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

Jean-Yves Tourneret, Nicolas Dobigeon, Chein-I Chang

Semi-Supervised Linear Spectral Unmixing Using a Hierarchical Bayesian Model for Hyperspectral Imagery

Nicolas Dobigeon, *Member, IEEE*, Jean-Yves Tourneret, *Senior Member, IEEE*, and Chein-I Chang, *Senior Member, IEEE*

***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*—Gibbs sampler, hierarchical Bayesian analysis, hyperspectral images, linear spectral unmixing, Markov chain Monte Carlo (MCMC) methods, 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 many 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 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.[1]

[1]Note that the proposed unmixing strategy is univariate in the sense that it is applied to each pixel of the image. Spatial correlation in the image could be considered by using hidden Markov models. This approach was for instance used in [7] for classification of hyperspectral images.

10.1109/TSP.2008.917851

The complexity of the posterior distributions for the unknown parameters requires to use appropriate simulation methods such as Markov chain Monte Carlo (MCMC) methods [8]. The prior distributions used in the present paper depend on hyperparameters 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 [9]. However, as explained in [10, 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 [11]–[15]. 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 [16]. Positivity constraints can be satisfied by choosing Gamma priors [17] or truncated Gaussian priors [18]. It is interesting to mention here that similar ideas have also been recently exploited to handle linear sparse approximation models. For instance, sparsity can be ensured by defining factoring mixtures with modified Rayleigh priors [19] or Student priors [20]. This paper defines a Bayesian model with priors satisfying positivity and additivity[2] 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 Sections 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.

[2]The term "additivity" comes from hyperspectral imagery [1] and corresponds to a "norm-1" constraint.

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 to 2.5 μm (see [1], [21], or more recently, [22]). However, the proposed analysis might be extended to nonlinear unmixing models, for instance, by using a basis function representation approach as in [23, p. 134]. The LMM assumes that the L-spectrum $\mathbf{y} = [y_1, \ldots, y_L]^T$ 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 $\mathbf{m}_r = [m_{r,1}, \ldots, m_{r,L}]^T$ denotes the spectrum of the rth material; α_r is the fraction of the rth material in the pixel; R is the number of pure materials (or *endmembers*) present in all the observed scene; L is the number of available spectral bands for the image; $\mathbf{n} = [n_1, \ldots, n_L]^T$ is the additive white noise sequence which is classically assumed to be an independent and identically distributed (i.i.d.) zero-mean Gaussian sequence[3] 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 $\boldsymbol{\alpha}^+ = [\alpha_1, \ldots, \alpha_R]^T$ satisfies the following positivity and additivity constraints

$$\forall r \in \{1, \ldots, R\},\ \alpha_r \geq 0 \text{ and } \sum_{r=1}^{R} \alpha_r = 1. \tag{2}$$

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], [25] or statistical procedures [26], [27]. 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 (2). This model is based on the likelihood of the observations and on prior distributions for the unknown parameters.

A. Likelihood

Equation (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 $\boldsymbol{\alpha}^+ = [\alpha_1, \ldots, \alpha_R]^T$. 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 $\|\boldsymbol{x}\|^2 = \boldsymbol{x}^T\boldsymbol{x}$ is the standard L^2 norm.

[3]More complicated noise structures could be considered. As an example, analyzing data contaminated by colored additive Gaussian noise has been studied in [24]. Following the ideas developed in [15], the case of an additive noise modeled as an AR process could also be handled. However, this would increase the computational cost of the algorithm.

B. Parameter Priors

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

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

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 [23, p. 237] for the definition of the Dirichlet distribution $\mathcal{D}_R(1, \ldots, 1)$).

A conjugate inverse-gamma distribution (with parameters $(\nu/2)$ and $(\gamma/2)$) is chosen as prior distribution for σ^2

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

The hyperparameter ν will be fixed to $\nu = 2$ (as in [13]) 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 noninformative Jeffrey's prior[4] for the hyperparameter γ

$$f(\gamma) \propto \frac{1}{\gamma} \mathbf{1}_{\mathbb{R}^+}(\gamma) \quad (6)$$

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:

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

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

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

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 $f(\boldsymbol{\alpha}, \sigma^2 | \mathbf{y})$.

[4]It is important to note that there is no difference between choosing a noninformative Jeffrey's prior as prior distribution for σ^2 and the hierarchical prior defined by $f(\sigma^2 \,|\, \nu, \gamma)$ and $f(\gamma)$ proposed in the paper since $f(\sigma^2) = \int f(\sigma^2 | \nu, \gamma) f(\gamma) d\gamma \propto (1/\sigma^2)$. However, the proposed hierarchical structure defined by $f(\sigma^2 | \nu, \gamma)$ and $f(\gamma)$ is interesting since it can be generalized to a colored Gaussian noise with a signal-to-noise ratio (SNR) that may change from a spectral range to another (see [24] for more details).

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 Sections IV-A and IV-B (see also Algorithm 1).

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

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

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

where

$$\begin{cases} \boldsymbol{\Lambda} = \left[\frac{1}{\sigma^2} (\mathbf{M} - \mathbf{m}_R \mathbf{u}^T)^T (\mathbf{M} - \mathbf{m}_R \mathbf{u}^T) \right]^{-1} \\ \boldsymbol{\mu} = \boldsymbol{\Lambda} \left[\frac{1}{\sigma^2} (\mathbf{M} - \mathbf{m}_R \mathbf{u}^T)^T (\mathbf{y} - \mathbf{m}_R) \right] \end{cases} \quad (10)$$

with $\mathbf{u} = [1, \ldots, 1]^T \in \mathbb{R}^{R-1}$. As a consequence, $\boldsymbol{\alpha} | \sigma^2, \mathbf{y}$ is distributed according to a truncated Gaussian distribution[5]

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

The generation of samples according to a truncated Gaussian distribution can be achieved using a standard accept-reject procedure, 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 [28] or [29] for more details).

B. Generation of Samples 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:

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

V. Convergence Diagnosis

The Gibbs sampler allows one to draw sample $(\boldsymbol{\alpha}^{(t)}, \sigma^{2(t)})$ asymptotically distributed according to $f(\boldsymbol{\alpha}, \sigma^2 \,|\, \mathbf{y})$. The abun-

[5]From a practical point of view, the component of $\boldsymbol{\alpha}^+$ to be discarded are randomly chosen at each iteration of the Gibbs sampler.

dance vector can then be estimated by the empirical average following 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 $(\boldsymbol{\alpha}^{(t)}, \sigma^{2(t)})$ 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 (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 [30]. 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 [31] and has been used in many studies including [30, p. 33], [32], and [33]. The main idea is to run M parallel chains of length $N_r + N_{\mathrm{bi}}$ 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

$$B = \frac{N_r}{M-1}\sum_{m=1}^{M}(\bar{\kappa}_m - \bar{\kappa})^2 \tag{14}$$

$$W = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{N_r}\sum_{t=1}^{N_r}\left(\kappa_m^{(t)} - \bar{\kappa}_m\right)^2 \tag{15}$$

with

$$\bar{\kappa}_m = \frac{1}{N_r}\sum_{t=1}^{N_r}\kappa_m^{(t)}, \qquad \bar{\kappa} = \frac{1}{M}\sum_{m=1}^{M}\bar{\kappa}_m \tag{16}$$

where κ is the parameter of interest and $\kappa_m^{(t)}$ is its estimate at the tth run of the mth chain. The convergence of the chain can then be monitored by the so-called *potential scale reduction factor* $\hat{\rho}$ defined as [34, 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 that a number of burn-in iterations N_{bi} is sufficient to obtain samples $(\boldsymbol{\alpha}^{(N_{\mathrm{bi}}+t)}, \sigma^{2(N_{\mathrm{bi}}+t)})$, $t = 1, \ldots, N_r$, 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 (13). An *ad hoc* approach consists of assessing convergence via appropriate graphical evaluations [30, p. 28]. This paper proposes to compute a reference estimate denoted as $\tilde{\boldsymbol{\alpha}}$ from a large number of iterations to ensure convergence of the sampler and good accuracy of the approximation in (13) ($\tilde{\boldsymbol{\alpha}} = (1/10\,000)\sum_{t=1}^{10\,000}\boldsymbol{\alpha}^{(N_{\mathrm{bi}}+t)}$ in our simulations). The mean square error (MSE) between this reference estimate $\tilde{\boldsymbol{\alpha}}$ and the estimate obtained after p iterations is then computed as

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

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.

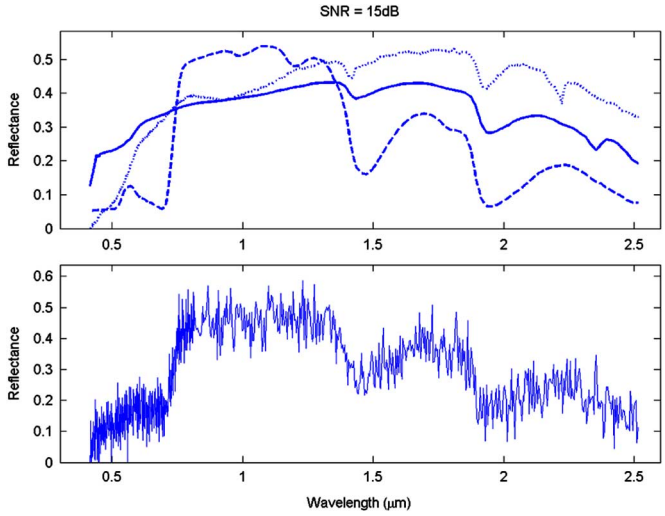


Fig. 1. Top: Endmember spectra: construction concrete (solid line), green grass (dashed line), dark yellowish brown micaceous loam (dotted line). Bottom: Resulting spectrum of the mixed pixel.

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 [35, p. 1035] and are representative of an 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 SNR is about $\mathrm{SNR} = 15\mathrm{dB}$, where $\mathrm{SNR} = L^{-1}\sigma^{-2}\|\sum_{r=1}^{R}\mathbf{m}_r\alpha_r\|^2$. The endmember spectra and the noisy spectrum of the mixed pixel are plotted in Fig. 1.

Fig. 2 shows the posterior distributions of the abundance coefficients $\alpha_r (r = 1, 2, 3)$ 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., $\boldsymbol{\alpha}^+ = [0.3, 0.6, 0.1]^T$. For comparison, the fully constrained least-squares (FCLS) algorithm detailed in [5], [36] has been run N_{MC} times for signals similar to Fig. 1(bottom) obtained with different noise sequences. Note that running $N_{\mathrm{MC}} = 20000$ times the FCLS algorithm on a pixel requires 6.23 s for

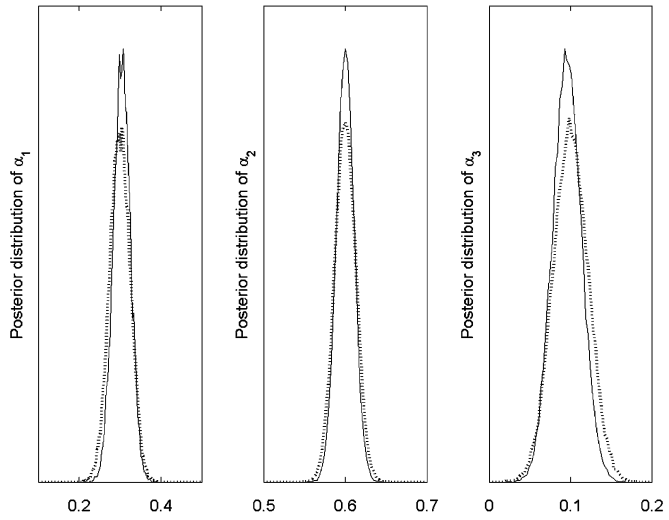


Fig. 2. Posterior distributions of the estimated abundances $[\alpha_1, \alpha_2, \alpha_3]^T$ (continuous lines) and histograms of FCLS estimates (dotted lines).

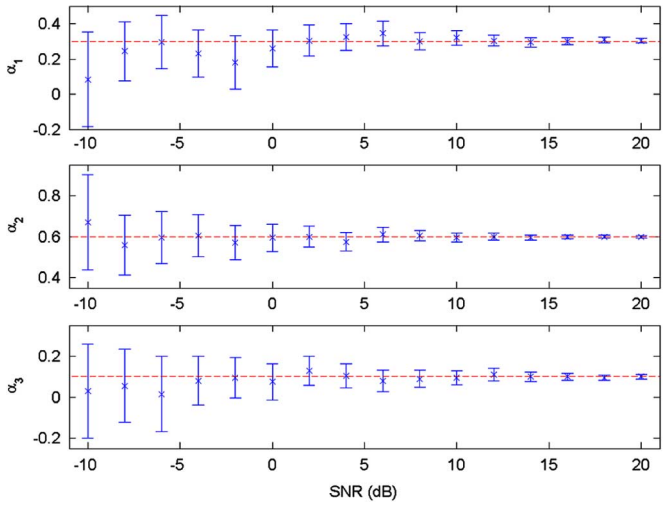


Fig. 3. MAP estimates (cross) and standard deviations (vertical bars) of $\alpha_r (r = 1, \ldots, 3)$ versus SNR.

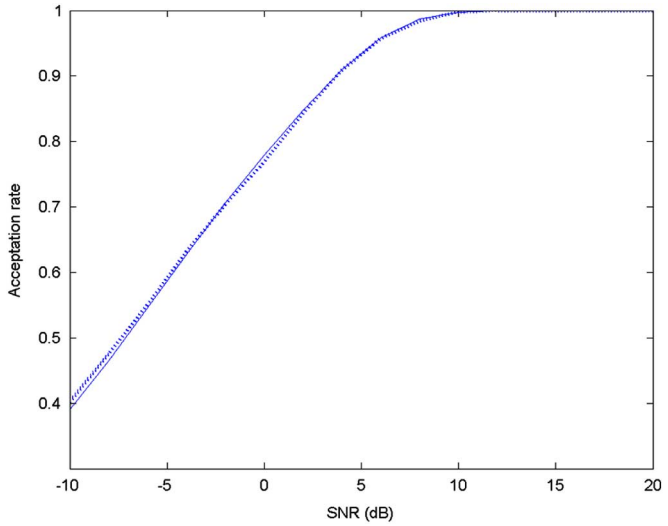


Fig. 4. Theoretical (solid) and experimental (dotted) acceptation rates of the accept-reject test versus SNR.

a MATLAB implementation on a 1.67-GHz Intel Core Duo. 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 FCLS 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 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 30 dB when the water absorption bands have been removed [37]. As a consequence, the results on Fig. 3 indicate that the proposed Bayesian algorithm performs satisfactorily for these SNRs. Fig. 3 also indicates that the proposed estimates of α_r converge (in the mean square sense) to the actual values of α_r when the SNR tends towards infinity.

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

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

where $\phi(\cdot \,|\, \boldsymbol{\mu}, \boldsymbol{\Lambda})$ is the probability density function (pdf) of a multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Lambda}$. Fig. 4 compares the theoretical acceptation rate $\mathrm{P}[\boldsymbol{\alpha} \in \mathbb{S}]$ resulting from a rectangle integration method which is compared with the experimental one estimated from the generation of 5000 Gaussian variables. These results have been obtained for a given value of $\boldsymbol{\alpha}^+ = [0.3, 0.6, 0.1]^T$ 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 ($\mathrm{SNR} > 30$ dB). It is interesting to mention here that we did not 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.

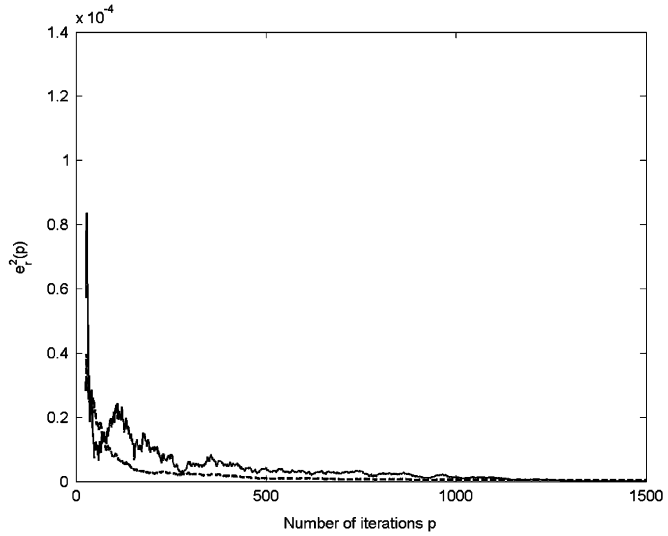


Fig. 5. MSE between the reference and estimated *a posteriori* change-point probabilities versus p (solid line). Averaged MSE computed from 10 chains (dashed line) ($N_{\mathrm{bi}} = 100$).

C. Sampler Convergence

The sampler convergence is monitored by computing the potential scale reduction factor introduced in Section V-A for an appropriate parameter of interest. Different choices for the 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 [32] for a similar choice). 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 $\sqrt{\hat{\rho}} \leq 1.2$ [34, p. 332]).

The number of iterations N_r necessary to compute an accurate estimate of $\boldsymbol{\alpha}$ according to the MMSE principle in (13) is determined by monitoring the MSE between a reference estimate $\tilde{\boldsymbol{\alpha}}$ (obtained with $N_r = 10000$) and the estimate obtained after $N_r = p$ iterations. Fig. 5 shows this MSE as a function of the number of iterations p (the number of burn-in iterations is $N_{\mathrm{bi}} = 100$). 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 (13) for this example. Note that, for such values of N_r and N_{bi}, unmixing this pixel takes approximately 0.3 s 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 [38]–[41]. The image depicted in Fig. 6 has 224 spectral bands, a nominal bandwidth of 10 nm, 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, CA (see [42] 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.

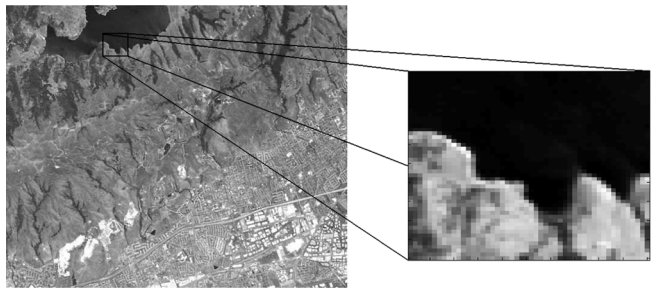


Fig. 6. Real hyperspectral data: Moffett Field acquired by AVIRIS in 1997 (left) and the region of interest at wavelength $\lambda = 0.66$ μm shown in gray scale (right).

The data set has been reduced from the original 224 bands to $L = 189$ bands by removing water absorption bands. Subimages of size 50×50 observed in L spectral bands have been processed by the proposed unmixing algorithm. This portion of the image is represented in gray scale at wavelength $\lambda = 0.66$ μm (band 30) in Fig. 6.

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 first two principal axes (associated to the two larger eigenvalues) which contain more than 95% of the information (i.e., $(\lambda_1 + \lambda_2 / \sum_{i=1}^{L} \lambda_i) > 0.95$). The vertices of the simplex defined by the centered-whitened data in the new 2–D 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. 7. It is interesting to note that other endmember extraction algorithms have been recently studied in [22] and [43]. The reader is invited to consult [44] for other simulation examples obtained with one of these algorithms.

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. 8 (bottom) which shows a typical Markov chain output for the three abundance coefficients. Consequently, the burn-in period has been fixed to $N_{\mathrm{bi}} = 10$ for all results presented in this section. The posterior distributions of the abundances $\alpha_r (r = 1, 2, 3)$ are represented in Fig. 8 (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.

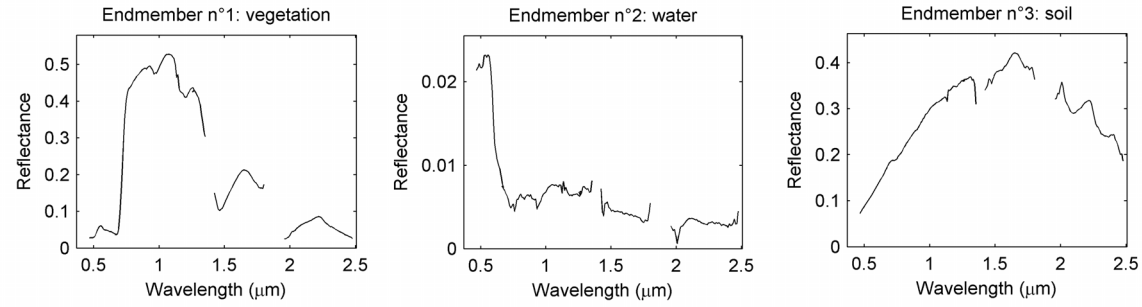


Fig. 7. The $R = 3$ endmember spectra obtained by the N-FINDR algorithm.

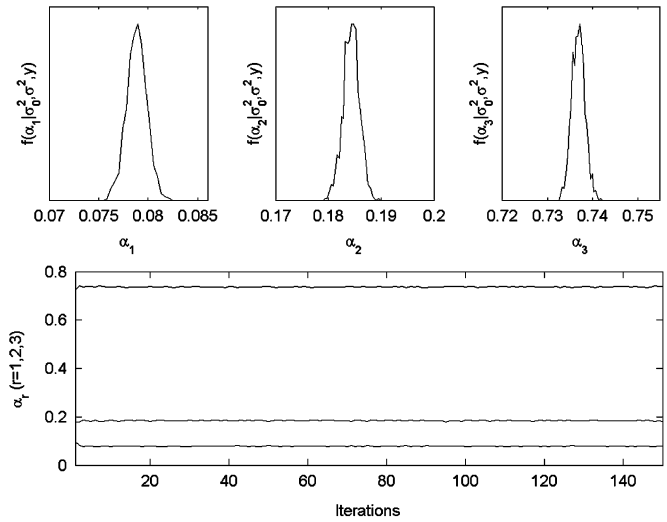


Fig. 8. Top: posteriors of the abundances $\alpha_r (r = 1, \ldots, 3)$ for the pixel #(43,35). Bottom: 150 first outputs of the sampler.

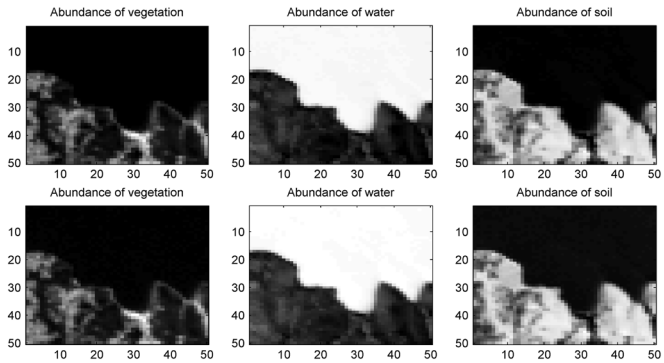


Fig. 9. Top: the fraction maps estimated by the proposed algorithm (black (respectively, white) means absence (respectively, presence) of the material). Bottom: the fraction maps recovered by the ENVI software.

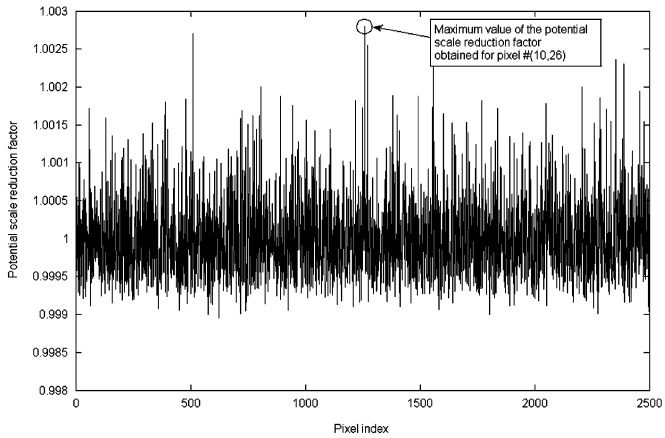


Fig. 10. Potential scale reduction factors computed for each pixel.

The image fraction maps estimated by the proposed algorithm for the $R = 3$ pure materials are represented in Fig. 9 (top). Note that a white (respectively, black) pixel in the map indicates a large (respectively, 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 approximately 18 min for a MATLAB implementation on a 2.8–GHz Pentium IV. The results obtained with the deterministic fraction mapping routine of the ENVI software [35, p. 739] are represented in Fig. 9 (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. 9 (top). However, the proposed Bayesian algorithm allows one to estimate the full posterior of the abundance coefficients and the noise variance. This posterior can be used to compute measures of confidence regarding the estimates.

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 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 pixel are represented in Fig. 10. 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.

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

ALGORITHM 2:
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 3),
    IF $b_{R^{(t-1)}} < u_1 \leq b_{\tilde{R}^{(t-1)}} + d_{\tilde{R}^{(t-1)}}$, THEN
      propose a DEATH move (see Algorithm 4),
    IF $u_1 > b_{\tilde{R}^{(t-1)}} + d_{\tilde{R}^{(t-1)}}$ THEN
      propose a SWITCH move (see Algorithm 5),
    draw $u_2 \sim \mathcal{U}_{[0,1]}$,
    IF $u_2 < \rho$ (see (33) or (25)) THEN
      set $\left(\boldsymbol{\alpha}^{(t)}, \mathbf{M}^{+(t)}, R^{(t)}\right) = \left(\boldsymbol{\alpha}^{\star}, \mathbf{M}^{+\star}, R^{\star}\right)$,
    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 $t \leftarrow t + 1$.

the L-spectrum $[s_{r,1}, \ldots, s_{r,L}]^T$ 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

$$f(\boldsymbol{\alpha}, \mathbf{M}^+, R, \sigma^2 \,|\, \mathbf{y}) \propto f(\mathbf{y} \,|\, \boldsymbol{\alpha}, \mathbf{M}^+, \sigma^2, R) \times f(\boldsymbol{\alpha} \,|\, R) f(\mathbf{M}^+ \,|\, R) f(\sigma^2) f(R) \quad (19)$$

where

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

and the dimensions of $\mathbf{M}(R)^+$ and $\boldsymbol{\alpha}(R)^+$ depend on the unknown parameter R. The priors $f(\boldsymbol{\alpha} \,|\, R)$ and $f(\sigma^2)$ 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

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

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

$$f(\mathbf{M}^+ \,|\, R) = \binom{R_{\max}}{R}^{-1} \quad (22)$$

with $\binom{R_{\max}}{R}^{-1} = \Gamma(R+1)\Gamma(R_{\max}-R+1)\Gamma(R_{\max}+1)^{-1}$.

B. Hybrid Metropolis-Within-Gibbs Algorithm

This section studies an hybrid Metropolis-within-Gibbs algorithm to sample according to $f(\boldsymbol{\alpha}, \mathbf{M}^+, \sigma^2, R \,|\, \mathbf{y})$. The vectors to be sampled belong to a space whose dimension depends on R, requiring to use a dimension matching strategy as in [11], [45]. More precisely, the proposed algorithm referred to as Algorithm 2 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 [11] and are detailed here.

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 [23, 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 [46] and then widely used in the signal processing literature (see [12], [13], or more recently, [47]). 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 $(\boldsymbol{\alpha}^{(t)}, \mathbf{M}^{+(t)}, \sigma^{2(t)}, R^{(t)})$. 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 3. A new spectrum $\mathbf{s}^{\star}$ is randomly chosen among the available endmembers of the library $\mathcal{S}$ to build $\mathbf{M}^{+\star} = [\mathbf{M}^{+(t)}, \mathbf{s}^{\star}]$. A new abundance coefficient vector $\boldsymbol{\alpha}^{+\star}$ is proposed according to a rule inspired by [11]:

ALGORITHM 3:
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

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

- 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

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

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

ALGORITHM 4:
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 $C = \sum_{r \neq j} \alpha_r^{(t)}$.

ALGORITHM 5:
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 $R^\star = R^{(t)}$.

— draw a new abundance coefficient $w^\star$ from the Beta distribution $\mathcal{B}e(1, R^{(t)})$;
— rescale 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 $\boldsymbol{\alpha}^{+\star} = [\alpha_1^\star, \ldots, \alpha_{R^{(t)}}^\star, w^\star]^T$.

- DEATH: A *death* move $R^\star = R^{(t)} - 1$ is proposed with the probability $d_{R^{(t)}}$ as explained in Algorithm 4. One of the spectra of $\mathbf{M}^{+(t)}$ is removed, as well as the corresponding abundance coefficient. The remaining abundances coefficients are rescaled to sum to 1.
- SWITCH: A *switch* move[6] is proposed with the probability $u_{R^{(t)}}$ (see Algorithm 5). A spectrum randomly chosen in $\mathbf{M}^{+(t)}$ is replaced by another spectrum randomly chosen in the library $\mathcal{S}$.

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, $b_2 = d_{R_{\max}} = u_2 = u_{R_{\max}} = (1/2)$ and $b_R = d_R = u_R = (1/3)$ 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 $\rho = \min\{1, A_{\mathrm{b}}^{-1}\}$ where A_{b} is given in Appendix I.

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

$$A_{\mathrm{s}} = \exp\left[-\frac{\|\mathbf{y} - \mathbf{M}^{+\star}\boldsymbol{\alpha}^{+\star}\|^2 - \left\|\mathbf{y} - \mathbf{M}^{+(t)}\boldsymbol{\alpha}^{+(t)}\right\|^2}{2}\right]. \tag{23}$$

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 $1/(R_{\max} - R^{(t)})$.

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:

$$\boldsymbol{\alpha} \,|\, \mathbf{M}^+, \sigma^2, R, \mathbf{y} \sim \mathcal{N}_{\mathbb{S}}(\boldsymbol{\mu}, \boldsymbol{\Lambda}). \tag{24}$$

[6]The "switch" move allows one to speed up the algorithm and to improve the mixing properties of the sampler. Assume the reversible jump algorithm has accepted the model $\mathbf{M}^+ = [\mathbf{s}_1, \mathbf{s}_2]$ whereas the correct model is $\mathbf{M}^+ = [\mathbf{s}_1, \mathbf{s}_3]$. In order to move from $[\mathbf{s}_1, \mathbf{s}_2]$ to $[\mathbf{s}_1, \mathbf{s}_3]$, the algorithm can certainly choose a new pure spectrum $\mathbf{s}_3$ (birth move) leading to $[\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3]$ and delete the component $\mathbf{s}_2$. However, the "switch" move allows one to move from $[\mathbf{s}_1, \mathbf{s}_2]$ to $[\mathbf{s}_1, \mathbf{s}_3]$ in a single step.

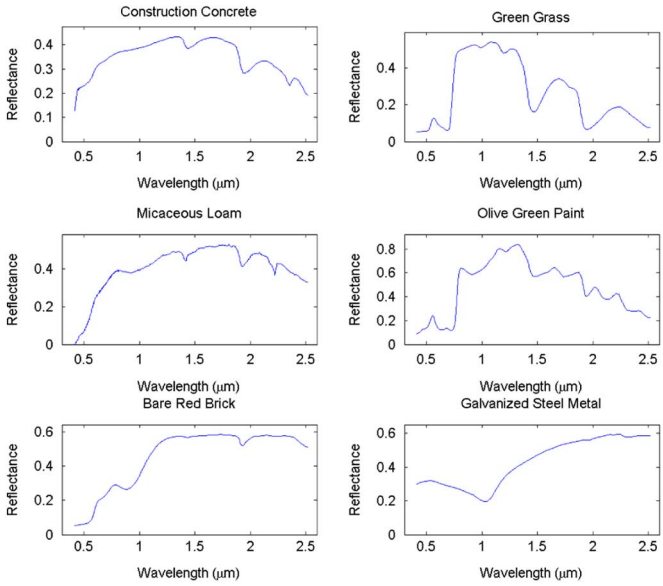


Fig. 11. Endmember spectra of the library.

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

$$\sigma^2 \,|\, \boldsymbol{\alpha}, \mathbf{M}^+, R, \mathbf{y} \sim \mathcal{IG}\left(\frac{L}{2}, \frac{\|\mathbf{y} - \mathbf{M}^+\boldsymbol{\alpha}^+\|^2}{2}\right). \tag{25}$$

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 $[0.3, 0.6, 0.1]^T$. The observation is corrupted by an additive Gaussian noise with $\mathrm{SNR} = 15$ 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, and galvanized steel metal. The spectra of these pure components are depicted in Fig. 11.

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. 12 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 matrices were 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. 13. These posteriors are clearly in good agreement with the actual values of the abundances $\boldsymbol{\alpha}^+ = [0.3, 0.6, 0.1]^T$. Note that unmixing this pixel with the values of N_{bi} and N_r defined above takes approximatively 50 s for a MATLAB implementation on a 2.8-GHz Pentium IV.

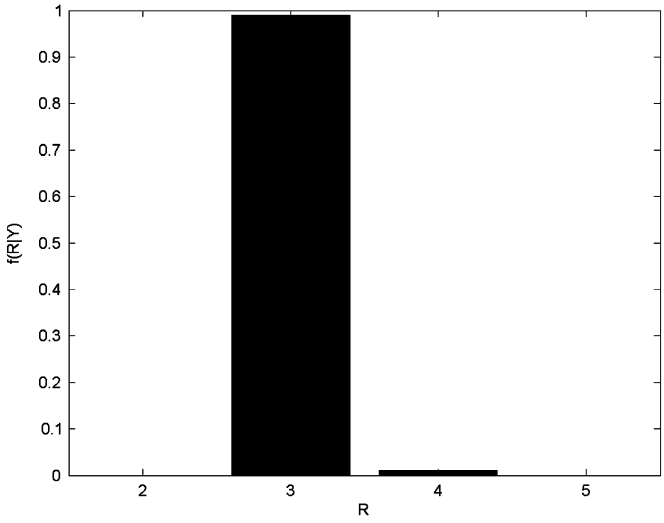


Fig. 12. Posterior distribution of the estimated model order R.

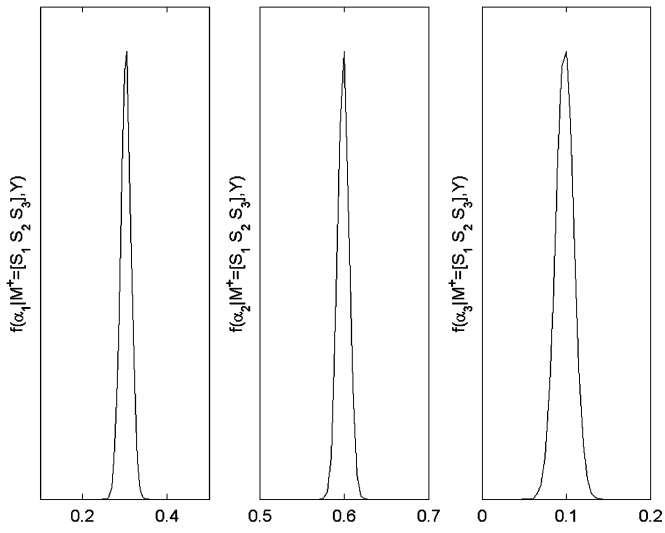


Fig. 13. Posterior distribution of the estimated abundances $\boldsymbol{\alpha}^{+} = [\alpha_1, \alpha_2, \alpha_3]^T$ conditioned upon $R = 3$ and $\mathbf{M}^{+} = [\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3]$.

IX. Conclusion

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.

The hierarchical Bayesian algorithm developed in this paper could be modified to handle more complicated models. For instance, it would be interesting to extend the proposed algorithm to unmix hyperspectral images composed of homogenous regions surrounded by sharp boundaries by introducing spatial correlation via hidden Markov models. Estimating the components of a mixture of endmembers embedded in other noise structures is also under investigation.

Appendix I
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 $\{\boldsymbol{\alpha}^{(t)}, \mathbf{M}^{+(t)}, R^{(t)}\}$ to the new state $\{\boldsymbol{\alpha}^{\star}, \mathbf{M}^{+\star}, R^{\star}\}$ with $\boldsymbol{\alpha}^{\star} = [(1-w^{\star})\alpha_1, \ldots, (1-w^{\star})\alpha_{R^{(t)}}]^T$, $\mathbf{M}^{+\star} = [\mathbf{M}^{+(t)}, s^{\star}]$ and $R^{\star} = R^{(t)} + 1$. The acceptance ratio associated to this "birth" move is

$$A_{\mathrm{b}} = \frac{f(\boldsymbol{\alpha}^{\star}, \mathbf{M}^{+\star}, R^{\star} \,|\, \mathbf{y})}{f\left(\boldsymbol{\alpha}^{(t)}, \mathbf{M}^{+(t)}, R^{(t)} \,|\, \mathbf{y}\right)} \frac{p_{R^{\star} \to R^{(t)}}}{p_{R^{(t)} \to R^{\star}}} \times \frac{q\left(\mathbf{M}^{+(t)}, \boldsymbol{\alpha}^{(t)} \,|\, \mathbf{M}^{+\star}, \boldsymbol{\alpha}^{\star}\right)}{q\left(\mathbf{M}^{+\star}, \boldsymbol{\alpha}^{\star} \,|\, \mathbf{M}^{+(t)}, \boldsymbol{\alpha}^{(t)}\right)} |J(w^{\star})| \quad (26)$$

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

$$\frac{q\left(\mathbf{M}^{+(t)}, \boldsymbol{\alpha}^{+(t)} \,|\, \mathbf{M}^{+\star}, \boldsymbol{\alpha}^{+\star}\right)}{q\left(\mathbf{M}^{+\star}, \boldsymbol{\alpha}^{+\star} \,|\, \mathbf{M}^{+(t)}, \boldsymbol{\alpha}^{+(t)}\right)} = \frac{1}{g_{1,R^{(t)}}(w^{\star})} \frac{R_{\max} - R^{(t)}}{R^{(t)} + 1} \quad (27)$$

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 $1/(R_{\max} - R^{(t)})$ and the probability of removing an element ("death" move) is $1/(R^{(t)}+1)$.

The posterior ratio appearing in (26) can be rewritten as

$$\frac{f\left(\boldsymbol{\alpha}^{\star}, \mathbf{M}^{+\star}, R^{\star} \,|\, \mathbf{y}\right)}{f\left(\boldsymbol{\alpha}^{(t)}, \mathbf{M}^{+(t)}, R^{(t)} \,|\, \mathbf{y}\right)} = \frac{f(\mathbf{y} \,|\, \boldsymbol{\alpha}^{\star}, \mathbf{M}^{+\star}, R^{\star})}{f\left(\mathbf{y} \,|\, \boldsymbol{\alpha}^{(t)}, \mathbf{M}^{+(t)}, R^{(t)}\right)} \times \frac{f\left(\boldsymbol{\alpha}^{\star} | R^{\star}\right)}{f\left(\boldsymbol{\alpha}^{(t)} | R^{(t)}\right)} \frac{f\left(\mathbf{M}^{+\star} | R^{\star}\right)}{f\left(\mathbf{M}^{+(t)} | R^{(t)}\right)} \frac{f(R^{\star})}{f\left(R^{(t)}\right)}. \quad (28)$$

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

$$\frac{f(\boldsymbol{\alpha}^{\star} \,|\, R^{\star})}{f\left(\boldsymbol{\alpha}^{(t)} \,|\, R^{(t)}\right)} = \frac{\Gamma\left(\delta R^{(t)} + \delta\right)}{\Gamma\left(\delta R^{(t)}\right)\Gamma(\delta)} w^{\star \delta - 1} (1 - w^{\star})^{(\delta-1)R^{(t)}}. \quad (29)$$

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

$$\frac{f(\mathbf{M}^{+\star}\,|\,R^{\star})}{f\left(\mathbf{M}^{+(t)}|R^{(t)}\right)} = \frac{\binom{R_{\max}}{R^{(t)}}}{\binom{R_{\max}}{R^{\star}}} = \frac{R^{(t)}+1}{R_{\max}-R}. \qquad (30)$$

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 is

$$\begin{aligned} A_{\mathrm{b}} = \exp&\left[-\frac{\|\mathbf{y}-\mathbf{M}^{+\star}\boldsymbol{\alpha}^{+\star}\|^2 - \left\|\mathbf{y}-\mathbf{M}^{+(t)}\boldsymbol{\alpha}^{+(t)}\right\|^2}{2}\right] \\ &\times \frac{d_{R^{(t)}+1}}{b_{R^{(t)}}}\frac{1}{g_{1,R^{(t)}}(w^{\star})}(1-w^{\star})^{R^{(t)}-1} \\ &\times \frac{\Gamma\left(\delta R^{(t)}+\delta\right)}{\Gamma\left(\delta R^{(t)}\right)\Gamma(\delta)} w^{\star\delta-1}(1-w^{\star})^{(\delta-1)R^{(t)}}. \qquad (31) \end{aligned}$$

Note that (31) is very similar to the equation given in [45] and that $\delta = 1$ when $\boldsymbol{\alpha}$ has a uniform prior on the simplex $\mathbb{S}$.

ACKNOWLEDGMENT

The authors would like to thank Prof. G. Letac (LSP, Toulouse, France) for his helpful comments on multivariate truncated normal distributions, M. Doisy (ENSEEIHT, Toulouse, France) for his feedback regarding reversible jump MCMCs and S. Moussaoui (IRCCyN, Nantes, France) for interesting discussions regarding this paper. The authors are also very grateful to the Jet Propulsion Laboratory (Pasadena, CA) for freely supplying the AVIRIS data.

REFERENCES

[1] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, 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. Applied Geologic Remote Sensing*, Vancouver, Canada, Apr. 1999, vol. 2, 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. Remote Sens.*, vol. 29, no. 3, pp. 529–545, Mar. 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. Remote Sens.*, vol. 36, no. 1, pp. 171–181, Jan. 1998.

[7] H. Snoussi, "Approche baysienne en sparation de sources, applications en imagerie," (in French) Ph.D. dissertation, Univ. Paris Sud, Orsay, France, 2003.

[8] 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, U.K.: Chapman & Hall, 1996, pp. 1–19.

[9] 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.

[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, pp. 259–273.

[11] 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.

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

[13] 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.

[14] 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 Process.*, vol. 55, no. 2, pp. 414–423, Feb. 2007.

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

[17] 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 Process.*, vol. 54, no. 11, pp. 4133–4145, Nov. 2006.

[18] 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 Process. Workshop*, Washington, DC, Aug. 2002, pp. 105–109.

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

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

[21] 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.

[22] J. M. Nascimento and J. M. B. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

[23] D. G. T. Denison, C. C. Holmes, B. K. Mallick, and A. F. M. Smith, *Bayesian Methods for Nonlinear Classification and Regression*. Chichester, U.K.: Wiley, 2002.

[24] N. Dobigeon, J.-Y. Tourneret, and A. O. Hero III, "Bayesian linear unmixing of hyperspectral images corrupted by colored gaussian noise with unknown covariance matrix," presented at the IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), Las Vegas, NV, Mar. 2008.

[25] M. Craig, "Minimum volume transforms for remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 3, pp. 542–552, May 1994.

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

[27] 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. Remote Sens.*, vol. 42, no. 10, pp. 2085–2095, Oct. 2004.

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

[29] C. P. Robert, "Simulation of truncated normal variables," *Statist. Comput.*, vol. 5, pp. 121–125, 1995.

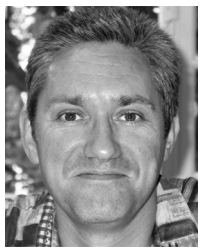
[30] C. P. Robert and D. Cellier, "Convergence control of MCMC algorithms," in *Discretization and MCMC Convergence Assessment*, C. P. Robert, Ed. New York: Springer-Verlag, 1998, pp. 27–46.

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

[32] S. Godsill and P. Rayner, "Statistical reconstruction and analysis of autoregressive signals in impulsive noise using the Gibbs sampler," *IEEE Trans. Speech, Audio Process.*, vol. 6, no. 4, pp. 352–372, 1998.
[33] P. M. Djurić and J.-H. Chun, "An MCMC sampling approach to estimation of nonstationary hidden Markov models," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1113–1123, May 2002.
[34] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*. London, U.K.: Chapman & Hall, 1995.
[35] ENVI User's Guide Version 4.0. Research Systems Inc. (RSI), Boulder, CO, Sep. 2003.
[36] C.-I Chang and B. Ji, "Weighted abundance-constrained linear spectral mixture analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 2, pp. 378–388, Feb. 2001.
[37] 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.
[38] E. Christophe, D. Léger, and C. Mailhes, "Quality criteria benchmark for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 9, pp. 2103–2114, Sep. 2005.
[39] F. W. Chen, "Archiving and distribution of 2-D geophysical data using image formats with lossless compression," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 1, pp. 64–68, Jan. 2005.
[40] X. Tang and W. A. Pearlman, "Lossy-to-lossless block-based compression of hyperspectral volumetric data," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2004, vol. 5, pp. 3283–3286.
[41] T. Akgun, Y. Altunbasak, and R. M. Mersereau, "Super-resolution reconstruction of hyperspectral images," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1860–1875, Nov. 2005.
[42] AVIRIS Free Data, Jet Propulsion Lab. (JPL), Calif. Inst. Technol., Pasadena, CA, 2006 [Online]. Available: http://aviris.jpl.nasa.gov/ html/aviris.freedata.html
[43] 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.
[44] N. Dobigeon, J.-Y. Tourneret, and C.-I Chang, "Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery," IRIT/ENSEEIHT/TeSA, Tech. Rep., Mar. 2007 [Online]. Available: http://www.enseeiht.fr/~dobigeon
[45] S. Richardson and P. J. Green, "Corrigendum: On Bayesian analysis of mixtures with unknown number of components," *J. Roy. Slot. Soc. B*, vol. 60, no. 3, p. 661, 1998.
[46] P. J. Green, "Reversible jump MCMC computation and Bayesian model determination," *Biometrika*, vol. 82, no. 4, pp. 711–732, Dec. 1995.
[47] M. Davy, S. Godsill, and J. Idier, "Bayesian analysis of polyphonic western tonal music," *J. Acoust. Soc. Amer.*, vol. 119, no. 4, pp. 2498–2517, Apr. 2006.

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 June 2004. In October 2007, he received the Ph.D. degree in signal processing also from the National Polytechnic Institute of Toulouse, within the Signal and Communication Group of the IRIT Laboratory.

Since November 2007, he has been a Postdoctoral Research associate with the Department of Electrical Engineering and Computer Science, University of Michigan. His research interests are centered around Bayesian inference and Markov chain Monte Carlo (MCMC) methods for signal and image processing.



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

He is currently a professor with 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 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, 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).

Chein-I Chang (S'81–M'82–SM'92) received the Ph.D. degree in electrical engineering from the University of Maryland, College Park, in 1987.

He has been with the University of Maryland, Baltimore County (UMBC), since 1987 and is currently a Professor in the Department of Computer Science and Electrical Engineering. He received a National Research Council (NRC) senior research associateship award from 2002–2003 and was a distinguished lecturer chair at the National Chung Hsing University from 2005 to 2006. He currently holds a chair professorship with the Environmental Restoration and Disaster Reduction Research Center and Department of Electrical Engineering, National Chung Hsing University, Taichung, Taiwan, R.O.C. He has four patents and was the guest editor and co-guest editor of four special issues on healthcare and hyperspectral imaging. He is author of the book *Hyperspectral Imaging: Techniques for Spectral Detection and Classification* (Boston, MA: Kluwer Academic, 2003) and editor of two books, *Recent Advances in Hyperspectral Signal and Image Processing* (Trivandrum, Kerala: Research Signpost, Trasworld Research Network, India, 2006) and *Hyperspectral Data Exploitation: Theory and Applications* (New York: Wiley, 2007). He is also co-editor (with A. Plaza) of the book *High Performance Computing in Remote Sensing* (Boca Raton, FL: CRC, 2007).

Dr. Chang was an Associate Editor in the area of hyperspectral signal processing for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING from 2001 to 2007 and is currently on editorial boards of three journals, the *Journal of High Speed Networks*, *Recent Patents on Mechanical Engineering*, and the *Open Remote Sensing Journal*. He is a Fellow of the SPIE.