorithms[3]. Note that the Bayesian LMM and FCLS algorithms require the *a priori* knowledge of deterministic endmembers $\boldsymbol{m}_1, \ldots, \boldsymbol{m}_R$ contained in the matrix $\boldsymbol{M}$. Consequently, the actual endmember matrix $\boldsymbol{M}$ is also used for computing the reconstruction error associated to the NCM algorithm for fair comparison. As shown in Table III, the Bayesian NCM yields the smaller reconstruction error.

[3]The MVC-NMF and NN-ICA algorithms have not been considered for this comparison since they estimate the endmembers and abundances jointly. Thus, small reconstruction errors for these algorithms do not indicate a good spectral unmixing.

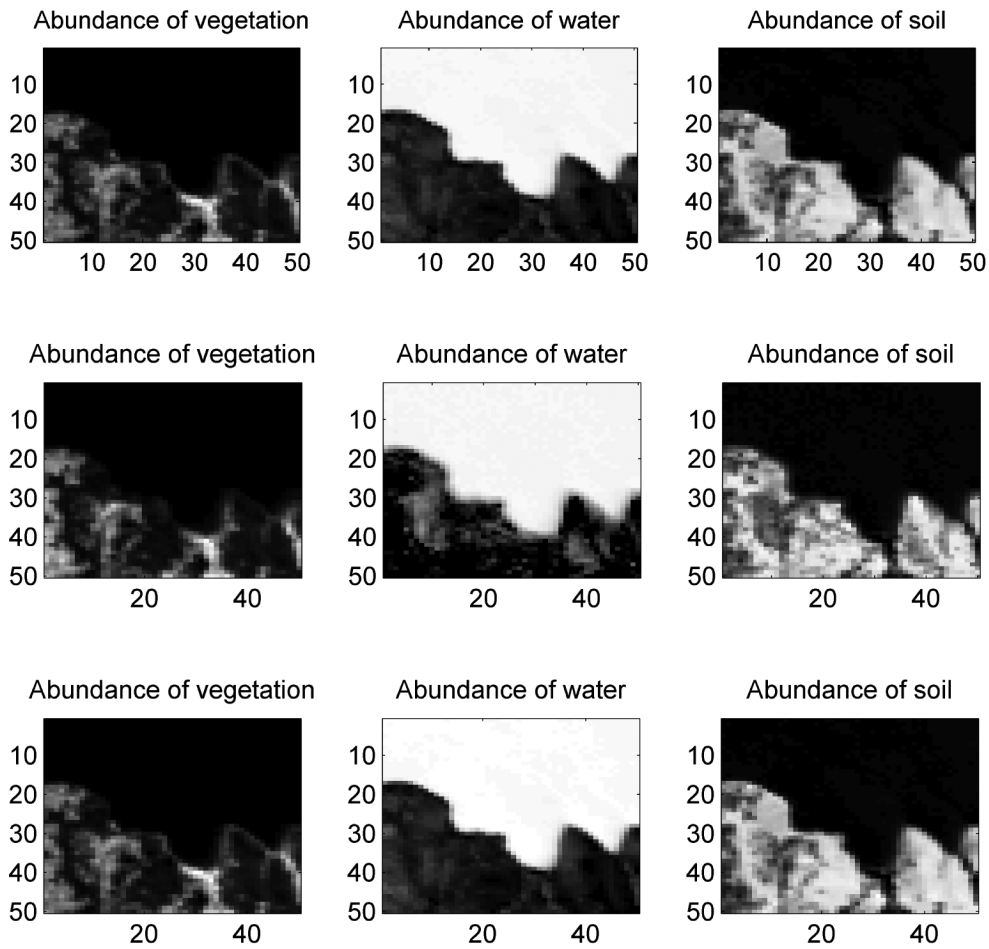

Fig. 13. Top: fraction maps estimated by the LMM algorithm (from [19]). Middle: fraction maps estimated by the FCLS algorithm [18]. Bottom: fraction maps estimated by the proposed algorithm (black (resp. white) means absence (resp. presence) of the material).

VII. Spectral Unmixing of an AVIRIS Image

This section considers a real hyperspectral image of size 50×50 depicted in Fig. 11 to evaluate the performance of the different algorithms. This image has been extracted from a larger image acquired in 1997 by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) over Moffett Field, CA. The data set has been reduced from the original 224 bands to $L = 189$ bands by removing water absorption bands. First, the image has been preprocessed by a PCA to determine the number of endmembers present in the scene as explained in [1]. Then, the N-FINDR algorithm has been applied to this image to estimate the endmember spectra. The $R = 3$ extracted endmembers (shown in Fig. 12) correspond to vegetation, water and soil, and have been used as the mean vectors $\boldsymbol{m}_1$, $\boldsymbol{m}_2$ and $\boldsymbol{m}_3$.

A. NCM Algorithm With a Single Endmember Variance

The image fraction maps estimated by the algorithm proposed in Sections II and III (for the $R = 3$ pure materials) are depicted in Fig. 13 (bottom). Note that a white (resp. black) pixel in the map indicates a large (resp. small) value of the abundance coefficient. Thus, the lake area (represented by white pixels in the water fraction map and by black pixels in the other maps) can be clearly recovered. These results have been compared to the fraction maps estimated with the LMM Bayesian algorithm (proposed in [7]) and the FCLS method [18]. As depicted in Fig. 13, the fraction maps obtained with the three algorithms are clearly in good agreement. Other results given by the MVC-NMF [19] and the NN-ICA [20] are detailed in [22].

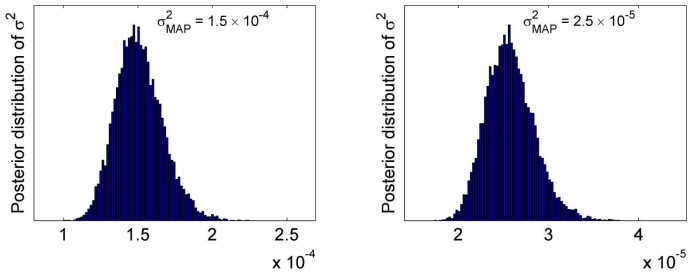

Fig. 14. Posterior distributions of the variance σ^2 for the pixels #(35,43) (left) and #(43,35) (right) estimated by the proposed algorithm.

Some results regarding the estimation of the endmember variance σ^2 are also presented. Fig. 14 shows the estimated posterior distributions of σ^2 for the pixels #(35,43) (left) and #(43,35) (right) of the image as well as their MAP estimates.

The proposed Bayesian algorithm can be used to estimate the probability of endmember presence defined as $\mathrm{P}[\alpha_i > \eta | \boldsymbol{m}_i]$,

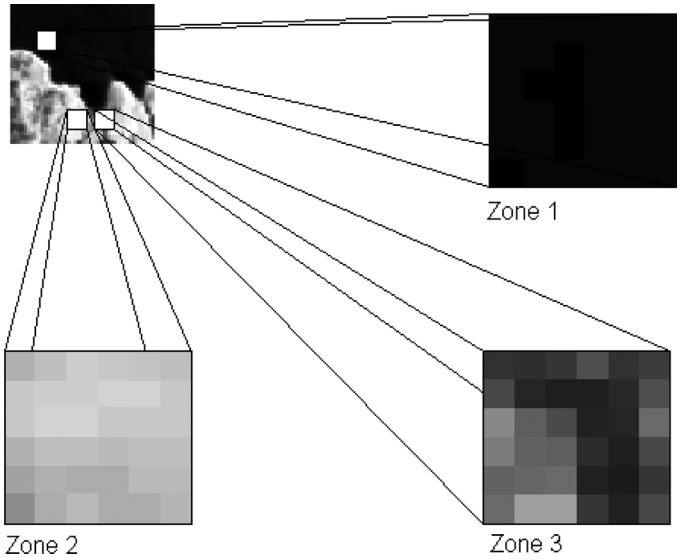

Fig. 15. Areas of water, soil, and vegetation analyzed for the probability of presence.

TABLE IV
PROBABILITY OF PRESENCE FOR EACH ENDMEMBER

	Zone 1 $\eta = 0.98$	Zone 2 $\eta = 0.9$	Zone 3 $\eta = 0.8$
$\mathrm{P}[\alpha_{\text{water}} > \eta \vert \boldsymbol{m}_{\text{water}}]$	0.9922	0	0
$\mathrm{P}[\alpha_{\text{soil}} > \eta \vert \boldsymbol{m}_{\text{soil}}]$	0	0.5147	0.0556
$\mathrm{P}[\alpha_{\text{vegetation}} > \eta \vert \boldsymbol{m}_{\text{veg.}}]$	0	0	0.2774

TABLE V
MMSE ESTIMATE OF $\sigma_r^2 (r = 1, \ldots, R)$

	Soil	Vegetation	Water
MMSE estimates	1×10^{-4}	6.9×10^{-3}	1×10^{-4}

where η is a given threshold. Three distinct zones of 6×6 pixels, depicted in Fig. 15, have been analyzed to estimate these probabilities. The first region (zone 1) has been extracted from the lake area and thus contains a majority of water pixels. Conversely, the other two regions (zones 2 and 3) are coastal areas containing soil and vegetation. Table IV shows the result obtained for different thresholds in each analyzed area.

B. NCM Algorithm With Distinct Endmember Variances

This hyperspectral image has also been analyzed by the algorithm detailed in Section IV to evaluate its performance. As the algorithm requires more than one pixel, the image has been divided into 256 blocks of 3×3 pixels. Thus, the analyzed area[4] has been reduced to 48×48. The estimated variances for the endmembers associated to the block centered around the pixel $\#(35, 43)$ are shown in Table V.

[4]Only the right and bottom edges of the image are not studied, which is a very small area compared to the full size of the image.

VIII. CONCLUSION

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, the endmembers variances and one hyperparameter. These samples were then used to estimate the parameters of interest. The proposed algorithm has several advantages versus the standard LMM-based algorithms. In particular, it allows one to extend the standard model to the case where endmember spectra have different variances. The simulation results on synthetic and real data showed very promising results.

Perspectives include the generalization of the NCM algorithm to more advanced models. For instance, the hyperspectral images could be considered as a set of homogenous regions surrounded by sharp boundaries. In this case, neighborhood conditions for the abundances could be introduced to improve unmixing.

APPENDIX
POSTERIOR DISTRIBUTION $f(\sigma_r^2 | \boldsymbol{\sigma}_{-r}, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{M})$

By using the Bayes' theorem, the posterior distribution $f\left(\sigma_r^2 | \boldsymbol{\sigma}_{-r}, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{M}\right)$ can be written

$$f\left(\sigma_r^2 | \boldsymbol{\sigma}_{-r}, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{M}\right) \propto f(\boldsymbol{Y} | \boldsymbol{A}, \boldsymbol{\sigma}, \boldsymbol{M}) f\left(\sigma_r^2 | \nu, \delta\right) \tag{36}$$

which leads to

$$\begin{aligned} f\left(\sigma_r^2 | \boldsymbol{\sigma}_{-r}, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{M}\right) \propto & \prod_{p=1}^{P} \left(\frac{1}{c(\boldsymbol{\alpha}_p)}\right)^{L/2} \\ & \times \exp\left(-\sum_{p=1}^{P} \frac{\|\boldsymbol{y}_p - \boldsymbol{\mu}(\boldsymbol{\alpha}_p)\|^2}{2c(\boldsymbol{\alpha}_p)}\right) \\ & \times \left(\frac{1}{\sigma_r^2}\right)^{\nu+1} \exp\left(-\frac{\delta}{\sigma_r^2}\right). \end{aligned}$$

This conditional posterior distribution can be rewritten

$$\begin{aligned} & f\left(\sigma_r^2 | \boldsymbol{\sigma}_{-r}, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{M}\right) \\ & \quad \propto \left(\frac{1}{\sigma_r^2}\right)^{\nu+1} \prod_{p=1}^{P} \left(\sigma_r^2 \alpha_{r,p}^2 + c(\boldsymbol{\alpha}_{-r})\right)^{-L/2} \\ & \qquad \times \exp\left(-\sum_{p=1}^{P} \frac{\|\boldsymbol{y}_p - \boldsymbol{\mu}(\boldsymbol{\alpha}_p)\|^2}{2\left(\sigma_r^2 \alpha_{r,p}^2 + c(\boldsymbol{\alpha}_{-r})\right)} - \frac{\delta}{\sigma_r^2}\right). \end{aligned} \tag{37}$$

REFERENCES

[1] N. Keshava and J. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–56, Jan. 2002.
[2] M. E. Winter, "Fast autonomous spectral endmember determination in hyperspectral data," in *Proc. 13th Int. Conf. Appl. Geologic Remote Sens.*, Vancouver, BC, Canada, Apr. 1999, vol. 2, 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, Apr. 2005.

[4] M. T. Eismann and D. Stein, "Stochastic mixture modeling," in *Hyperspectral Data Exploitation: Theory and Applications*, C. Chang, Ed. New York: Wiley, 2007, ch. 5.
[5] J. Settle, "On the relationship between spectral unmixing and subspace projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 4, pp. 1045–1046, Jul. 1996.
[6] D. Manolakis, C. Siracusa, and G. Shaw, "Hyperspectral subpixel target detection using the linear mixing model," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1392–1409, Jul. 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 Process.*, vol. 56, no. 7, pp. 2684–2696, Jul. 2008.
[8] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. New York: Springer Verlag, 2004.
[9] D. Stein, "Application of the normal compositional model to the analaysis of hyperspectral imagery," in *Proc. IEEE Workshop Adv. Tech. for Anal. of Remotely Sens. Data*, Oct. 2003, pp. 44–51.
[10] 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, U.K.: Chapman & Hall, 1996.
[11] S. Moussaoui, H. Hauksdóttir, F. Schmidt, C. Jutten, J. Chanussot, D. Brie, S. Douté, and J. A. Benediksson, "On the decomposition of mars hyperspectral data by ICA and Bayesian positive source separation," *Neurocomputing*, vol. 71, no. 10–12, pp. 2194–2208, Jun. 2008.
[12] N. Bali and A. Mohammad-Djafari, "Bayesian approach with hidden Markov modeling and mean field approximation for hyperspectral data analysis," *IEEE Trans. Image Process.*, vol. 17, no. 2, pp. 217–225, Feb. 2008.
[13] H. Snoussi, "Efficient Bayesian spectral matching separation in noisy mixtures," *IEEE Trans. Image Process.*, to be published.
[14] E. Punskaya, C. Andrieu, A. Doucet, and W. Fitzgerald, "Bayesian curve fitting using MCMC with applications to signal segmentation," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 747–758, Mar. 2002.
[15] N. Dobigeon, J.-Y. Tourneret, and M. Davy, "Joint segmentation of piecewise constant autoregressive processes by using a hierarchical model and a Bayesian sampling approach," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1251–1263, Apr. 2007.
[16] "ENVI User's Guide Version 4.0," RSI (Research Systems Inc), Boulder, CO, 2003.
[17] R. O. Green *et al.*, "Imaging spectroscopy and the airborne visible/ infrared imaging spectrometer (AVIRIS)," *Remote Sens. Environ.*, vol. 65, no. 3, pp. 227–248, Sep. 1998.
[18] 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. Remote Sens.*, vol. 39, no. 3, pp. 529–545, Mar. 2001.
[19] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 765–776, Mar. 2007.
[20] M. D. Plumbley and E. Oja, "A nonnegative-PCA algorithm for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 15, no. 1, pp. 66–76, Jan. 2004.
[21] J. W. Boardman, F. Kruse, and R. O. Green, "Mapping target signatures via partial unmixing of AVIRIS data," in *Summaries of Fifth Annual JPL Airborne Earth Science Workshop*. Pasadena, CA: JPL, 1995, pp. 23–26.
[22] O. Eches, N. Dobigeon, C. Mailhes, and J. Y. Tourneret, "Unmixing hyperspectral images using a normal compositional model and MCMC methods," Univ. Toulouse, Toulouse, France, Mar. 2009 [Online]. Available: http://eches.perso.enseeiht.fr

Olivier Eches was born in Villefranche-de-Rouergue, France, in 1984. He received the Eng. degree in electrical engineering from ENSEEIHT, Toulouse, France, and the M.Sc. degree in signal processing from the National Polytechnic Institute of Toulouse, both in June 2007. He is currently pursuing the Ph.D. degree at the University of Toulouse (IRIT/INP-ENSEEIHT) on the study of Bayesian algorithms and MCMC methods for the analysis of hyperspectral images.

Nicolas Dobigeon (S'05–M'08) was born in Angoulême, France, in 1981. He received the Eng. degree in electrical engineering from ENSEEIHT, Toulouse, France, and the M.Sc. degree in signal processing from the National Polytechnic Institute of Toulouse, both in 2004, and the Ph.D. degree in signal processing from the National Polytechnic Institute of Toulouse in 2007.

From 2007 to 2008, he was a Postdoctoral Research Associate in the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor. Since 2008, he has been an Assistant Professor with the National Polytechnic Institute of Toulouse (ENSEEIHT—University of Toulouse), within the Signal and Communication Group of the IRIT Laboratory. His research interests are centered around statistical signal and image processing with a particular interest to Bayesian inference and Markov chain Monte Carlo (MCMC) methods.

Corinne Mailhes (M'87) was born in France in 1965. She received the Eng. degree in electronics and signal processing and the Ph.D. degree in signal processing from the University of Toulouse (ENSEEIHT), Toulouse, France, in 1986 and 1990, respectively.

She is currently a Professor with the University of Toulouse (ENSEEIHT) and a member of the IRIT Laboratory (UMR 5505 of the CNRS) and of TeSA Lab (http://www.tesa.prd.fr). Her research activities are centered on statistical signal processing, with particular interests in spectral analysis, data compression, and biomedical signal processing.

Prof. Mailhes was member of the Organizing Committee for the International Conference ICASSP'06, held in Toulouse in 2006.

Jean-Yves Tourneret (M'94–SM'08) received the ingénieur degree in electrical engineering from École Nationale Supérieure d'Électronique, d'Électrotechnique, d'Informatique et d'Hydraulique in Toulouse (ENSEEIHT), France, in 1989 and the Ph.D. degree from the National Polytechnic Institute from Toulouse in 1992.

He is currently a Professor at the University of Toulouse (ENSEEIHT) and a member of the IRIT Laboratory (UMR 5505 of the CNRS). His research activities are centered around statistical signal processing with a particular interest to classification and Markov Chain Monte Carlo methods.

Dr. Tourneret was the program chair of the European conference on signal processing (EUSIPCO), which was held in Toulouse in 2002. He was also member of the organizing committee for the international conference ICASSP'06 which was held in Toulouse in 2006. He has been a member of different technical committees including the Signal Processing Theory and Methods (SPTM) committee of the IEEE Signal Processing Society (2001–2007, 2010–present). He is currently serving as an associate editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING.