Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery

Nicolas Dobigeon[†], Jean-Yves Tourneret[†] and Chein-I Chang[*]

E-mail : {Nicolas.Dobigeon, Jean-Yves.Tourneret}@enseeiht.fr, cchang@umbc.edu

TECHNICAL REPORT – 2007, March

[†]IRIT/ENSEEIHT/TéSA, 2 rue Camichel, BP 7122, 31071 Toulouse cedex 7, France

[*]University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA

Abstract

This paper proposes a hierarchical Bayesian model that can be used for semi-supervised hyperspectral image unmixing. The model assumes that the pixel reflectances result from linear combinations of pure component spectra contaminated by an additive Gaussian noise. The abundance parameters appearing in this model satisfy positivity and additivity constraints. These constraints are naturally expressed in a Bayesian context by using appropriate abundance prior distributions. The posterior distributions of the unknown model parameters are then derived. A Gibbs sampler allows one to draw samples distributed according to the posteriors of interest and to estimate the unknown abundances. An extension of the algorithm is finally studied for mixtures with unknown numbers of spectral components belonging to a know library. The performance of the different unmixing strategies is evaluated via simulations conducted on synthetic and real data.

Index Terms

Hyperspectral images, linear spectral unmixing, hierarchical Bayesian analysis, MCMC methods, Gibbs sampler, reversible jumps.

I. Introduction

Spectral unmixing has been widely used in remote sensing signal processing for data analysis [1]. Its underlying assumption is based on the fact that all data sample vectors are mixed by a number of so-called endmembers assumed to be present in the data. By virtue of this assumption, two models have been investigated in the past to model how mixing activities take place. One is the macrospectral mixture that describes a mixed pixel as a linear mixture of endmembers opposed to the other model suggested by Hapke [2], referred to as intimate mixture that models a mixed pixel as a nonlinear mixture. Nonetheless, it has been shown in [3] that the intimate model could be linearized to simplify analysis. Accordingly, only linear spectral unmixing is considered in this paper. In order for linear spectral unmixing to be effective, three key issues must be addressed. One is the number of endmembers assumed to be in the data for linear mixing. Another is how to estimate these endmembers once the number of endmembers is determined. The third issue is algorithms designed for linear unmixing (also referred to as inversion algorithms). While much work in linear spectral unmixing is devoted to the third issue, the first and second issues have been largely ignored or avoided by assuming availability of prior knowledge. Therefore, most linear unmixing techniques currently being developed in the literature are supervised, that is the knowledge of endmembers is assumed to be given a priori. This paper considers a semi-supervised linear spectral unmixing approach which determines how endmembers from a given spectral library should be present in the data and uses the desired endmembers for linear spectral unmixing. In some real applications, the endmembers must be obtained directly from the data itself without prior knowledge. In this case, the proposed algorithm has to be combined with an endmember extraction algorithm such as the well-known N-finder algorithm (N-FINDR) developed by Winter [4] to find desired endmembers which will be used to form a base of the linear mixing model (LMM).

As explained above, the inversion step of an unmixing algorithm has already received much attention in the literature (see for example [1] and references therein). The LMM is classically used to model the spectrum of a pixel in the observed scene. This model assumes that the spectrum of a given pixel is related to endmember spectra via linear relations whose coefficients are referred to as abundance coefficients or abundances. The inversion problem then reduces to estimate the abundances from the observed pixel spectrum. The abundances satisfy the constraints

of non-negativity and full additivity. Consequently, their estimation requires to use a quadratic programming algorithm with linear equalities and inequalities as constraints. Different estimators including constrained Least squares and minimum variance estimators were developed by using these ideas [5], [6]. This paper studies a hierarchical Bayesian estimator which allows one to estimate the abundances in an LMM. The proposed algorithm defines appropriate prior distributions for the unknown signal parameters (here the abundance coefficients and the noise variance) and estimates these unknown parameters from their posterior distributions.

The complexity of the posterior distributions for the unknown parameters requires to use appropriate simulation methods such as Markov Chain Monte Carlo (MCMC) methods [7]. The prior distributions used in the present paper depend on hyper-parameters which have to be determined. There are mainly two approaches which can be used to estimate these hyperparameters. The first approach couples MCMCs with an expectation maximization (EM) algorithm which allows one to estimate the unknown hyperparameters [8]. However, as explained in [9, p. 259], the EM algorithm suffers from the initialization issue and can converge to local maxima or saddle points of the log-likelihood function. The second approach defines non-informative prior distributions for the hyperparameters introducing a second level of hierarchy within the Bayesian paradigm. The hyperparameters are then integrated out from the joint posterior distribution or estimated from the observed data [10]–[13]. This second strategy results in a hierarchical Bayesian estimator which will show interesting properties for unmixing hyperspectral images. Another advantage of the hierarchical Bayesian estimator is that it allows one to estimate the full posterior distribution of the unknown parameters and hyperparameters. As a result, these posterior distributions can be used to derive confidence intervals for the unknown parameters, providing information on the significance of the estimations.

The proposed spectral unmixing problem is formulated as a constrained linear regression problem. Bayesian models are particularly appropriate for these problems since the constraints can be included in the prior distribution. The support of the posterior then reduces to the constrained parameter space. Examples of constraints recently studied in the literature include monotone constraints and positivity constraints. Monotony can be handled efficiently by using truncated Gaussian priors [14] whereas positivity constraints are satisfied when choosing Gamma priors [15] or truncated Gaussian priors [16]. It is interesting to mention here that similar ideas have been recently exploited to handle linear sparse approximation models. For instance, sparsity

can be ensured by defining factoring mixture priors [17] or Student priors [18]. This paper defines a Bayesian model with priors satisfying positivity and additivity constraints as required in hyperspectral imagery. To our knowledge, this is the first time a Bayesian model based on these constraints is proposed in the literature. The parameters of this model are estimated by an appropriate Gibbs sampler. Interestingly, the proposed sampler can handle mixtures with unknown numbers of spectral components belonging to a known library.

The paper is organized as follows. Section II presents the usual LMM for hyperspectral images. Section III describes the different elements of the proposed hierarchical model for unmixing these hyperspectral images. Section IV studies a Gibbs sampler which allows one to generate samples distributed according to the posteriors of the unknown parameters to be estimated. The sampler convergence is investigated in Section V. Some simulation results on synthetic and real data are presented in Section VI and VII. Section VIII shows that the number of endmembers contained in the mixing model can be estimated by including a reversible jump MCMC algorithm. Conclusions are reported in Section IX.

II. Linear Mixing Model

This section defines the classical analytical model which will be used to perform spectral unmixing. This paper concentrates on the most commonly used linear unmixing problem which constitutes a good approximation in the reflective domain ranging from $0.4\mu m$ to $2.5\mu m$ (see [1], [19] or more recently [20]). However, the proposed analysis might be extended to nonlinear unmixing models, for instance, by using a basis function representation approach as in [21, p. 134]. The LMM assumes that the L-spectrum $\mathbf{y} = [y_1, \ldots, y_L]^{\mathsf{T}}$ of a mixed pixel is a linear combination of R spectra $\mathbf{m}_r$ contaminated by additive white noise:

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

where

- $\mathbf{m}_r = [m_{r,1}, \ldots, m_{r,L}]^{\mathsf{T}}$ denotes the spectrum of the r^{th} material,
- α_r is the fraction of the r^{th} material in the pixel,
- R is the number of pure materials (or *endmembers*) present in all the observed scene,
- L is the number of available spectral bands for the image,

- $\mathbf{n} = [n_1, \ldots, n_L]^\mathsf{T}$ is the additive white noise sequence which is classically assumed to be an independent and identically distributed (i.i.d.) zero-mean Gaussian sequence with variance σ^2, denoted as $\mathbf{n} \sim \mathcal{N}(\mathbf{0}_L, \sigma^2 \mathbf{I}_L)$, where $\mathbf{I}_L$ is the identity matrix of dimension $L \times L$.

Due to physical considerations, the fraction vector $\boldsymbol{\alpha}^+ = [\alpha_1, \ldots, \alpha_R]^\mathsf{T}$ satisfies the following positivity and additivity constraints:

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

The R endmembers spectra $\mathbf{m}_r$ are assumed to be known in the first part of this paper. As a consequence, the proposed methodology has to be coupled with one of the many identification techniques to estimate these endmember spectra. These techniques include geometrical methods [4], [22] or statistical procedures [23], [24]. The second part of the paper extends the algorithm to mixtures containing an unknown number of spectra belonging to a known library.

III. Hierarchical Bayesian model

This section introduces a hierarchical Bayesian model to estimate the unknown parameter vector $\boldsymbol{\alpha}^+$ under the constraints specified in Eq. (2). This model is based on the likelihood of the observations and on prior distributions for the unknown parameters.

A. Likelihood

Eq. (1) shows that $\mathbf{y} \sim \mathcal{N}(\mathbf{M}^+\boldsymbol{\alpha}^+, \sigma^2\mathbf{I}_L)$, where $\mathbf{M}^+ = [\mathbf{m}_1, \ldots, \mathbf{m}_R]$ and $\boldsymbol{\alpha}^+ = [\alpha_1, \ldots, \alpha_R]^\mathsf{T}$. Consequently, the likelihood function of $\mathbf{y}$ can be expressed as:

$$f(\mathbf{y}|\boldsymbol{\alpha}^+, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{L}{2}} \exp\left[-\frac{\|\mathbf{y} - \mathbf{M}^+\boldsymbol{\alpha}^+\|^2}{2\sigma^2}\right], \tag{3}$$

where $\|\boldsymbol{x}\|^2 = \boldsymbol{x}^\mathsf{T}\boldsymbol{x}$ is the standard L^2 norm.

B. Parameter priors

The abundance vector can be written as $\boldsymbol{\alpha}^+ = \left[\boldsymbol{\alpha}^\mathsf{T}, \alpha_R\right]^\mathsf{T}$ with $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_{R-1}]^\mathsf{T}$ and $\alpha_R = 1 - \sum_{r=1}^{R-1} \alpha_r$. The LMM constraints (2) impose that $\boldsymbol{\alpha}$ belongs to the simplex $\mathbb{S}$ defined by:

$$\mathbb{S} = \left\{ \boldsymbol{\alpha} \,\middle|\, \alpha_r \geq 0, \ \forall r = 1, \ldots, R-1, \ \sum_{r=1}^{R-1} \alpha_r \leq 1 \right\}. \tag{4}$$

A uniform distribution on $\mathbb{S}$ is chosen for $\boldsymbol{\alpha}$ in order to reflect the absence of prior knowledge regarding this unknown parameter vector. Note that choosing this prior distribution for $\boldsymbol{\alpha}$ is equivalent to choosing a prior Dirichlet distribution $\mathcal{D}_R(1,\ldots,1)$ for $\boldsymbol{\alpha}^+$ (see [21, p. 237] for the definition of the Dirichlet distribution $\mathcal{D}_R(1,\ldots,1)$).

A non-informative conjugate prior is chosen for σ^2, i.e. an inverse-gamma distribution with parameters $\frac{\nu}{2}$ and $\frac{\gamma}{2}$:

$$\sigma^2 \sim \mathcal{IG}\left(\frac{\nu}{2}, \frac{\gamma}{2}\right). \tag{5}$$

The hyperparameter ν will be fixed to $\nu = 2$ (as in [12]) whereas γ is an adjustable hyperparameter.

C. Hyperparameter prior

The hyperparameter associated to the parameter priors defined above is γ. Of course, the quality of the unmixing procedure depends on the value of this hyperparameter. The hierarchical Bayesian approach developed in this paper uses a non-informative Jeffrey's prior for the hyperparameter γ:

$$f(\gamma) = \frac{1}{\gamma}\mathbf{1}_{\mathbb{R}^+}(\gamma), \tag{6}$$

where $\mathbf{1}_{\mathbb{R}^+}(\cdot)$ is the indicator function defined on $\mathbb{R}^+$.

D. Posterior distribution of $\boldsymbol{\theta}$

The posterior distribution of the unknown parameter vector $\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \sigma^2\}$ can be computed from the following hierarchical structure:

$$f(\boldsymbol{\theta}|\mathbf{y}) \propto \int f(\mathbf{y}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\gamma) f(\gamma) d\gamma, \tag{7}$$

where $\propto$ means "proportional to" and $f(\mathbf{y}|\boldsymbol{\theta})$ and $f(\gamma)$ are defined in (3) and (6) respectively. By assuming the independence between σ^2 and $\boldsymbol{\alpha}$, i.e. $f(\boldsymbol{\theta}|\gamma) = f(\boldsymbol{\alpha}) f(\sigma^2|\nu,\gamma)$, the hyperparameter γ can be integrated out from the joint distribution $f(\boldsymbol{\theta},\gamma|\mathbf{y})$, yielding:

$$f\left(\boldsymbol{\alpha},\sigma^2|\mathbf{y}\right) \propto \frac{1}{\sigma^{L+2}} \exp\left[-\frac{\|\mathbf{y}-\mathbf{M}^+\boldsymbol{\alpha}^+\|^2}{2\sigma^2}\right] \mathbf{1}_{\mathbb{S}}(\boldsymbol{\alpha}), \tag{8}$$

where $\mathbf{1}_{\mathbb{S}}(\cdot)$ is the indicator function defined on the simplex $\mathbb{S}$. The next section shows that an appropriate Gibbs sampling strategy allows one to generate samples distributed according to the joint distribution $f(\boldsymbol{\alpha},\sigma^2|\mathbf{y})$.

ALGORITHM 1:

Gibbs sampling algorithm for hyperspectral image unmixing

- Initialization:
  - Sample parameters $\sigma^{2(0)}$ and $\boldsymbol{\alpha}^{(0)}$,
  - Set $t \leftarrow 1$,
- Iterations: for $t = 1, 2, \ldots,$ do
  - Sample $\boldsymbol{\alpha}^{(t)}$ from the pdf in (11),
  - Sample $\sigma^{2(t)}$ from the pdf in (12),
  - Set $t \leftarrow t + 1$.

IV. A GIBBS SAMPLER FOR ABUNDANCE ESTIMATION

Sampling according to $f(\boldsymbol{\alpha}, \sigma^2|\mathbf{y})$ can be achieved by a Gibbs sampler whose steps are detailed in Subsections IV-A and IV-B (see also Algorithm 1).

A. Generation of samples distributed according to $f(\boldsymbol{\alpha}|\sigma^2, \mathbf{y})$

By denoting $\mathbf{M} = [\mathbf{m}_1, \ldots \mathbf{m}_{R-1}]$, straightforward computations detailed in Appendix II yield:

$$f\left(\boldsymbol{\alpha}|\sigma^2, \mathbf{y}\right) \propto \exp\left[-\frac{(\boldsymbol{\alpha}-\boldsymbol{\mu})^{\mathsf{T}} \boldsymbol{\Lambda}^{-1} (\boldsymbol{\alpha}-\boldsymbol{\mu})}{2}\right] \mathbf{1}_{\mathbb{S}}(\boldsymbol{\alpha}), \tag{9}$$

where

$$\left\{ \begin{array}{ll} \boldsymbol{\Lambda} & = \left[\frac{1}{\sigma^2}\left(\mathbf{M} - \mathbf{m}_R\mathbf{u}^{\mathsf{T}}\right)^{\mathsf{T}}\left(\mathbf{M} - \mathbf{m}_R\mathbf{u}^{\mathsf{T}}\right)\right]^{-1}, \\ \boldsymbol{\mu} & = \boldsymbol{\Lambda}\left[\frac{1}{\sigma^2}\left(\mathbf{M} - \mathbf{m}_R\mathbf{u}^{\mathsf{T}}\right)^{\mathsf{T}}(\mathbf{y} - \mathbf{m}_R)\right], \end{array} \right. \tag{10}$$

with $\mathbf{u} = [1, \ldots, 1]^{\mathsf{T}} \in \mathbb{R}^{R-1}$. As a consequence, $\boldsymbol{\alpha}|\sigma^2, \mathbf{y}$ is distributed according to a truncated Gaussian distribution (defined in Appendix I):

$$\boldsymbol{\alpha}|\sigma^2, \mathbf{y} \sim \mathcal{N}_{\mathbb{S}}\left(\boldsymbol{\mu}, \boldsymbol{\Lambda}\right). \tag{11}$$

The generation of samples according to a truncated Gaussian distribution can be achieved using a standard accept-reject procedure described in Algorithm 2, when the number of endmembers is relatively small (as in the examples studied in this paper). However, it is interesting to mention here that a more efficient simulation technique based on Gibbs moves can be used for high dimension problems (see [25] or [26] for more details).

ALGORITHM 2:

Generation according to a truncated normal distribution

1. sample $\boldsymbol{\alpha}^* \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Lambda})$
2. accept/reject test:
   2.1. if $\boldsymbol{\alpha}^* \in \mathbb{S}$, set $\boldsymbol{\alpha} = \boldsymbol{\alpha}^*$ (*accept*),
   2.2. if $\boldsymbol{\alpha}^* \notin \mathbb{S}$, go to step 1. (*reject*),
3. set $\boldsymbol{\alpha}^+ = \left[\boldsymbol{\alpha}^\mathsf{T}, 1 - \sum_{r=1}^{R-1} \alpha_r\right]^\mathsf{T}$.

B. Generation of samples distributed according to $f(\sigma^2|\boldsymbol{\alpha}, \mathbf{y})$

Looking carefully at the joint distribution $f(\sigma^2, \boldsymbol{\alpha}|\mathbf{y})$, the conditional distribution of $\sigma^2|\boldsymbol{\alpha}, \mathbf{y}$ is clearly the following inverse gamma distribution:

$$\sigma^2|\boldsymbol{\alpha}, \mathbf{y} \sim \mathcal{IG}\left(\frac{L}{2}, \frac{\|\mathbf{y} - \mathbf{M}^+\boldsymbol{\alpha}^+\|^2}{2}\right). \tag{12}$$

V. CONVERGENCE DIAGNOSIS

The Gibbs sampler allows to draw samples $\left(\boldsymbol{\alpha}^{(t)}, \sigma^{2(t)}\right)$ asymptotically distributed according to $f(\boldsymbol{\alpha}, \sigma^2|\mathbf{y})$. The abundance vector can then be estimated by the empirical average according to the minimum mean square error (MMSE) principle:

$$\hat{\boldsymbol{\alpha}}_{\mathrm{MMSE}} = \frac{1}{N_r}\sum_{t=1}^{N_r} \boldsymbol{\alpha}^{(N_{\mathrm{bi}}+t)}, \tag{13}$$

where N_{bi} and N_r are the numbers of burn-in and computation iterations, respectively. However, two important questions have to be addressed: 1) When can we decide that the samples $\left(\boldsymbol{\alpha}^{(t)}, \sigma^{2(t)}\right)$ are actually distributed according to the target distribution $f(\boldsymbol{\alpha}, \sigma^2|\mathbf{y})$? 2) How many samples are necessary to obtain an accurate estimate of $\boldsymbol{\alpha}$ when using Eq. (13)? This section surveys some works allowing to determine appropriate values for parameters N_{r} and N_{bi}.

A. Determination of the burn-in period N_{bi}

Running multiple chains with different initializations allows to define various convergence measures for MCMC methods [27]. The popular between-within variance criterion has shown

interesting properties for diagnosing convergence of MCMC methods. This criterion was initially studied by Gelman and Rubin in [28] and has been used in many studies including [27, p. 33], [29], [30]. The main idea is to run M parallel chains of length N_{r} for each data set with different starting values and to evaluate the dispersion of the estimates obtained from the different chains. The between-sequence variance B and within-sequence variance W for the M Markov chains are defined by

$$B = \frac{N_{\mathrm{r}}}{M-1} \sum_{m=1}^{M} \left(\overline{\kappa}_m - \overline{\kappa}\right)^2, \tag{14}$$

and

$$W = \frac{1}{M} \sum_{m=1}^{M} \frac{1}{N_{\mathrm{r}}} \sum_{t=1}^{N_{\mathrm{r}}} \left(\kappa_m^{(t)} - \overline{\kappa}_m\right)^2, \tag{15}$$

with

$$\begin{cases} \overline{\kappa}_m = \frac{1}{N_{\mathrm{r}}} \sum\limits_{t=1}^{N_{\mathrm{r}}} \kappa_m^{(t)}, \\ \overline{\kappa} = \frac{1}{M} \sum\limits_{m=1}^{M} \overline{\kappa}_m, \end{cases} \tag{16}$$

where κ is the parameter of interest and $\kappa_m^{(t)}$ is its estimate at the t^{th} run of the m^{th} chain. The convergence of the chain can then be monitored by the so-called *potential scale reduction factor* $\hat{\rho}$ defined as [31, p. 332]:

$$\sqrt{\hat{\rho}} = \sqrt{\frac{1}{W}\left(\frac{N_r - 1}{N_r} W + \frac{1}{N_r} B\right)}. \tag{17}$$

A value of $\sqrt{\hat{\rho}}$ close to 1 indicates a good convergence of the sampler. In other words, a value of $\sqrt{\hat{\rho}}$ close to 1 shows that the number of burn-in iterations N_{bi} is sufficient to obtain samples $\left(\boldsymbol{\alpha}^{(N_{\mathrm{bi}}+t)}, \sigma^{2(N_{\mathrm{bi}}+t)}\right)$, $t = 1, \ldots, N_r$, distributed according to the target distribution.

B. Determination of the number of computation iterations N_r

Once the number of burn-in N_{bi} iterations has been adjusted, it is important to determine the appropriate number of iterations N_r to obtain an accurate estimate of $\boldsymbol{\alpha}$ when using Eq. (13). An *ad hoc* approach consists of assessing convergence via appropriate graphical evaluations [27, p. 28]. This paper proposes to compute a reference estimate denoted as $\widetilde{\boldsymbol{\alpha}}$ from a large number of iterations to ensure convergence of the sampler and good accuracy of the approximation in

Eq. (13). The mean square error (MSE) between this reference estimate $\widetilde{\boldsymbol{\alpha}}$ and the estimate obtained after p iterations is then computed as follows:

$$e_r^2(p) = \left\| \widetilde{\boldsymbol{\alpha}} - \frac{1}{p} \sum_{t=1}^{p} \boldsymbol{\alpha}^{(N_{\mathrm{bi}}+t)} \right\|^2 .$$

The number of iterations N_r is finally determined as the value of p ensuring the MSE $e_r^2(p)$ is below a predefined threshold.

VI. Simulation Results on Synthetic Data

A. Abundance Estimation

The accuracy of the proposed abundance estimation procedure is first illustrated by unmixing a synthetic pixel resulting from the combination of three pure components. These components have been extracted from the spectral libraries that are distributed with the ENVI software [32, p. 1035] and are representative of a urban or suburban environment: construction concrete, green grass and dark yellowish brown micaceous loam. The proportions of these components are defined by $\alpha_1 = 0.3$, $\alpha_2 = 0.6$ and $\alpha_3 = 0.1$. The observations have been corrupted by an additive Gaussian noise with variance $\sigma^2 = 0.025$ (i.e., the signal to noise ratio is about SNR $= 15$dB). The endmember spectra and the resulting noisy spectrum of the mixed pixel are plotted in Fig. 1.

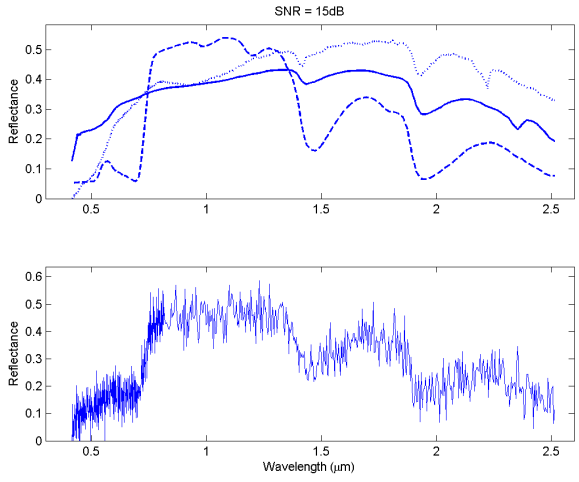

Fig. 1. Top: endmember spectra: construction concrete (solid line), green grass (dashed line), dark yellowish brown micaceous loam (dotted line). Bottom: resulting spectrum of the mixed pixel.

Fig. 2 shows the posterior distributions of the abundance coefficients α_r ($r = 1, 2, 3$) obtained for $N_{\mathrm{MC}} = 20000$ iterations (including $N_{\mathrm{bi}} = 100$ burn-in iterations). These distributions are in good agreement with the actual values of abundances, i.e. $\boldsymbol{\alpha}^+ = [0.3, 0.6, 0.1]^{\mathsf{T}}$. For comparison, the fully constrained least-squares (FCLS) algorithm detailed in [5], [33] has been run N_{MC} times for signals similar to Fig. 1 (bottom) obtained with different noise sequences. The histograms of the N_{MC} FCLS abundance estimates are depicted in Fig. 2 (dotted lines). These histograms are clearly in good agreement with the corresponding posterior distributions obtained from the proposed hierarchical Bayesian algorithm. However, it is important to point out that the abundance posteriors shown in Fig. 2 (continuous lines) have been obtained from a given pixel spectrum, whereas the FLCS algorithm has to be run N_{MC} times to compute the abundance histograms.

Fig. 3 shows the abundance MAP estimates of α_r and the corresponding standard-deviations (computed from the proposed Bayesian algorithm) as a function of the signal-to-noise ratio (SNR). These figures allow us to evaluate the estimation performance for a given SNR. Note that the SNRs of the actual spectrometers like AVIRIS are not below 30dB when the water absorption bands have been removed [34]. As a consequence, the results on Fig. 3 indicate that the proposed Bayesian algorithm performs satisfactorily for these SNRs.

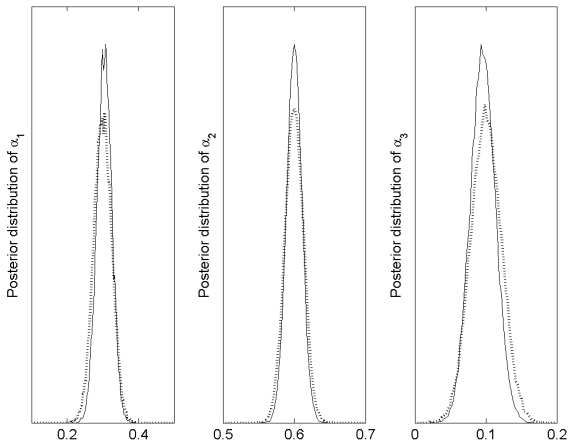

Fig. 2. Posterior distributions of the estimated abundances $[\alpha_1, \alpha_2, \alpha_3]^{\mathsf{T}}$ (continuous lines) and histograms of FCLS estimates (dotted lines).

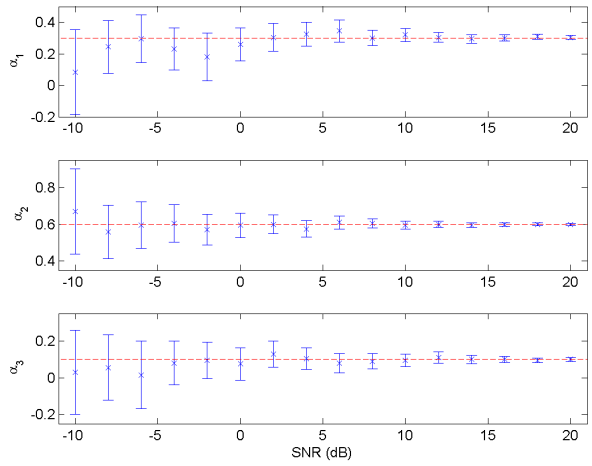

Fig. 3. MAP estimates (cross) and standard deviations (vertical bars) of α_r ($r = 1, \ldots, 3$) versus SNR.

B. Acceptance rate of the sampler

The computational efficiency of the proposed Gibbs sampler is governed by the acceptation rate of the accept-reject procedure for simulating according to a truncated Gaussian distribution. The probability of accepting a sample distributed according to a truncated Gaussian distribution is denoted $\mathrm{P}[\boldsymbol{\alpha} \in \mathbb{S}]$, where $\boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ and $\boldsymbol{\mu}$ and $\boldsymbol{\Lambda}$ have been defined in (10). Straightforward computations allow us to obtain the following results:

$$\begin{aligned} \mathrm{P}[\boldsymbol{\alpha} \in \mathbb{S}] &= \int_{\mathbb{S}} \phi\left(\boldsymbol{\alpha} \middle| \boldsymbol{\mu}, \boldsymbol{\Lambda}\right) d\boldsymbol{\alpha} \\ &= \int_0^1 \int_0^{1-\alpha_1} \int_0^{1-\alpha_1-\alpha_2} \\ &\quad \ldots \int_0^{1-\sum_{r=1}^{R-2}\alpha_r} \phi\left(\boldsymbol{\alpha} \middle| \boldsymbol{\mu}, \boldsymbol{\Lambda}\right) d\alpha_{R-1} d\alpha_{R-2} \ldots d\alpha_1, \end{aligned} \tag{18}$$

where ϕ is the Gaussian probability density function (pdf) defined in Section III-B. Fig. 4 displays the theoretical acceptation rate $\mathrm{P}[\boldsymbol{\alpha} \in \mathbb{S}]$ which is compared with the experimental one computed from the generation of 5000 Gaussian variables. These results have been obtained for a given value of $\boldsymbol{\alpha} = [0.3, 0.6, 0.1]^{\mathsf{T}}$ as a function of the SNR. However, these results do not change significantly for other values of $\boldsymbol{\alpha}$. Fig. 4 shows that the acceptation rate $\mathrm{P}[\boldsymbol{\alpha} \in \mathbb{S}]$ is an increasing function of SNR, as expected. It also shows that the acceptation rate is very satisfactory for typical SNRs encountered in hyperspectral imagery (SNR > 30dB). It is interesting to mention

here that we didn't experience any problem in our simulations regarding the time required for simulating according to the truncated Gaussian distribution, since the number of endmembers present in the image is relatively small. However, in higher dimensions or for smaller SNRs, the accept/reject algorithm can be relatively inefficient. In such case, an appropriate Gibbs sampler can be used for simulating according to the truncated Gaussian distribution (see [25] and [26] for more details).

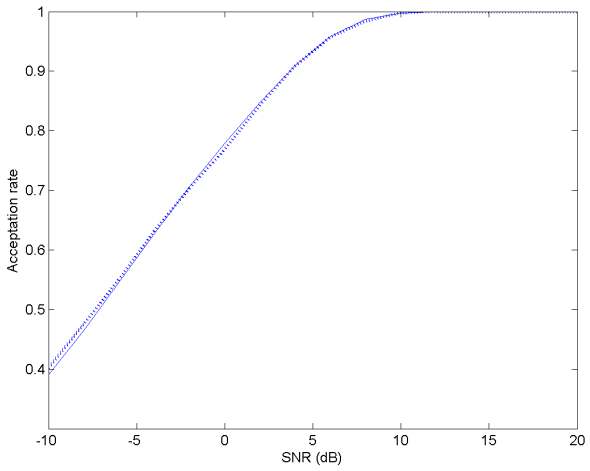

Fig. 4. Theoretical (solid) and experimental (dotted) acceptation rates of the accept-reject test versus SNR.

C. Sampler convergence

The sampler convergence is monitored by computing the potential scale reduction factor defined in Eq. (17). Different choices for parameter κ could be considered for the proposed unmixing procedure. This paper proposes to monitor the convergence of the Gibbs sampler by checking the noise variance σ^2 (see [29] for a similar choice). As an example, the outputs of 5 chains for parameter σ^2 are depicted in Fig. 5. The chains clearly converge to a similar value that is in agreement with the actual variance noise $\sigma^2 = 0.025$. The potential scale reduction factor for parameter σ^2 computed from $M = 10$ Markov chains is equal to 0.9996. This value of $\sqrt{\hat{\rho}}$ confirms the good convergence of the sampler (a recommendation for convergence assessment is a value of $\sqrt{\hat{\rho}} \leq 1.2$ [31, p. 332]).

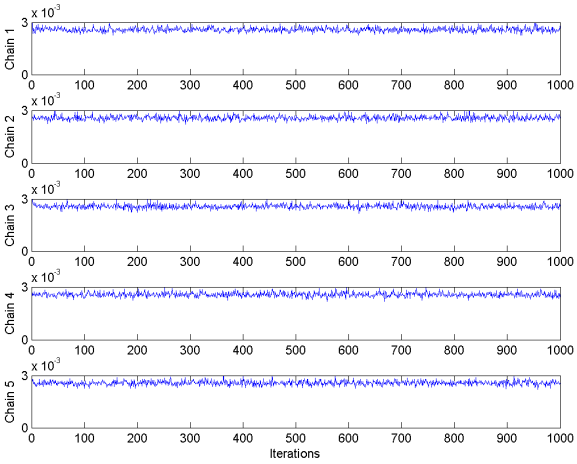

Fig. 5. Convergence assessment with five realizations of the Markov chain.

The number of iterations N_r necessary to compute an accurate estimate of $\boldsymbol{\alpha}$ according to the MMSE principle in Eq. (13) is determined by monitoring the MSE between a reference estimate $\widetilde{\boldsymbol{\alpha}}$ (obtained with $N_r = 10000$) and the estimate obtained after $N_r = p$ iterations. Fig. 6 shows this MSE as a function of the number of iterations p (the number of burn-in iterations is $N_{\text{bi}} = 100$). This figure indicates that a number of iterations equal to $N_r = 500$ is sufficient to ensure an accurate estimation of the empirical average in Eq. (13) for this example. Note that, for such values of N_r and N_{bi}, unmixing this pixel takes approximatively 0.3 seconds for a MATLAB implementation on a 2.8 GHz Pentium IV.

VII. Spectral unmixing of an AVIRIS image

To evaluate the performance of the proposed algorithm for actual data, this section presents the analysis of an hyperspectral image that has received much attention in the remote sensing and image processing communities [35]–[38]. The image depicted in Fig. 7 has 224 spectral bands, a nominal bandwidth of $10nm$, and was acquired in 1997 by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) over Moffett Field, at the southern end of the San Francisco Bay, California (see [39] for more details). It consists of a large water point (a part of a lake that appears in dark pixel at the top of the image) and a coastal area composed of vegetation and soil.

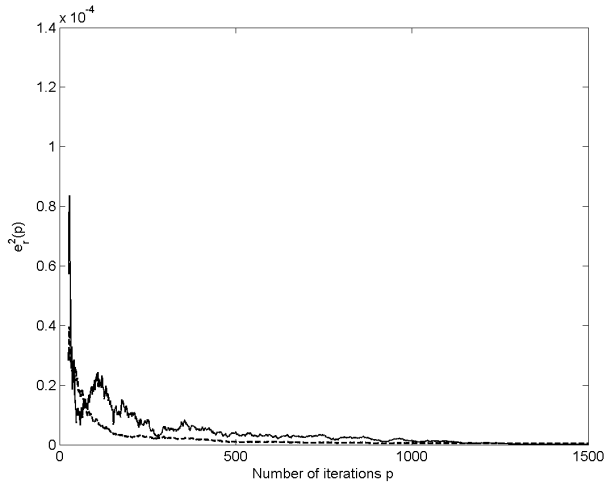

Fig. 6. MSE between the reference and estimated *a posteriori* change-point probabilities versus p (solid line). Averaged MSE computed from 10 chains (dashed line) ($N_{\mathrm{bi}} = 100$).

The data set has been reduced from the original 224 bands to $L = 189$ bands by removing water absorption bands. A 50×50 part of the image represented in gray scale at wavelength $\lambda = 0.66\mu m$ (band 30) has been processed by the proposed unmixing algorithm.

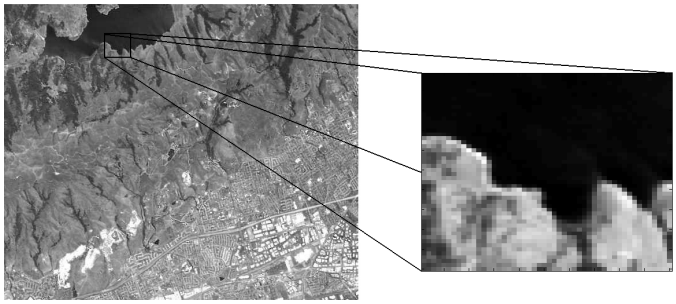

Fig. 7. Real hyperspectral data: Moffett Field acquired by AVIRIS in 1997 (left) and the region of interest at wavelength $\lambda = 0.66\mu m$ shown in gray scale (right).

A. Endmember determination

The first step of the analysis identifies the pure materials that are present in the scene. Note that a preliminary knowledge of the ground geology would allow us to use a supervised method for endmember extraction (e. g. by averaging the pixel spectra on appropriate regions of interest).

Such data being not available, a fully automatic procedure has been implemented. This procedure includes a principal component analysis (PCA) which allows one to reduce the dimensionality of the data and to know the number of endmembers present in the scene as explained in [1]. After computing the cumulative normalized eigenvalues, the data have been projected on the two first principal axes (associated to the two larger eigenvalues). The vertices of the simplex defined by the centered-whitened data in the new 2 dimensional space are determined by the N-FINDR algorithm [4]. The $R = 3$ resulting endmember spectra corresponding to vegetation, water and soil are plotted in Fig. 8. It is interesting to note that other endmember extraction algorithms have been recently studied in the literature [20], [40].

B. Abundance estimation

The Bayesian unmixing algorithm defined in sections III and IV has been applied on each pixel of the hyperspectral image (using the endmember spectra resulting from VII-A). Various convergence diagnosis have shown that a short burn-in is sufficient for this example. This is confirmed in Fig. 9 (bottom) which shows a typical Markov chain output for the 3 abundance coefficients. Consequently, the burn-in period has been fixed to $N_{\text{bi}} = 10$ for all results presented in this section. The posterior distributions of the abundances α_r ($r = 1, 2, 3$) are represented in Fig. 9 (top) for the pixel $\#(43, 35)$. These posterior distributions indicate that the pixel is composed of soil essentially, reflecting that the pixel is located on a coast area containing very few vegetation.

The image fraction maps estimated by the proposed algorithm for the $R = 3$ pure materials are represented in Fig. 10 (top). Note that a white (resp. black) pixel in the map indicates a large (resp. small) value of the abundance coefficient. Note also that the estimates have been obtained by averaging the last $N_r = 900$ simulated samples for each pixel, according to the MMSE principle. The lake area (represented by white pixels in the water fraction map and by black pixels in the other maps) can be clearly recovered. Note that the analysis of this image takes approximatively 18 minutes for a MATLAB implementation on a 2.8 GHz Pentium IV. The results obtained with the deterministic fraction mapping routine of the ENVI software [32, p. 739] are represented in Fig. 10 (bottom) for comparison. These figures obtained with a constrained least-squares algorithm (satisfying the additivity and positivity constraints) are clearly in good agreement with Fig. 10 (top). However, the proposed Bayesian algorithm allows one to estimate

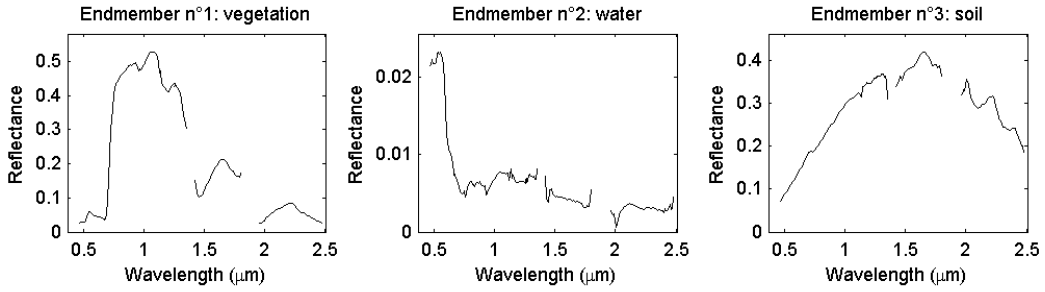

Fig. 8. The $R = 3$ endmember spectra obtained by the N-FINDR algorithm.

the full posterior of the abundance coefficients and the noise variance. This posterior can be used to compute measures of confidence regarding the estimates.

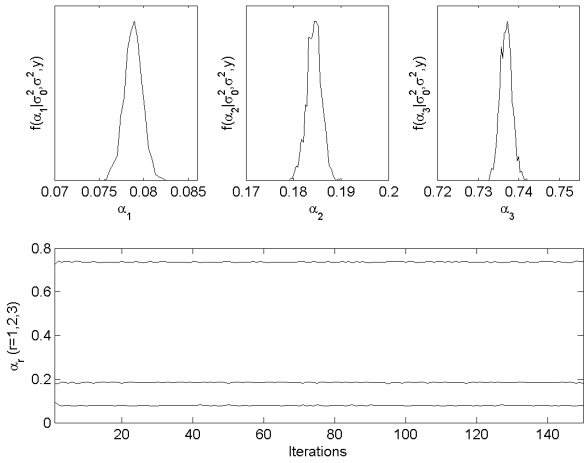

Fig. 9. Top: posteriors of the abundances α_r ($r = 1, \ldots, 3$) for the pixel $\#(43, 35)$. Bottom: 150 first outputs of the sampler.

C. Convergence of the sampler

As explained in Section V, the convergence of the sampler can be checked by monitoring some key parameters such as the parameter σ^2. The outputs of 5 different Markov chains for parameter σ^2 are depicted in Fig. 11 for the pixel $\#(43, 35)$. All chains clearly converge to a similar value. The potential scalar reduction factor associated with the noise variance σ^2 is computed from $M = 10$ Markov chains for each pixel. The values of $\sqrt{\hat{\rho}}$ computed for each

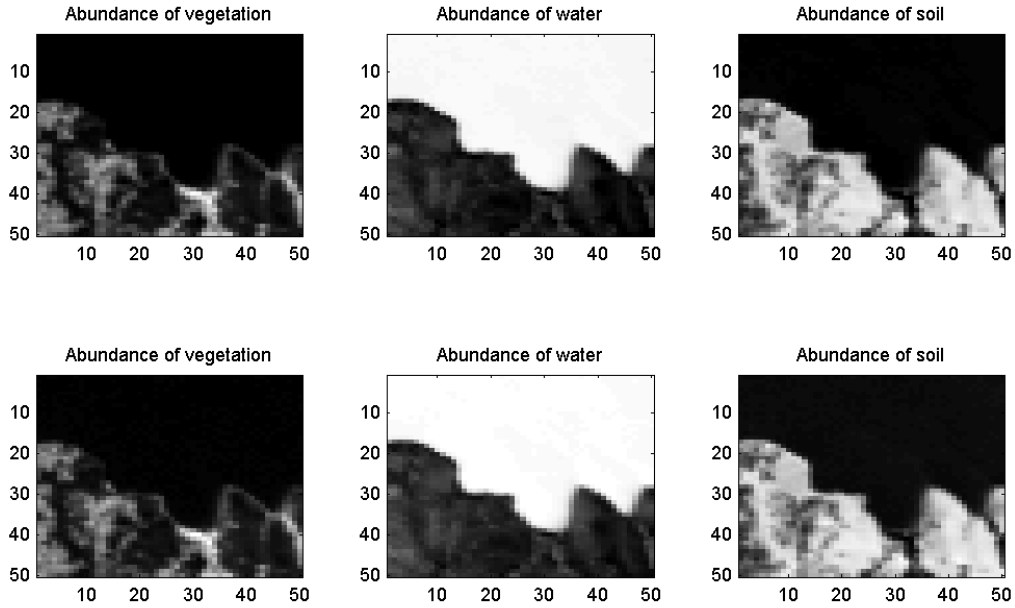

Fig. 10. Top: the fraction maps estimated by the proposed algorithm (black (resp. white) means absence (resp. presence) of the material). Bottom: the fraction maps recovered by the ENVI software.

pixel are represented in Fig. 12. All these values are below 1.0028 (the value obtained for the pixel $\#(10, 26)$) which indicate a good convergence of the sampler for each pixel.

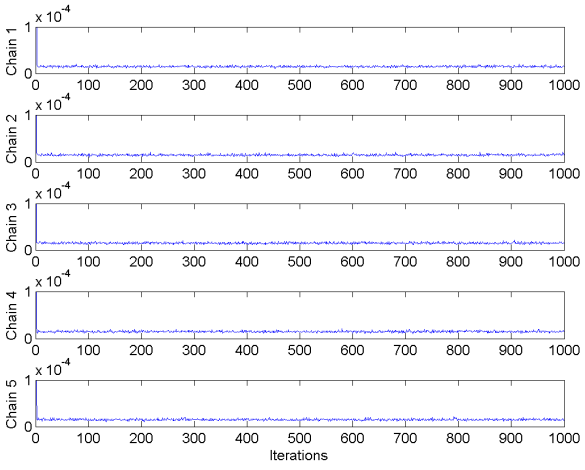

Fig. 11. Convergence assessment with five realizations of the Markov chain.

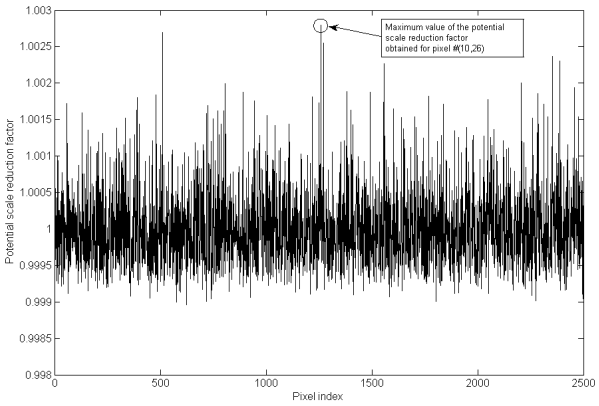

Fig. 12. Potential scale reduction factors computed for each pixel.

D. Sensibility to the endmember extraction step

The proposed unmixing procedure does not seem sensitive to the endmember extraction step decribed in Subsection VII-A. To illustrate this point, simulations have been performed on the real data with a different endmember extraction procedure. First, the Moffett Field hyperspectral image has been segmented by an unsupervised K-means algorithm initialized with 3 classes. The segmentation results are depicted in Figure 13 showing that the "water", "vegetation" and "soil" classes can be easily recovered.

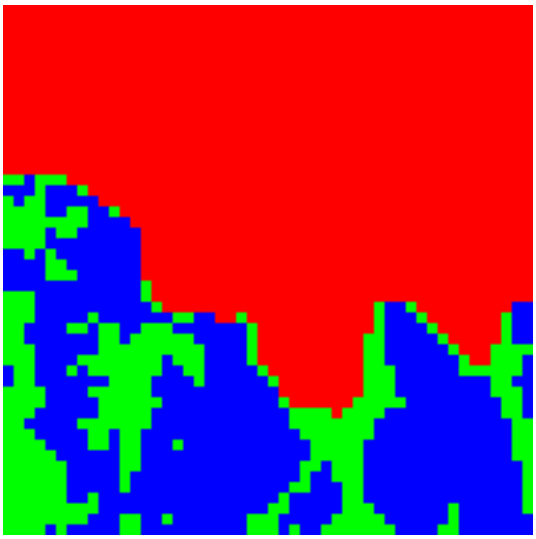

Fig. 13. Output of the K-means algorithm applied on Moffett Field image.

The purest pixels belonging to each class have been identified thanks to the purety pixel index (PPI) algorithm (with 15000 iterations). For each class, the pixels with highest PPI scores have been retained. Their spectra have been averaged to define the three new spectra of the pure components. The resulting new endmember spectra have been computed and are compared to the previous spectra in Figure 14 (top). Finally, the unmixing procedure proposed in the paper has been performed with these new endmembers. The abundance maps, depicted in Figure 14 (bottom), are clearly similar to those obtained with the N-FINDR procedure.

VIII. Estimating the number of endmembers using a reversible jump sampler

This section generalizes the previous hierarchical Bayesian sampler to linear mixtures with an unknown number of components R. We assume here that the R endmember spectra belong to a known library $\mathcal{S} = \{\mathbf{s}_1, \ldots, \mathbf{s}_{R_{\max}}\}$ (where $\mathbf{s}_r$ denotes the L-spectrum $[s_{r,1}, \ldots, s_{r,L}]^{\mathsf{T}}$ of the endmember $\#r$). However, the number of components R as well as the corresponding spectra belonging to $\mathcal{S}$ are unknown.

A. Extended Bayesian model

The posterior distribution of the unknown parameter vector $\{\boldsymbol{\alpha}, \mathbf{M}^+, R, \sigma^2\}$ can be written:

$$\begin{aligned} f\left(\boldsymbol{\alpha}, \mathbf{M}^+, R, \sigma^2 | \mathbf{y}\right) \propto \quad & f\left(\mathbf{y} | \boldsymbol{\alpha}, \mathbf{M}^+, \sigma^2, R\right) \\ & \times f\left(\boldsymbol{\alpha} | R\right) f\left(\mathbf{M}^+ | R\right) f\left(\sigma^2\right) f\left(R\right), \end{aligned} \tag{19}$$

where

$$f(\mathbf{y} | \boldsymbol{\alpha}^+, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{L}{2}} \exp\left[-\frac{\|\mathbf{y} - \mathbf{M}(R)^+ \boldsymbol{\alpha}(R)^+\|^2}{2\sigma^2}\right]. \tag{20}$$

and the dimensions of $\mathbf{M}(R)^+$ and $\boldsymbol{\alpha}(R)$ depend on the unknown parameter R. The priors $f\left(\boldsymbol{\alpha}|R\right)$ and $f\left(\sigma^2\right)$ have been defined in section III-B. A discrete uniform distribution on $[2, \ldots, R_{\max}]$ is chosen for the prior associated to the number of mixture components R:

$$f\left(R\right) = \frac{1}{R_{\max} - 1}, \quad R = 2, \ldots, R_{\max}. \tag{21}$$

Moreover, all combinations of R spectra belonging to the library $\mathcal{S}$ are assumed to be equiprobable conditional upon R:

$$\begin{aligned} f(\mathbf{M}^+ \mid R) &= \frac{1}{\binom{R_{\max}}{R}}, \\ &= \frac{\Gamma\left(R+1\right)\Gamma\left(R_{\max} - R + 1\right)}{\Gamma\left(R_{\max}+1\right)}. \end{aligned} \tag{22}$$

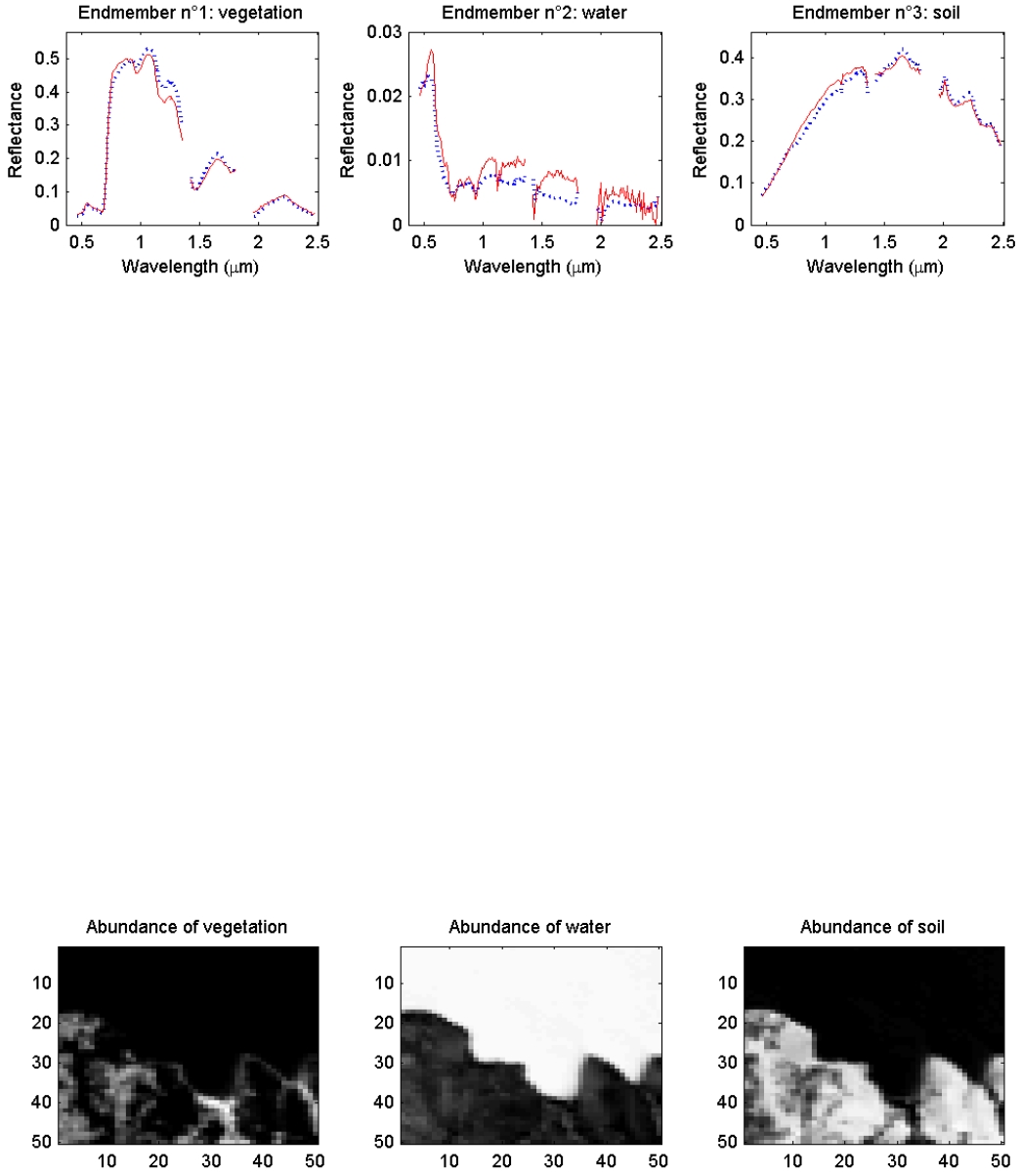

Fig. 14. Top: the $R = 3$ endmember spectra obtained by the combined K-means/PPI procedure (red solid line) and by the NFINDR procedure (blue dotted lines). Bottom: the fraction maps of the endmember recovered by the combined K-means/PPI procedure in the scene (black (resp. white) means absence (resp. presence) of the material)

B. Hybrid Metropolis-within-Gibbs algorithm

This section studies an hybrid Metropolis-within-Gibbs algorithm to sample according to $f\left(\boldsymbol{\alpha}, \mathbf{M}^{+}, \sigma^{2}, R|\mathbf{y}\right)$. The vectors to be sampled belong to a space whose dimension depends

ALGORITHM 3:

Hybrid Metropolis-within-Gibbs sampler for hyperspectral image unmixing

- Initialization:
  - Sample parameter $R^{(0)}$,
  - Choose $R^{(0)}$ spectra in the library $\mathcal{S}$ to build $\mathbf{M}^{+(0)}$,
  - Sample parameters $\sigma^{2(0)}$ and $\boldsymbol{\alpha}^{(0)}$,
  - Set $t \leftarrow 1$,
- Iterations: for $t = 1, 2, \ldots,$ do
  - Update the spectrum matrix $\mathbf{M}^{+(t)}$:

    draw $u_1 \sim \mathcal{U}_{[0,1]}$,

    IF $u_1 \leq b_{R^{(t-1)}}$, THEN

    propose a BIRTH move (see Algorithm 4),

    IF $b_{R^{(t-1)}} < u_1 \leq b_{\widetilde{R}^{(t-1)}} + d_{\widetilde{R}^{(t-1)}}$, THEN

    propose a DEATH move (see Algorithm 5),

    IF $u_1 > b_{\widetilde{R}^{(t-1)}} + d_{\widetilde{R}^{(t-1)}}$ THEN

    propose a SWITCH move (see Algorithm 6),

    draw $u_2 \sim \mathcal{U}_{[0,1]}$,

    IF $u_2 < \rho$ (see (35) or (25)) THEN

    set $\left(\boldsymbol{\alpha}^{(t)}, \mathbf{M}^{+(t)}, R^{(t)}\right) = (\boldsymbol{\alpha}^{*}, \mathbf{M}^{+*}, R^{*})$,

    ELSE

    set $\left(\boldsymbol{\alpha}^{(t)}, \mathbf{M}^{+(t)}, R^{(t)}\right) = \left(\boldsymbol{\alpha}^{(t-1)}, \mathbf{M}^{+(t-1)}, R^{(t-1)}\right)$,
  - Sample $\boldsymbol{\alpha}^{(t)}$ from the pdf in (26),
  - Sample $\sigma^{2(t)}$ from the pdf in (27),
  - Set $t \leftarrow t + 1$.

on R, requiring to use a dimension matching strategy as in [10]. More precisely, the proposed algorithm referred to as Algorithm 3 consists of three moves:

1) updating the endmember spectra $\mathbf{M}^{+}$,

2) updating the abundance vector $\boldsymbol{\alpha}$,

3) updating the noise variance σ^2.

The three moves are scanned systematically as in [10] and are detailed below.

1) Updating the endmember spectra $\mathbf{M}^{+}$*:* The endmember spectra involved in the mixing model are updated by using three types of move, referred to as "BIRTH", "DEATH" and "SWITCH" moves, as in [21, p. 53]. The first two of these moves consist of increasing or decreasing the number of pure components R by 1. Therefore, they require the use of the reversible jump MCMC method introduced by Green [41] and then widely used in the signal processing literature (see [11], [12] or more recently [42]). Conversely, the dimension of R is not changed in the third move, requiring the use of a standard Metropolis-Hastings acceptance procedure. Assume that at iteration t, the current model is defined by $\left(\boldsymbol{\alpha}^{(t)}, \mathbf{M}^{+(t)}, \sigma^{2(t)}, R^{(t)}\right)$. The "BIRTH", "DEATH" and "SWITCH" moves are defined as follows:

- BIRTH: a *birth* move $R^{\star} = R^{(t)} + 1$ is proposed with the probability $b_{R^{(t)}}$ as explained in Algorithm 4. A new spectrum $\mathbf{s}^{\star}$ is randomly chosen among the available endmembers of the library $\mathcal{S}$ to build $\mathbf{M}^{+\star} = \left[\mathbf{M}^{+(t)}, \mathbf{s}^{\star}\right]$. A new abundance coefficient vector $\boldsymbol{\alpha}^{+\star}$ is proposed according to a rule inspired by [10]:
  - draw a new abundance coefficient $w^{\star}$ from the Beta distribution $\mathcal{B}e\left(1, R^{(t)}\right)$,
  - re-scale the existing weights so that all weights sum to 1, using $\alpha_r^{\star} = \alpha_r^{(t)}(1 - w^{\star})$, $r = 1, \ldots, R^{(t)}$,
  - build $\boldsymbol{\alpha}^{+\star} = \left[\alpha_1^{\star}, \ldots, \alpha_{R^{(t)}}^{\star}, w^{\star}\right]^{\mathsf{T}}$,

---

ALGORITHM 4:

BIRTH move

- set $R^{\star} = R^{(t)} + 1$,
- choose $\mathbf{s}^{\star}$ in $\mathcal{S}$ such as $\mathbf{s}^{\star} \neq \mathbf{m}_r^{(t)}$, $r = 1, \ldots, R^{(t)}$,
- add $\mathbf{s}^{\star}$ to $\mathbf{M}^{+(t)}$, i.e. set

$$\mathbf{M}^{+\star} = \left[\mathbf{m}_1^{(t)}, \ldots, \mathbf{m}_{R^{(t)}}^{(t)}, \mathbf{s}^{\star}\right], \tag{23}$$

- draw $w^{\star} \sim \mathcal{B}e\left(1, R^{(t)}\right)$,
- add $w^{\star}$ to $\boldsymbol{\alpha}^{+(t)}$ and re-scale the other coefficient abundances, i.e. set

$$\boldsymbol{\alpha}^{+\star} = \left[\frac{\alpha_1^{(t)}}{C}, \ldots, \frac{\alpha_{R^{(t)}}^{(t)}}{C}, w^{\star}\right]^{\mathsf{T}}, \tag{24}$$

with $C = \frac{1}{(1-w^{\star})}$.

- DEATH: a *death* move $R^\star = R^{(t)} - 1$ is proposed with the probability $d_{R^{(t)}}$ as explained in Algorithm 5. One of the spectra of $\mathbf{M}^{+(t)}$ is removed, as well as the corresponding abundance coefficient. The remaining abundances coefficients are re-scaled to sum to 1,

---

ALGORITHM 5:

DEATH move

– set $R^\star = R^{(t)} - 1$,

– draw $j \sim \mathcal{U}_{\{1,\ldots,R^{(t)}\}}$,

– remove $\mathbf{m}_j^{(t)}$ from $\mathbf{M}^{+(t)}$, i.e. set

$$\mathbf{M}^{+\star} = \left[\mathbf{m}_1^{(t)}, \ldots, \mathbf{m}_{j-1}^{(t)}, \mathbf{m}_{j+1}^{(t)}, \ldots, \mathbf{m}_{R^{(t)}}^{(t)}\right],$$

– remove $\alpha_j^{(t)}$ from $\boldsymbol{\alpha}^{+(t)}$ and re-scale the remaining abundance coefficients, i.e. set

$$\boldsymbol{\alpha}^{+\star} = \left[\frac{\alpha_1^{(t)}}{C}, \ldots, \frac{\alpha_{j-1}^{(t)}}{C}, \frac{\alpha_{j+1}^{(t)}}{C}, \ldots, \frac{\alpha_{R^{(t)}}^{(t)}}{C}\right]^\mathsf{T},$$

with $C = \sum_{r \neq j} \alpha_r^{(t)}$.

- SWITCH: a *switch* move is proposed with the probability $u_{R^{(t)}}$ (see Algorithm 6). A spectrum randomly chosen in $\mathbf{M}^{+(t)}$ is replaced by another spectrum randomly chosen in the library $\mathcal{S}$.

ALGORITHM 6:

SWITCH move

– draw $j \sim \mathcal{U}_{\{1,\ldots,R^{(t)}\}}$,

– choose $\mathbf{s}^\star$ in $\mathcal{S}$ such as $\mathbf{s}^\star \neq \mathbf{m}_r^{(t)}$, $r = 1, \ldots, R^{(t)}$,

– replace $\mathbf{m}_j^{(t)}$ in $\mathbf{M}^{+(t)}$ by $\mathbf{s}^\star$, i.e. set

$$\mathbf{M}^{+\star} = \left[\mathbf{m}_1^{(t)}, \ldots, \mathbf{m}_{j-1}^{(t)}, \mathbf{s}^\star, \mathbf{m}_{j+1}^{(t)}, \ldots, \mathbf{m}_{R^{(t)}}^{(t)}\right],$$

– let $\boldsymbol{\alpha}^{+\star} = \boldsymbol{\alpha}^{+(t)}$ and $R^\star = R^{(t)}$.

At each iteration, one of the moves "BIRTH", "DEATH" and "SWITCH" is randomly chosen with probabilities $b_{R^{(t)}}$, $d_{R^{(t)}}$ and $u_{R^{(t)}}$ with $b_{R^{(t)}} + d_{R^{(t)}} + u_{R^{(t)}} = 1$. Of course, the death move is not

allowed for $R = 2$ and the birth move is impossible for $R = R_{\max}$ (i.e. $d_2 = b_{R_{\max}} = 0$). As a consequence, $b_2 = d_{R_{\max}} = u_2 = u_{R_{\max}} = \frac{1}{2}$ and $b_R = d_R = u_R = \frac{1}{3}$ for $R \in \{3, \ldots, R_{\max} - 1\}$. The acceptance probabilities for the "birth" and "death" moves are $\rho = \min\{1, A_{\mathrm{b}}\}$ and $\rho = \min\left\{1, A_{\mathrm{b}}^{-1}\right\}$ where A_{b} is given in Appendix III.

The acceptance probability for the "switch" move is the standard Metropolis Hastings ratio $\rho = \min\{1, A_{\mathrm{s}}\}$ with

$$A_{\mathrm{s}} = \exp\left[-\frac{\left\|\mathbf{y} - \mathbf{M}^{+\star}\boldsymbol{\alpha}^{+\star}\right\|^2 - \left\|\mathbf{y} - \mathbf{M}^{+(t)}\boldsymbol{\alpha}^{+(t)}\right\|^2}{2}\right]. \tag{25}$$

Note that the proposal ratio associated to this switch move is 1, since in each direction the probability of selecting one spectrum from the library is $1 \,/\left(R_{\max} - R^{(t)}\right)$.

2) Generating samples according to $f(\boldsymbol{\alpha}|\mathbf{M}^+, R, \sigma^2, \mathbf{y})$: As in the initial model, the following posterior is obtained:

$$\boldsymbol{\alpha}|\mathbf{M}^+, \sigma^2, R, \mathbf{y} \sim \mathcal{N}_{\mathbb{S}}\left(\boldsymbol{\mu}, \boldsymbol{\Lambda}\right). \tag{26}$$

3) Generating σ^2 according to $f(\sigma^2|\boldsymbol{\alpha}, \mathbf{M}^+, R, \mathbf{y})$: This is achieved as follows:

$$\sigma^2|\boldsymbol{\alpha}, \mathbf{M}^+, R, \mathbf{y} \sim \mathcal{IG}\left(\frac{L}{2}, \frac{\left\|\mathbf{y} - \mathbf{M}^+\boldsymbol{\alpha}^+\right\|^2}{2}\right). \tag{27}$$

C. Simulations

The accuracy of the Metropolis-within-Gibbs sampler is studied by considering the synthetic pixel spectrum used in Section VI. Recall here that this pixel results from the combination of three endmembers (construction concrete, green grass, micaceous loam) with the abundance vector $[0.3, 0.6, 0.1]^{\mathsf{T}}$. The observation is corrupted by an additive Gaussian noise with SNR $= 15$dB. The results are obtained for $N_{\mathrm{MC}} = 20000$ iterations, including $N_{\mathrm{bi}} = 200$ burn-in iterations. This simulation uses a spectrum library containing six elements: construction concrete, green grass, micaceous loam, olive green paint, bare red brick, galvanized steel metal. The spectra of these pure components are depicted in Fig. 15.

The first step of the analysis estimates the model order R (i.e. the number of endmembers used for the mixture) using the maximum *a posteriori* (MAP) estimator. The posterior distribution of R depicted in Fig. 16 is clearly in good agreement with the actual value of R since its maximum is obtained for $R = 3$. The second step of the analysis estimates the posterior probabilities of all endmember combinations, conditioned to $R = 3$. For this experiment, only two vectors were

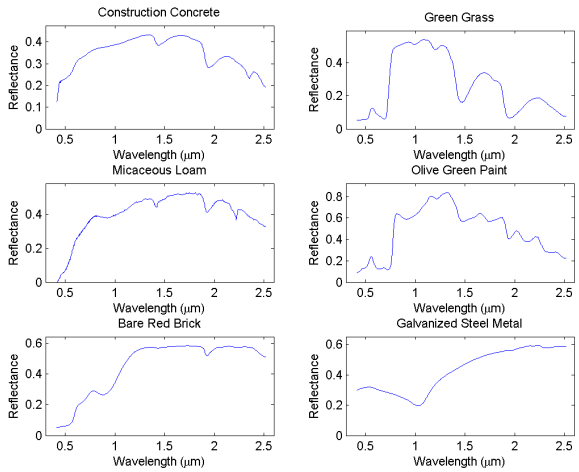

Fig. 15. Endmember spectra of the library.

generated $[\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3]$ and $[\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_5]$ with the probabilities $\mathrm{P}_{1,2,3} = 0.84$ and $\mathrm{P}_{1,2,5} = 0.16$. The maximum probability corresponds to the actual spectra involved in the mixture. The posterior distributions of the corresponding abundance coefficients are finally estimated and depicted in Fig. 17. These posteriors are clearly in good agreement with the actual values of the abundances $\boldsymbol{\alpha} = [0.3, 0.6, 0.1]^{\mathsf{T}}$. Note that unmixing this pixel with the values of N_{bi} and N_r defined above takes approximatively 50 seconds for a MATLAB implementation on a 2.8 GHz Pentium IV.

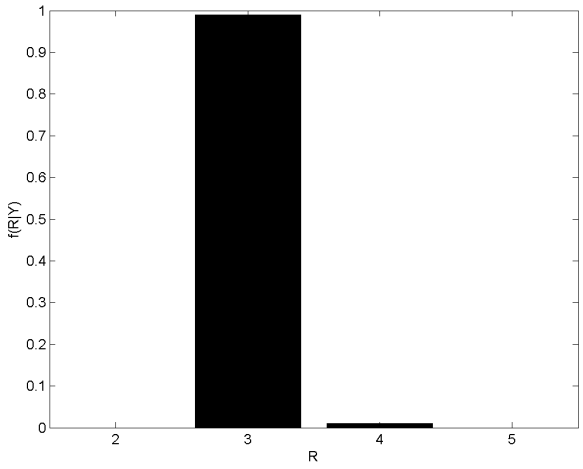

Fig. 16. Posterior distribution of the estimated model order R.

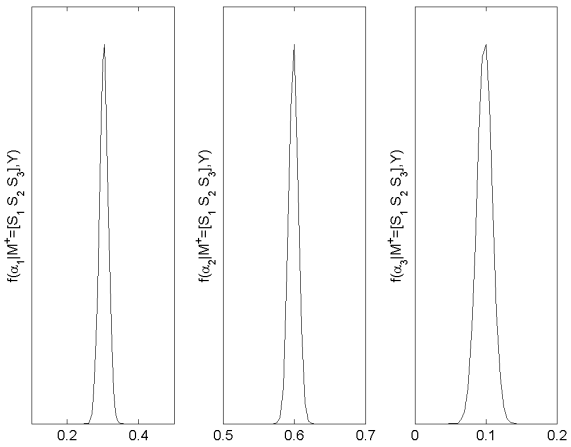

Fig. 17. Posterior distribution of the estimated abundances $\boldsymbol{\alpha}^{+} = [\alpha_1, \alpha_2, \alpha_3]^{\mathsf{T}}$ conditioned upon $R = 3$ and $\mathbf{M}^{+} = [\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3]$.

IX. Conclusions

This paper studied a hierarchical Bayesian model for hyperspectral image unmixing. The relationships between the different image spectra were naturally expressed in a Bayesian context by the prior distributions adopted for the model and their parameters. The posterior distributions of the unknown parameters related to this model were estimated by a Gibbs sampling strategy. These posterior distributions provided estimates of the unknown parameters but also information about their uncertainties such as standard deviations or confidence intervals. Two algorithms were developed depending whether the endmembers belonging to the mixture are known or belong to a known library. Simulation results conducted on synthetic and real images illustrated the performance of the proposed methodologies. It is interesting to note that the hierarchical Bayesian algorithm developed in this paper could be modified to handle more complicated models. Estimating the components of a mixture embedded in a correlated noise sequence is for instance under investigation.

Acknowledgments

The authors would like to thank Prof. Gérard Letac (LSP, Toulouse, France) for his helpful comments on multivariate truncated normal distributions and Saïd Moussaoui (IRCCyN, Nantes, France) for interesting discussions regarding this work. The authors are also very grateful to the Jet Propulsion Laboratory (Pasadena, CA, USA) for freely supplying the AVIRIS data.

APPENDIX I
TRUNCATED MULTIVARIATE NORMAL DISTRIBUTION

Let $\mathbb{E}$ be a Euclidean space with scalar product $\langle x, y\rangle$ and norm $\|x\| = \sqrt{\langle x, x\rangle}$. If $m \in \mathbb{E}$ and if Σ is a non singular positive definite operator on $\mathbb{E}$, the normal distribution on $\mathbb{E}$ is defined as follows:

$$\phi(dx|m, \Sigma) = (2\pi)^{-\dim \mathbb{E}/2} \exp\left[-\frac{1}{2}\big\langle x - m, \Sigma(x - m)\big\rangle\right] dx, \tag{28}$$

where the Lebesgue measure dx gives mass 1 to the cube built on any orthonormal basis of $\mathbb{E}$. The *standard* normal distribution is $\phi_{\mathbb{E}}(\cdot) = \phi(\cdot|0, \mathbf{I}_{\mathbb{E}})$. If U is a non empty open subset of $\mathbb{E}$, consider the following distribution on U:

$$\phi_U(dx|m, \Sigma) = \frac{\mathbf{1}_U(x)}{\phi(U|m, \Sigma)}\phi(dx|m, \Sigma). \tag{29}$$

We will say that $\phi_U(\cdot|m, \Sigma)$ is the U truncated normal distribution with *hidden mean* m and *hidden covariance* Σ. The reason of these terms is that m and Σ are not in general the mean and the covariance of the distribution $\phi_U(\cdot|m, \Sigma)$. For instance, if U is convex, a case which arises in most practical situations, then the mean of $\phi_U(\cdot|m, \Sigma)$ is necessarily in U although m can be outside of U. In general if U is known and fixed, the estimation of the hidden parameters m and Σ from a sample is not easy. If $\mathbb{E} = \mathbb{R}$ this estimation problem is studied for $U = (0, 1)$ in [43, example 9.16] and for $U = (0, \infty)$ in [44] and in [45, Chapter 2, Theorem 1.1].

Since the image of $\phi(\cdot|m, \Sigma)$ by $x \mapsto z = f(x) = \Sigma^{-1/2}(x - m)$ is the standard normal distribution $\phi_{\mathbb{E}}(\cdot)$, the image of $\phi_U(\cdot|m, \Sigma)$ by f is $\phi_{f(U)}(\cdot|0, \mathbf{I}_{\mathbb{E}})$. This important remark can be used to simulate $\phi_U(\cdot|m, \Sigma)$. Indeed, introduce $g(z) = m + \Sigma^{1/2}z$. One simulates easily i.i.d. random variables $z_1, \ldots, z_N, \ldots$ with the standard normal distribution $\mathcal{N}_{\mathbb{E}}$. Denote as $k_1 < k_2 < \ldots$ the set of integers k such that $z_k \in U$. One can prove that $(z_{k_j})_{j=1}^{\infty}$ are i.i.d. random variables with distribution $\phi_{f(U)}(\cdot|0, \mathbf{I}_{\mathbb{E}})$. This implies that $(x_{k_j})_{j=1}^{\infty} = (g(z_{k_j}))_{j=1}^{\infty}$ are i.i.d. random variables distributed according to the distribution $\phi_U(\cdot|m, \Sigma)$.

APPENDIX II
POSTERIOR DISTRIBUTION $f(\boldsymbol{\alpha}|\sigma^2, \mathbf{y})$

By using the Bayes theorem, the posterior distribution of $f(\boldsymbol{\alpha}|\sigma^2, \mathbf{y})$ can be written:

$$\begin{aligned} f(\boldsymbol{\alpha}|\sigma^2, \mathbf{y}) &\propto f(\mathbf{y}|\boldsymbol{\alpha}, \sigma^2) f(\boldsymbol{\alpha}), \\ &\propto \exp\left[-\frac{\|\mathbf{y} - \mathbf{M}^+\boldsymbol{\alpha}^+\|^2}{2\sigma^2}\right] \mathbf{1}_{\mathbb{S}}(\boldsymbol{\alpha}), \\ &\propto \exp\left[-\frac{C\left(\alpha|\mathbf{y}, \sigma^2\right)}{2\sigma^2}\right] \mathbf{1}_{\mathbb{S}}(\boldsymbol{\alpha}), \end{aligned}$$

with

$$C\left(\alpha|\mathbf{y}, \sigma^2\right) = \left\|\mathbf{y} - \mathbf{M}^+\boldsymbol{\alpha}^+\right\|^2.$$

Straightforward computations yield

$$\begin{aligned} \left\|\mathbf{y} - \mathbf{M}^+\boldsymbol{\alpha}^+\right\|^2 &= \sum_{l=1}^{L}\left[y_l - \sum_{r=1}^{R-1} m_{r,l}\alpha_r - m_{R,l}\alpha_R\right]^2, \\ &= \sum_{l=1}^{L}\left[\left(y_l - \sum_{r=1}^{R-1} m_{r,l}\alpha_r\right)^2 - 2\left(y_l - \sum_{r=1}^{R-1} m_{r,l}\alpha_r\right) m_{R,l}\alpha_R + \left(m_{R,l}\alpha_R\right)^2\right] \\ &= \left\|\mathbf{y} - \mathbf{M}\boldsymbol{\alpha}\right\|^2 - 2\left(\mathbf{y} - \mathbf{M}\boldsymbol{\alpha}\right)^{\mathsf{T}} \mathbf{m}_R(1 - \mathbf{u}^{\mathsf{T}}\boldsymbol{\alpha}) + \left\|\mathbf{m}_R(1 - \mathbf{u}^{\mathsf{T}}\boldsymbol{\alpha})\right\|^2, \end{aligned}$$

with $\mathbf{u} = [1, \ldots, 1]^{\mathsf{T}} \in \mathbb{R}^{R-1}$, hence

$$\begin{aligned} C\left(\alpha|\mathbf{y}, \sigma^2\right) &= \left\|\mathbf{y} - \mathbf{M}^+\boldsymbol{\alpha}^+\right\|^2 \\ &= \left[\left\|\mathbf{y} - \mathbf{M}\boldsymbol{\alpha}\right\|^2 - 2\left(\mathbf{y} - \mathbf{M}\boldsymbol{\alpha}\right)^{\mathsf{T}} \mathbf{m}_R(1 - \mathbf{u}^{\mathsf{T}}\boldsymbol{\alpha}) + \left\|\mathbf{m}_R\left(1 - \mathbf{u}^{\mathsf{T}}\boldsymbol{\alpha}\right)\right\|^2\right] \\ &= \left[\left(\mathbf{y}^{\mathsf{T}}\mathbf{y} - \boldsymbol{\alpha}^{\mathsf{T}}\mathbf{M}^{\mathsf{T}}\mathbf{y} - \mathbf{y}^{\mathsf{T}}\mathbf{M}\boldsymbol{\alpha} + \boldsymbol{\alpha}^{\mathsf{T}}\mathbf{M}^{\mathsf{T}}\mathbf{M}\boldsymbol{\alpha}\right) + 2\left(\mathbf{M}\boldsymbol{\alpha} - \mathbf{y}\right)^{\mathsf{T}}\left(\mathbf{m}_R - \mathbf{m}_R\mathbf{u}^{\mathsf{T}}\boldsymbol{\alpha}\right)\right] \\ &+ \left[\left\|\mathbf{m}_R\right\|^2\left(1 - 2\mathbf{u}^{\mathsf{T}}\boldsymbol{\alpha} + \boldsymbol{\alpha}^{\mathsf{T}}\mathbf{u}\mathbf{u}^{\mathsf{T}}\boldsymbol{\alpha}\right)\right] \end{aligned}$$

Reorganizing the different terms leads to

$$\begin{aligned} C\left(\alpha|\mathbf{y}, \sigma^2\right) &\propto \boldsymbol{\alpha}^{\mathsf{T}}\left[\left(\mathbf{M}^{\mathsf{T}}\mathbf{M} - \mathbf{M}^{\mathsf{T}}\mathbf{m}_R\mathbf{u}^{\mathsf{T}} - \mathbf{u}\mathbf{m}_R^{\mathsf{T}}\mathbf{M} + \left\|\mathbf{m}_R\right\|^2\mathbf{u}\mathbf{u}^{\mathsf{T}}\right)\right]\boldsymbol{\alpha} \\ &+ \boldsymbol{\alpha}^{\mathsf{T}}\left[\left(-\mathbf{M}^{\mathsf{T}}\mathbf{y} + \mathbf{M}^{\mathsf{T}}\mathbf{m}_R + \mathbf{u}\mathbf{m}_R^{\mathsf{T}}\mathbf{y} - \left\|\mathbf{m}_R\right\|^2\mathbf{u}\right)\right] \\ &+ \left[\left(-\mathbf{M}^{\mathsf{T}}\mathbf{y} + \mathbf{M}^{\mathsf{T}}\mathbf{m}_R + \mathbf{u}\mathbf{m}_R^{\mathsf{T}}\mathbf{y} - \left\|\mathbf{m}_R\right\|^2\mathbf{u}\right)\right]^{\mathsf{T}}\boldsymbol{\alpha}, \end{aligned}$$

or equivalently

$$\begin{aligned} C\left(\alpha|\mathbf{y},\sigma^2\right) \propto \boldsymbol{\alpha}^{\mathsf{T}}\left[\left(\mathbf{M}-\mathbf{m}_R\mathbf{u}^{\mathsf{T}}\right)^{\mathsf{T}}\left(\mathbf{M}-\mathbf{m}_R\mathbf{u}^{\mathsf{T}}\right)\right]\boldsymbol{\alpha} \\ -\boldsymbol{\alpha}^{\mathsf{T}}\left[\left(\mathbf{M}-\mathbf{m}_R\mathbf{u}^{\mathsf{T}}\right)^{\mathsf{T}}\left(\mathbf{y}-\mathbf{m}_R\right)\right] \\ -\left[\left(\mathbf{M}-\mathbf{m}_R\mathbf{u}^{\mathsf{T}}\right)^{\mathsf{T}}\left(\mathbf{y}-\mathbf{m}_R\right)\right]^{\mathsf{T}}\boldsymbol{\alpha}. \end{aligned}$$

By denoting

$$\begin{aligned} \boldsymbol{\Lambda} &= \left[\frac{1}{\sigma^2}\left(\mathbf{M}-\mathbf{m}_R\mathbf{u}^{\mathsf{T}}\right)^{\mathsf{T}}\left(\mathbf{M}-\mathbf{m}_R\mathbf{u}^{\mathsf{T}}\right)\right]^{-1}, \\ \mathbf{E} &= \boldsymbol{\Lambda}\left[\frac{1}{\sigma^2}\left(\mathbf{M}-\mathbf{m}_R\mathbf{u}^{\mathsf{T}}\right)^{\mathsf{T}}\left(\mathbf{y}-\mathbf{m}_R\right)\right], \end{aligned}$$

the posterior distribution of $f(\boldsymbol{\alpha}|\sigma^2,\mathbf{y})$ satisfies the following relation

$$f\left(\boldsymbol{\alpha}|\sigma^2,\mathbf{y}\right) \propto \exp\left[-\frac{\left(\boldsymbol{\alpha}-\mathbf{E}\right)^{\mathsf{T}}\boldsymbol{\Lambda}^{-1}\left(\boldsymbol{\alpha}-\mathbf{E}\right)}{2}\right]\mathbf{1}_{\mathbb{S}}(\boldsymbol{\alpha}).$$

APPENDIX III
ACCEPTANCE PROBABILITIES FOR THE "BIRTH" AND "DEATH" MOVES

This section derives the acceptance probabilities for the "birth" and "death" moves introduced in section VIII. At iteration index t, consider the birth move from the state $\left\{\boldsymbol{\alpha}^{(t)},\mathbf{M}^{+(t)},R^{(t)}\right\}$ to the new state $\left\{\boldsymbol{\alpha}^{\star},\mathbf{M}^{+\star},R^{\star}\right\}$ with $\boldsymbol{\alpha}^{\star}=[(1-w^{\star})\alpha_1,\ldots,(1-w^{\star})\alpha_{R^{(t)}}]^{\mathsf{T}}$, $\mathbf{M}^{+\star}=\left[\mathbf{M}^{+(t)},s^{\star}\right]$ and $R^{\star}=R^{(t)}+1$. The acceptance ratio associated to this "birth" move is:

$$\begin{aligned} A_{\mathrm{b}} = \frac{f\left(\boldsymbol{\alpha}^{\star},\mathbf{M}^{+\star},R^{\star}\mid\mathbf{y}\right)}{f\left(\boldsymbol{\alpha}^{(t)},\mathbf{M}^{+(t)},R^{(t)}\mid\mathbf{y}\right)}\frac{p_{R^{\star}\rightarrow R^{(t)}}}{p_{R^{(t)}\rightarrow R^{\star}}} \\ \times\frac{q\left(\mathbf{M}^{+(t)},\boldsymbol{\alpha}^{(t)}\mid\mathbf{M}^{+\star},\boldsymbol{\alpha}^{\star}\right)}{q\left(\mathbf{M}^{+\star},\boldsymbol{\alpha}^{\star}\mid\mathbf{M}^{+(t)},\boldsymbol{\alpha}^{(t)}\right)}\left|J\left(w^{\star}\right)\right|, \end{aligned} \tag{30}$$

where $q\left(\cdot|\cdot\right)$ refers to the proposal distribution, $\left|J\left(w^{\star}\right)\right|$ is the Jacobian of the transformation and $p_{\cdot\rightarrow\cdot}$ denotes the transition probability, i.e. $p_{R^{\star}\rightarrow R^{(t)}}=d_{R^{\star}}$ and $p_{R^{(t)}\rightarrow R^{\star}}=b_{R^{(t)}}$. According to the moves specified in section VIII, the proposal ratio is:

$$\frac{q\left(\mathbf{M}^{+(t)},\boldsymbol{\alpha}^{+(t)}\mid\mathbf{M}^{+\star},\boldsymbol{\alpha}^{+\star}\right)}{q\left(\mathbf{M}^{+\star},\boldsymbol{\alpha}^{+\star}\mid\mathbf{M}^{+(t)},\boldsymbol{\alpha}^{+(t)}\right)} = \frac{1}{g_{1,R^{(t)}}\left(w^{\star}\right)}\frac{R_{\max}-R^{(t)}}{R^{(t)}+1}, \tag{31}$$

where $g_{a,b}(\cdot)$ denotes the pdf of a Beta distribution $\mathcal{B}e(a,b)$. Indeed, the probability of choosing a new element in the library ("birth" move) is $1/\left(R_{\max}-R^{(t)}\right)$ and the the probability of removing an element ("death" move) is $1/\left(R^{(t)}+1\right)$.

The posterior ratio appearing in (30) can be rewritten as:

$$\frac{f\left(\boldsymbol{\alpha}^{\star},\mathbf{M}^{+\star},R^{\star}\mid\mathbf{y}\right)}{f\left(\boldsymbol{\alpha}^{(t)},\mathbf{M}^{+(t)},R^{(t)}\mid\mathbf{y}\right)}=\frac{f\left(\mathbf{y}\mid\boldsymbol{\alpha}^{\star},\mathbf{M}^{+\star},R^{\star}\right)}{f\left(\mathbf{y}\mid\boldsymbol{\alpha}^{(t)},\mathbf{M}^{+(t)},R^{(t)}\right)}\times\frac{f\left(\boldsymbol{\alpha}^{\star}|R^{\star}\right)}{f\left(\boldsymbol{\alpha}^{(t)}|R^{(t)}\right)}\frac{f\left(\mathbf{M}^{+\star}|R^{\star}\right)}{f\left(\mathbf{M}^{+(t)}|R^{(t)}\right)}\frac{f\left(R^{\star}\right)}{f\left(R^{(t)}\right)}. \quad (32)$$

Since the abundance coefficient vector $\boldsymbol{\alpha}^{+}$ has a Dirichlet prior $\mathcal{D}_R(\delta,\ldots,\delta)$, the prior ratio can be expressed as:

$$\frac{f\left(\boldsymbol{\alpha}^{\star}|R^{\star}\right)}{f\left(\boldsymbol{\alpha}^{(t)}|R^{(t)}\right)}=\frac{\Gamma\left(\delta R^{(t)}+\delta\right)}{\Gamma\left(\delta R^{(t)}\right)\Gamma\left(\delta\right)}\times w^{\star\delta-1}\left(1-w^{\star}\right)^{(\delta-1)R^{(t)}}. \quad (33)$$

By choosing a priori equiprobable configurations for $\mathbf{M}^{+}$ conditional upon R, the prior ratio related to the spectrum matrix is:

$$\frac{f\left(\mathbf{M}^{+\star}|R^{\star}\right)}{f\left(\mathbf{M}^{+(t)}|R^{(t)}\right)}=\frac{\binom{R_{\max}}{R^{(t)}}}{\binom{R_{\max}}{R^{\star}}}=\frac{R^{(t)}+1}{R_{\max}-R}. \quad (34)$$

The prior ratio related to the number of mixtures R associated to the uniform distribution specified in (21) reduces to 1.

Finally, the acceptance ratio for the BIRTH move can be written:

$$\begin{aligned}A_{\mathrm{b}}=&\exp\left[-\frac{\left\|\mathbf{y}-\mathbf{M}^{+\star}\boldsymbol{\alpha}^{+\star}\right\|^{2}-\left\|\mathbf{y}-\mathbf{M}^{+(t)}\boldsymbol{\alpha}^{+(t)}\right\|^{2}}{2}\right]\\&\times\frac{d_{R^{(t)}+1}}{b_{R^{(t)}}}\frac{1}{g_{1,R^{(t)}}\left(w^{\star}\right)}\left(1-w^{\star}\right)^{R^{(t)}-1}\\&\times\frac{\Gamma\left(\delta R^{(t)}+\delta\right)}{\Gamma\left(\delta R^{(t)}\right)\Gamma\left(\delta\right)}w^{\star\delta-1}\left(1-w^{\star}\right)^{(\delta-1)R^{(t)}},\end{aligned} \quad (35)$$

Note that (35) is very similar to Eq. (12) of [10] and that $\delta=1$ when $\boldsymbol{\alpha}$ has a uniform prior on the simplex $\mathbb{S}$.

REFERENCES

[1] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, pp. 44–57, Jan. 2002.

[2] B. W. Hapke, "Bidirectional reflectance spectroscopy. I. Theory," *J. Geophys. Res.*, vol. 86, pp. 3039–3054, 1981.

[3] P. E. Johnson, M. O. Smith, S. Taylor-George, and J. B. Adams, "A semiempirical method for analysis of the reflectance spectra of binary mineral mixtures," *J. Geophys. Res.*, vol. 88, pp. 3557–3561, 1983.

[4] M. 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.

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

[6] T. M. Tu, C. H. Chen, and C.-I Chang, "A noise subspace projection approach to target signature detection and extraction in an unknown background for hyperspectral images," *IEEE Trans. Geosci. and Remote Sensing*, vol. 36, no. 1, pp. 171–181, Jan. 1998.

[7] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, "Introducing Markov Chain Monte Carlo," in *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, Eds. London: Chapman & Hall, 1996, pp. 1–19.

[8] E. Kuhn and M. Lavielle, "Coupling a stochastic approximation version of EM with an MCMC procedure," *ESAIM Probab. Statist.*, vol. 8, pp. 115–131, 2004.

[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, pp. 259–273.

[10] S. Richardson and P. J. Green, "On Bayesian analysis of mixtures with unknown number of components," *J. Roy. Stat. Soc. B*, vol. 59, no. 4, pp. 731–792, 1997.

[11] C. Andrieu and A. Doucet, "Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC," *IEEE Trans. Signal Processing*, vol. 47, no. 10, pp. 19–37, Oct. 1999.

[12] 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.

[13] N. Dobigeon, J.-Y. Tourneret, and J. D. Scargle, "Joint segmentation of multivariate astronomical time series: Bayesian sampling with a hierarchical model," *IEEE Trans. Signal Processing*, vol. 55, no. 2, pp. 414–423, Feb. 2007.

[14] M.-H. Chen and J. J. Deely, "Bayesian analysis for a constrained linear multiple regression problem for predicting the new crop of apples," *J. of Agricultural, Biological and Environmental Stat.*, vol. 1, pp. 467–489, 1996.

[15] S. Moussaoui, D. Brie, A. Mohammad-Djafari, and C. Carteret, "Separation of non-negative mixture of non-negative sources using a Bayesian approach and MCMC sampling," *IEEE Trans. Signal Processing*, vol. 54, no. 11, pp. 4133–4145, Nov. 2006.

[16] G. Rodriguez-Yam, R. A. Davis, and L. Scharf, "A Bayesian model and Gibbs sampler for hyperspectral imaging," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop*, Washington, D.C., Aug. 2002, pp. 105–109.

[17] T. Blumensath and M. E. Davies, "Monte-Carlo methods for adaptive sparse approximations of time-series," *IEEE Trans. Signal Processing*, vol. 55, no. 9, pp. 4474–4486, Sept. 2007.

[18] C. Févotte and S. J. Godsill, "A Bayesian approach for blind separation of sparse sources," *IEEE Trans. Audio, Speech, Language Processing*, vol. 14, no. 6, pp. 2174–2188, Nov. 2006.

[19] 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.

[20] J. M. Nascimento and J. M. B. 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.

[21] D. G. T. Denison, C. C. Holmes, B. K. Mallick, and A. F. M. Smith, *Bayesian methods for nonlinear classification and regression*. Chichester, England: Wiley, 2002.

[22] M. Craig, "Minimum volume transforms for remotely sensed data," *IEEE Trans. Geosci. and Remote Sensing*, pp. 542–552, 1994.

[23] A. Strocker and P. Schaum, "Application of stochastic mixing models to hyperspectral detection problems," in *Proc. SPIE, Algorithms for Multispectral and Hyperspectral Imagery III*, vol. 3071, Orlando, FL, 1997, pp. 47–60.

[24] M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Dunne, and J. F. Huntington, "ICE: A statistical approach to identifying endmembers in hyperspectral images," *IEEE Trans. Geosci. and Remote Sensing*, vol. 42, no. 10, pp. 2085–2095, Oct. 2004.

[25] N. Dobigeon and J.-Y. Tourneret, "Efficient sampling according to a multivariate Gaussian distribution truncated on a simplex," IRIT/ENSEEIHT/TéSA, Tech. Rep., March 2007. [Online]. Available: http://www.enseeiht.fr/˜dobigeon

[26] C. P. Robert, "Simulation of truncated normal variables," *Statistics and Computing*, vol. 5, pp. 121–125, 1995.

[27] C. P. Robert and S. Richardson, "Markov Chain Monte Carlo methods," in *Discretization and MCMC Convergence Assessment*, C. P. Robert, Ed. New York: Springer Verlag, 1998, pp. 1–25.

[28] A. Gelman and D. Rubin, "Inference from iterative simulation using multiple sequences," *Statistical Science*, vol. 7, no. 4, pp. 457–511, 1992.

[29] S. Godsill and P. Rayner, "Statistical reconstruction and analysis of autoregressive signals in impulsive noise using the Gibbs sampler," *IEEE Trans. Speech, Audio Processing*, vol. 6, no. 4, pp. 352–372, 1998.

[30] P. M. Djurić and J.-H. Chun, "An MCMC sampling approach to estimation of nonstationary hidden markov models," *IEEE Trans. Signal Processing*, vol. 50, no. 5, pp. 1113–1123, 2002.

[31] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*. London: Chapman & Hall, 1995.

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

[33] C.-I Chang and B. Ji, "Weighted abundance-constrained linear spectral mixture analysis," *IEEE Trans. Geosci. and Remote Sensing*, vol. 44, no. 2, pp. 378–388, Feb. 2001.

[34] 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, Sept. 1998.

[35] E. Christophe, D. Léger, and C. Mailhes, "Quality criteria benchmark for hyperspectral imagery," *IEEE Trans. Geosci. and Remote Sensing*, vol. 43, no. 9, pp. 2103–2114, Sept. 2005.

[36] F. W. Chen, "Archiving and distribution of 2-D geophysical data using image formats with lossless compression," *IEEE Geosci. and Remote Sensing Lett.*, vol. 2, no. 1, pp. 64–68, Jan. 2005.

[37] X. Tang and W. A. Pearlman, "Lossy-to-lossless block-based compression of hyperspectral volumetric data," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, vol. 5, Oct. 2004, pp. 3283–3286.

[38] T. Akgun, Y. Altunbasak, and R. M. Mersereau, "Super-resolution reconstruction of hyperspectral images," *IEEE Trans. Image Processing*, vol. 14, no. 11, pp. 1860–1875, Nov. 2005.

[39] AVIRIS Free Data. (2006) Jet Propulsion Lab. (JPL). California Inst. Technol., Pasadena, CA. [Online]. Available: http://aviris.jpl.nasa.gov/html/aviris.freedata.html

[40] F. Chaudhry, C.-C. Wu, W. Liu, C.-I Chang, and A. Plaza, "Pixel purity index-based algorithms for endmember extraction

from hyperspectral imagery," in *Recent Advances in Hyperspectral Signal and Image Processing*, C.-I Chang, Ed. Trivandrum, Kerala, India: Research Signpost, 2006, ch. 2.

[41] P. J. Green, "Reversible jump MCMC computation and bayesian model determination," *Biometrika*, vol. 82, no. 4, pp. 711–732, Dec. 1995.

[42] M. Davy, S. Godsill, and J. Idier, "Bayesian analysis of polyphonic western tonal music," *J. Acoust. Soc. Am.*, vol. 119, no. 4, pp. 2498–2517, April 2006.

[43] O. E. Barndorff-Nielsen, *Information and Exponential Families in Statistical Theory*. New-York: Wiley, 1978.

[44] J. del Castillo, "The singly truncated normal distribution: a non-steep exponential family," *Ann. Inst. Statist. Math*, vol. 46, no. 1, pp. 57–66, 1994.

[45] G. Letac, *Lectures on natural exponential families and their variance functions*. Rio de Janeiro, Brazil: Instituto de de Matemática Pura e Aplicada: Monografias de Matemática, 1992, vol. 50.