MCMC ALGORITHMS FOR SUPERVISED AND UNSUPERVISED LINEAR UNMIXING OF HYPERSPECTRAL IMAGES

N. Dobigeon[1], S. Moussaoui[2], M. Coulon[1], J.-Y. Tourneret[1] and A.O. Hero[3]

Abstract. In this paper, we describe two fully Bayesian algorithms that have been previously proposed to unmix hyperspectral images. These algorithms relies on the widely admitted linear mixing model, *i.e.* each pixel of the hyperspectral image is decomposed as a linear combination of pure endmember spectra. First, the unmixing problem is addressed in a supervised framework, *i.e.*, when the endmembers are perfectly known, or previously identified by an endmember extraction algorithm. In such scenario, the unmixing problem consists of estimating the mixing coefficients under positivity and additivity constraints. Then the previous algorithm is extended to handle the unsupervised unmixing problem, *i.e.*, to estimate the endmembers and the mixing coefficients jointly. This blind source separation problem is solved in a lower-dimensional space, which effectively reduces the number of degrees of freedom of the unknown parameters. For both scenarios, appropriate distributions are assigned to the unknown parameters, that are estimated from their posterior distribution. Markov chain Monte Carlo (MCMC) algorithms are then developed to approximate the Bayesian estimators.

1 Abstract

For several decades, hyperspectral imagery has been demonstrating its high interest in numerous research works devoted to Earth monitoring. This interest can be easily explained by the high spectral resolution of the images provided by the

---

[1] University of Toulouse, IRIT/INP-ENSEEIHT, 2 rue Camichel, BP. 7122, 31071 Toulouse Cedex 7, France

[2] IRCCyN - CNRS UMR 6597, ECN, 1 rue de la Noë, BP. 92101, 44321 Nantes Cedex 3, France

[3] University of Michigan, Department of EECS, 1301 Beal Avenue, Ann Arbor, 48109-2122, USA

hyperspectral sensors. For instance, hyperspectral images can provide automatic classification maps for mineralogic surveys, avoiding long and tedious sampling campaigns (Jackson & Landgrebe 2002; Rellier *et al.* 2004). When environmental issues are on the front of the stage, hyperspectral imaging enables to provide crucial information related to macroscopic parameters, *e.g.*, the status of ecosystems or plants. Obviously, the price to pay for extracting the information contained in these images is to develop new methods exploiting the data provided by hyperspectral sensors efficiently.

Since the first hyperspectral images were acquired, spectral unmixing has been of considerable interest, not only in the remote sensing community, but also in the signal and image processing community. Solving this problem can indeed provide answers to various important issues such as classification (Chang 2003), material quantification (Plaza *et al.* 2005) and sub-pixel detection (Manolakis *et al.* 2001). Spectral unmixing consists of decomposing each pixel of the observed scene into a collection of reference spectra, usually referred to as endmembers, and estimating their proportions, or abundances, in each pixel (Bioucas-Dias *et al.* 2012). To formally describe the mixture, the most frequently encountered model is the macroscopic model which gives a good approximation of the nonlinear model introduced by Hapke (Hapke 1981) in the reflective spectral domain from visible to near-infrared (0.4 μm to 2.5 μm) (Johnson *et al.* 1983). This linear model assumes that the observed pixel spectrum is a weighted linear combination of the endmember spectra.

As noticed in (Keshava & Mustard 2002), linear spectral unmixing has often been handled as a two-step procedure: the endmember extraction step and the inversion step, respectively. In the first step of the analysis, the macroscopic materials that are present in the observed scene are identified by using an endmember extraction algorithm (EEA). The most popular EEAs include pixel purity index (PPI), N-FINDR (Winter 1999), and more recently the VCA algorithm (Nascimento & Bioucas-Dias 2005a) which proposes to recover the vertices of the biggest simplex in the observed data. A common assumption in these EEAs is that they require the presence of pure pixels in the observed image. Conversely, (Craig 1994) and (Bowles *et al.* 1995) proposed minimum volume transforms to recover the smallest simplex that contains all the dataset.

The second step of spectral unmixing is devoted to the abundance estimation. These abundances have to ensure constraints inherent to hyperspectral imagery: as they represent proportions, the abundances have to satisfy positivity and additivity constraints. Several algorithms proposed in the literature to solve this inversion step rely on constrained optimization techniques (Heinz & Chang 2001; Theys *et al.* 2009; Tu *et al.* 1998).

This paper studies alternatives based on Bayesian inference for supervised and unsupervised unmixing problems. In the first part of this work, the endmembers are assumed to be previously identified, *e.g.*, using *a priori* knowledge regarding the observed scene or using results provided by an EEA. In this case, the unmixing algorithm performs the inversion step, *i.e.*, it estimates the abundance coefficients under positivity and additivity constraints. In a second part of this paper, we

introduce a spectral unmixing algorithm in a fully unsupervised framework to estimate the pure component spectra and their proportions jointly.

In both frameworks, Bayesian formulation allows the constraints within the model to be satisfied. Indeed, appropriate prior distributions are chosen to take into account the positivity and additivity of the abundances, as well as the positivity of the endmember spectra. To overcome the complexity of the posterior distribution, Markov chain Monte Carlo algorithms are proposed to approximate the standard minimum mean squared error estimator. Moreover, as the full posterior distribution of all the unknown parameters is available, confidence intervals can be easily computed. These measures allow the accuracy of the different estimates to be quantified.

2 Linear mixing model and problem statement

Let consider P pixels of an hyperspectral image acquired in L spectral bands. According to the linear mixing model, described for instance in (Bioucas-Dias *et al.* 2012), the observed spectrum $\mathbf{y}_p = [y_{p,1}, \ldots, y_{p,L}]^T$ of the pth pixel $(p = 1, \ldots, P)$ is written as an the linear combination of R spectral signatures $\mathbf{m}_r$, corrupted by an additive noise $\mathbf{n}_p$:

$$\mathbf{y}_p = \sum_{r=1}^{R} \mathbf{m}_r a_{p,r} + \mathbf{n}_p, \tag{2.1}$$

where $\mathbf{m}_r = [m_{r,1}, \ldots, m_{r,L}]^T$ is the pure spectrum that is characteristic of the rth material and $a_{p,r}$ is the abundance of the rth material in the pth pixel. Moreover, in (2.1), $\mathbf{n}_p = [n_{p,1}, \ldots, n_{p,L}]^T$ is an noise sequence whose components are assumed to be independent and identically distributed (i.i.d.) according to a centered Gaussian distribution with covariance matrix[4] $\boldsymbol{\Sigma}_{\mathrm{n}} = \sigma^2 \mathbf{I}_L$, where $\mathbf{I}_L$ is the identity matrix of size $L \times L$

$$\mathbf{n}_p | \sigma^2 \sim \mathcal{N}\left(\mathbf{0}_L, \boldsymbol{\Sigma}_{\mathrm{n}}\right). \tag{2.2}$$

Due to physical considerations (Keshava & Mustard 2002), the abundance vectors $\mathbf{a}_p = [a_{p,1}, \ldots, a_{p,R}]^T$ in (2.1) satisfy the following positivity and additivity constraints

$$\begin{cases} a_{p,r} \geq 0, \ \forall r = 1, \ldots, R, \\ \sum_{r=1}^{R} a_{p,r} = 1. \end{cases} \tag{2.3}$$

In other words, the P abundance vectors belong to the space

$$\mathcal{A} = \{\mathbf{a}_p : \ \|\mathbf{a}\|_1 = 1 \text{ and } \mathbf{a}_p \succeq \mathbf{0}\}, \tag{2.4}$$

where $\|\cdot\|_1$ is the ℓ_1 norm such that $\|\mathbf{x}\|_1 = \sum_i |x_i|$, and $\mathbf{a}_p \succeq \mathbf{0}$ stands for the set of inequalities $\{a_{p,r} \geq 0\}_{r=1,\ldots,R}$. In addition, the spectral signatures $\mathbf{m}_r$

---

[4]The proposed model can be easily extended to more complex noise models, following for instance (Dobigeon *et al.* 2008a).

correspond to reflectance measures and, as a consequence, need to ensure the positivity constraints

$$m_{r,l} \geq 0, \ \forall r = 1, \ldots, R, \ \forall l = 1, \ldots, L. \tag{2.5}$$

If we consider all the pixels in the hyperspectral image, the set of Equations (2.1) can be rewritten using the following matrix notations

$$\mathbf{Y} = \mathbf{MA} + \mathbf{N} \tag{2.6}$$

where $\mathbf{Y}$ is a $L \times P$ matrix that contains all the observations associated with the image pixels, $\mathbf{M}$ is the $L \times R$ matrix of the spectral signatures, $\mathbf{A}$ is the $R \times P$ matrix of the abundances and $\mathbf{N}$ is a $L \times P$ matrix of the noise vectors

$$
\begin{aligned}
\mathbf{Y} &= [\mathbf{y}_1, \ldots, \mathbf{y}_P], & \mathbf{M} &= [\mathbf{m}_1, \ldots, \mathbf{m}_R], \\
\mathbf{A} &= [\mathbf{a}_1, \ldots, \mathbf{a}_P], & \mathbf{N} &= [\mathbf{n}_1, \ldots, \mathbf{n}_P].
\end{aligned}
\tag{2.7}
$$

This paper proposes a Bayesian approach to first estimate the abundance coefficient under the constraints (2.3) when the spectral signatures are known. Then, the spectra of the pure components will be assumed unknown and will be included within the estimation procedure.

3   Supervised unmixing: The spectral components are known

When the pure spectral components (also known as *endmembers*) are perfectly known, the problem of linear unmixing reduces to the inversion step, *i.e.*, the constrained estimation of the abundances. This problem can be formulated as a linear regression under constraints whose resolution can be conducted within a Bayesian framework. Indeed, Bayesian models are very convenient in such situation since the constraints are conveniently handled when defining a priori distributions for the unknown parameters. Several constraints have been studied in the literature, including monotony (Chen & Deely 1996), positivity (Moussaoui *et al.* 2006) or sparsity (Blumensath & Davies 2007; Févotte & Godsill 2006). Constraints inherent to hyperspectral imagery are positivity and additivity, as explained in Section 2. In what follows, the Bayesian model to solve the supervised spectral unmixing model is described. Note that in this supervised scenario, spectral unmixing is conducted pixel-by-pixel. As consequence, to lighten the notations, the dependence of the quantity $\mathbf{y}_p, \mathbf{a}_p, \mathbf{c}_p$ on the pixel p will be omitted.

3.1   Bayesian model

3.1.1   Likelihood

The linear mixing model defined by (2.1) and the statistical properties (2.2) of the noise vector $\mathbf{n}$ lead to a Gaussian distribution for the observed spectrum for the pth pixel:

$$\mathbf{y}|\mathbf{a}, \sigma^2 \sim \mathcal{N}\left(\mathbf{Ma}, \sigma^2 \mathbf{I}_L\right). \tag{3.1}$$

As a consequence, the likelihood function of the vector $\mathbf{y}$ can be written

$$f\left(\mathbf{y}\big|\mathbf{a},\sigma^2\right) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{L}{2}} \exp\left[-\frac{\|\mathbf{y}-\mathbf{Ma}\|^2}{2\sigma^2}\right], \tag{3.2}$$

where $\|\mathbf{x}\| = \left(\mathbf{x}^T\mathbf{x}\right)^{\frac{1}{2}}$ is the ℓ_2-norm of the vector $\mathbf{x}$.

3.1.2 Parameter prior distributions

When the spectra of the pure components $\mathbf{m}_1,\ldots,\mathbf{m}_R$ are known, the vector of unknown vector denoted as $\boldsymbol{\theta}$ is composed of the abundance vector and the noise variance $\boldsymbol{\theta} = \left\{\mathbf{a},\sigma^2\right\}$.

Abundance coefficients. For each pixel p, thanks to the additivity constraints enounced in (2.3), the abundance vector $\mathbf{a}$ can be rewritten[5]

$$\mathbf{a} = \left[\begin{array}{c} \mathbf{c} \\ a_R \end{array}\right] \qquad \text{with} \qquad \mathbf{c} = \left[\begin{array}{c} a_1 \\ \vdots \\ a_{R-1} \end{array}\right], \tag{3.3}$$

and $a_R = 1 - \sum_{r=1}^{R-1} a_r$. According to the model proposed in (Dobigeon *et al.* 2008b), the prior distribution chosen for $\mathbf{c}$ is a uniform distribution defined on the simplex $\mathcal{S}$

$$\mathcal{S} = \left\{\mathbf{c};\ \|\mathbf{c}\|_1 \leq 1 \text{ and } \mathbf{c} \succeq \mathbf{0}\right\}. \tag{3.4}$$

Choosing this prior distribution for $\mathbf{c}$ is fully equivalent of choosing a Dirichlet prior $\mathcal{D}\left(1,\ldots,1\right)$ for $\mathbf{a}$, *i.e.*, a uniform distribution on the the set $\mathcal{A}$ of admissible values for $\mathbf{a}$ (defined by (2.4)) (Robert 2007, Appendix A).

Noise variance. A conjugate inverse-gamma distribution is chosen as a prior distribution for the noise variance σ^2

$$\sigma^2\left|\nu,\gamma\right. \sim \mathcal{IG}\left(\frac{\nu}{2},\frac{\gamma}{2}\right), \tag{3.5}$$

where $\mathcal{IG}\left(\frac{\nu}{2},\frac{\gamma}{2}\right)$ is an inverse-gamma distribution of parameters $\frac{\nu}{2}$ and $\frac{\gamma}{2}$. This distribution has been successfully used in several works of the literature, *e.g.*, (Punskaya *et al.* 2002) and (Dobigeon *et al.* 2007). As in the references above, the hyperparameter ν will be fixed to $\nu = 2$.

Moreover, γ is an hyperparameter assumed to be unknown, and a non-informative Jeffreys' distribution is chosen as prior distribution (Jeffreys 1961)

$$f\left(\gamma\right) \propto \frac{1}{\gamma}\mathbf{1}_{\mathbb{R}^+}\left(\gamma\right), \tag{3.6}$$

where $\propto$ stands for "proportional to".

---

[5]For writing conciseness, the last component of $\mathbf{a}$ will be always expressed as a function of the others. Note however that in the algorithm described in the following section, the discarded component can be randomly chosen at each iteration of the Gibbs sampler.

3.1.3   Posterior distribution

The posterior distribution of the unknown parameter vector $\boldsymbol{\theta} = \left\{ \mathbf{c}, \sigma^2 \right\}$ is computed from the following hierarchical model

$$f(\boldsymbol{\theta}|\mathbf{y}) \propto \int f(\mathbf{y}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\gamma) f(\gamma) d\gamma, \qquad (3.7)$$

where $f(\mathbf{y}|\boldsymbol{\theta})$ and $f(\gamma)$ are defined in (3.2) and (3.6), respectively. By assuming prior independence between σ^2 and $\mathbf{c}$, *i.e.* $f(\boldsymbol{\theta}|\gamma) = f(\mathbf{c}) f(\sigma^2|\gamma)$, the hyperparameter γ can be integrated out from the joint distribution $f(\boldsymbol{\theta}, \gamma|\mathbf{y})$ in (3.7), which leads to

$$f\left(\mathbf{c}, \sigma^2|\mathbf{y}\right) \propto \frac{1}{\sigma^{L+2}} \exp\left[-\frac{\|\mathbf{y} - \mathbf{Ma}\|^2}{2\sigma^2} \right] \mathbf{1}_{\mathcal{S}}\left(\mathbf{c}\right). \qquad (3.8)$$

Note that this posterior distribution is defined on the simplex $\mathcal{S} \times \mathbb{R}^+$, *i.e.*, $\mathbf{c}$ satisfies the constraints resulting from the positivity and additivity constraints of $\mathbf{a}$. We introduce in the following section a Gibbs sampler that allows samples to be generated according to the joint distribution $f\left(\mathbf{c}, \sigma^2|\mathbf{y}\right)$.

3.2   Gibbs sampler

Samples (denoted as $\cdot^{(t)}$ where t is the iteration index) can be generated according to $f\left(\mathbf{c}, \sigma^2|\mathbf{y}\right)$ thanks to a Gibbs sampler described below. It successively generates samples according to the conditional distributions $f\left(\mathbf{c}|\sigma^2, \mathbf{y}\right)$ and $f(\sigma^2|\mathbf{c}, \mathbf{y})$. The main steps of this algorithm are detailed below and are summarized by Algo. 1. The interested reader can refer to (Robert & Casella 1999) for more details on MCMC methods.

Algorithm 1 Gibbs sampler for supervised unmixing

1: *% Initialization*
2: Sampling the parameters $\widetilde{\sigma}^{2(0)}$ and $\widetilde{\mathbf{c}}^{(0)}$ from the prior distributions defined in Section 3.1.2,
3: *% Iterations*
4: **for** $t = 1, 2, \ldots,$ **do**
5:     Sampling $\widetilde{\mathbf{c}}^{(t)}$ according to the distribution (3.11),
6:     Sampling $\widetilde{\sigma}^{2(t)}$ according to the distribution (3.12),
7: **end for**

3.2.1   Sampling according to $f\left(\mathbf{c}|\sigma^2, \mathbf{y}\right)$

The conditional posterior distribution of the partial abundance vector is

$$f\left(\mathbf{c}|\sigma^2, \mathbf{y}\right) \propto \exp\left[-\frac{(\mathbf{c} - \boldsymbol{v})^T \boldsymbol{\Sigma}^{-1} (\mathbf{c} - \boldsymbol{v})}{2} \right] \mathbf{1}_{\mathcal{S}}\left(\mathbf{c}\right), \qquad (3.9)$$

where

$$\begin{cases} \boldsymbol{\Sigma} = \left[\left(\mathbf{M}_{\text{-}R} - \mathbf{m}_R \mathbf{1}_{R-1}^T \right)^T \boldsymbol{\Sigma}_{\text{n}}^{-1} \left(\mathbf{M}_{\text{-}R} - \mathbf{m}_R \mathbf{1}_{R-1}^T \right) \right]^{-1}, \\ \boldsymbol{v} = \boldsymbol{\Sigma} \left[\left(\mathbf{M}_{\text{-}R} - \mathbf{m}_R \mathbf{1}_{R-1}^T \right)^T \boldsymbol{\Sigma}_{\text{n}}^{-1} \left(\mathbf{y} - \mathbf{m}_R \right) \right], \end{cases} \quad (3.10)$$

with $\boldsymbol{\Sigma}_{\text{n}}^{-1} = \frac{1}{\sigma^2}\mathbf{I}_L$, $\mathbf{1}_{R-1} = [1, \ldots, 1]^T \in \mathbb{R}^{R-1}$ and where $\mathbf{M}_{\text{-}R}$ is the matrix $\mathbf{M}$ whose Rth column has been removed. As a consequence, the vector $\mathbf{c}|\sigma^2, \mathbf{y}$ is distributed according to a multivariate Gaussian distribution truncated on the simplex $\mathcal{S}$ defined by (3.4)

$$\mathbf{c}\,|\sigma^2, \mathbf{y} \sim \mathcal{N}_{\mathcal{S}}\left(\boldsymbol{v}, \boldsymbol{\Sigma}\right). \quad (3.11)$$

Sampling according to this truncated Gaussian distribution can be conducted following the strategy described in (Dobigeon & Tourneret 2007).

3.2.2 Sampling according to $f(\sigma^2|\mathbf{c}, \mathbf{y})$

By looking at the joint distribution $f\left(\sigma^2, \mathbf{c}|\mathbf{y}\right)$, it can be stated that the conditional distribution of $\sigma^2|\mathbf{c}, \mathbf{y}$ is the following inverse-gamma distribution

$$\sigma^2|\mathbf{c}, \mathbf{y} \sim \mathcal{IG}\left(\frac{L}{2}, \frac{\|\mathbf{y} - \mathbf{Ma}\|^2}{2}\right). \quad (3.12)$$

3.3 Simulation results on synthetic data

To illustrate the algorithm performance, a synthetic mixture of $R = 3$ pure components is generated. These spectral signatures are extracted from the library provided with the ENVI software (RSI (Research Systems Inc.) 2003, p. 1035) and are characteristics of a urban or sub-urban scene: construction concrete, green grass and micaceous loam. The mixing coefficients are defined as $a_1 = 0.3$, $a_2 = 0.6$ and $a_2 = 0.1$. The observed spectrum has been corrupted by an additive Gaussian noise with variance $\sigma^2 = 0.025$, which corresponds to a signal-to-noise ratio RSB $\approx$ 15dB where RSB $= L^{-1}\sigma^{-2} \left\|\sum_{r=1}^{R} \mathbf{m}_r a_r\right\|^2$. The endmembers and the resulting observed spectrum are represented in Figure 1.

Figure 2 shows the posterior distributions of the abundance coefficients a_r ($r = 1, 2, 3$) estimated by the proposed Gibbs sampler for $N_{\text{MC}} = 20\,000$ iterations (with $N_{\text{bi}} = 100$ burn-in iterations). These distributions are in good agreement with the actual values of the coefficients $\mathbf{a} = [0.3, 0.6, 0.1]^T$. As a comparison, the results obtained with the FCLS algorithm (Chang & Ji 2001; Heinz & Chang 2001) have been also depicted in this figure for N_{MC} Monte Carlo simulations (*i.e.*, for N_{MC} realizations of the noise sequence).

3.4 Results on real data

This paragraph presents the analysis of an hyperspectral image that has been widely studied in the literature (Akgun *et al.* 2005; Chen 2005; Christophe *et al.* 2005; Tang & Pearlman 2004). This image, depicted in Figure 3, is initially
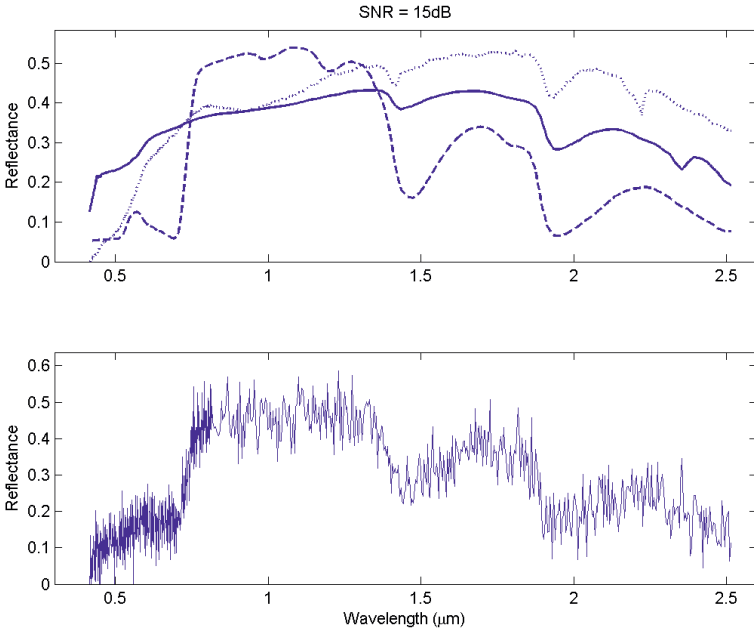
Fig. 1. *Top*: endmember spectra: construction concrete (line), green grass (dashed line), loam (dotted line). *Bottom*: spectrum of the observed pixel.

composed of 189 spectral bands (when the water absorption bands have been removed). It has been acquired by the spectro-imager AVIRIS (Jet Propulsion Lab. (JPL) 2006) in 1997 over Moffett Field, CA. It is composed of a lake and a coastal area. The spectral unmixing algorithm has been applied on a 50×50 scene. The analyzed image area is depicted in Figure 3.

3.4.1   Endmember identification

First, the pure materials that are present in the image have been identified. Since no prior knowledge is available for the analyzed scene, an endmember extraction algorithm has been used to recover to identify the endmember spectra. More precisely, N-FINDR (Winter 1999) has been used to identify $R = 3$ endmembers that are represented in Figure 4: vegetation, water and soil. Note that the number of endmembers has been determined by a principal component analysis, as explained in (Keshava & Mustard 2002).

3.4.2   Abundance estimation

The supervised unmixing algorithm introduced in Sections 3.1 and 3.2 has been applied on each pixel of the AVIRIS hyperspectral image using the endmembers
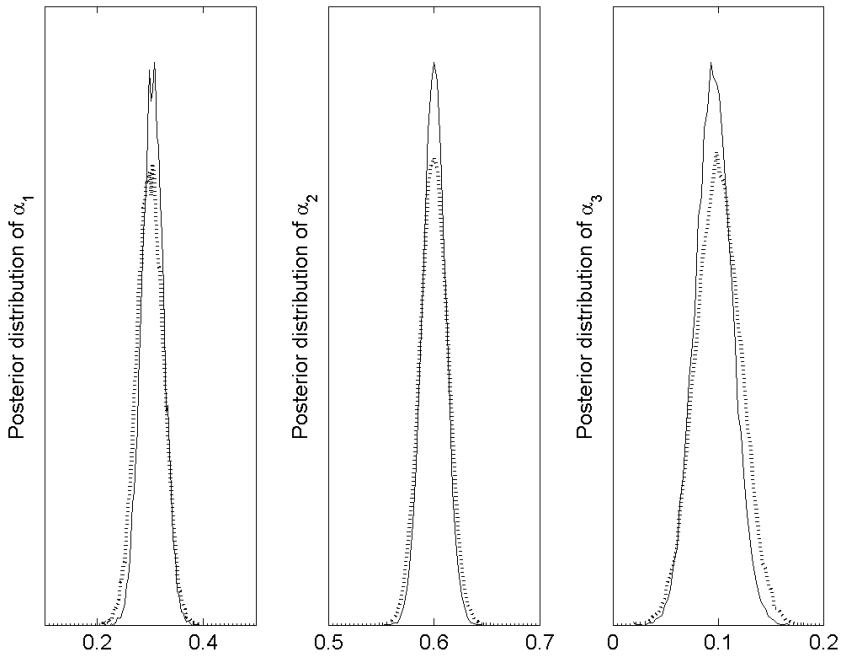
Fig. 2. Posterior distributions of the abundance coefficients $[a_1, a_2, a_3]^T$ estimated by the proposed algorithm (continuous lines) and histograms of the estimated values by FCLS (dashed lines).

previously identified. The abundance maps estimated by the proposed algorithm for the $R = 3$ materials are depicted in Figure 5 (top). Note that a white (resp. black) pixel corresponds to a high (resp. low) proportion of the corresponding material. The lake area (that appears as white pixels on the water abundance map) has been clearly recovered. The results obtained with the unmixing algorithm provided with the ENVI software (RSI (Research Systems Inc.) 2003, p. 739) are also depicted in Figure 5 (bottom). These results obtained with constrained least square algorithm are in good agreement with those of Figure 5 (top). Note however that the proposed algorithm also allows posterior distributions to be estimated. These posterior distributions can be useful to derive confidence intervals.

4 Unsupervised unmixing

As explained in (Bioucas-Dias *et al.* 2012; Keshava & Mustard 2002), linear spectral unmixing has been often addressed in a two-step procedure: i) endmember identification by an EEA and ii) abundance estimation. However, solving the unmixing problem in two distinct and successive steps may lead to poor estimation performance. In particular, when no pure pixels are present in the image,
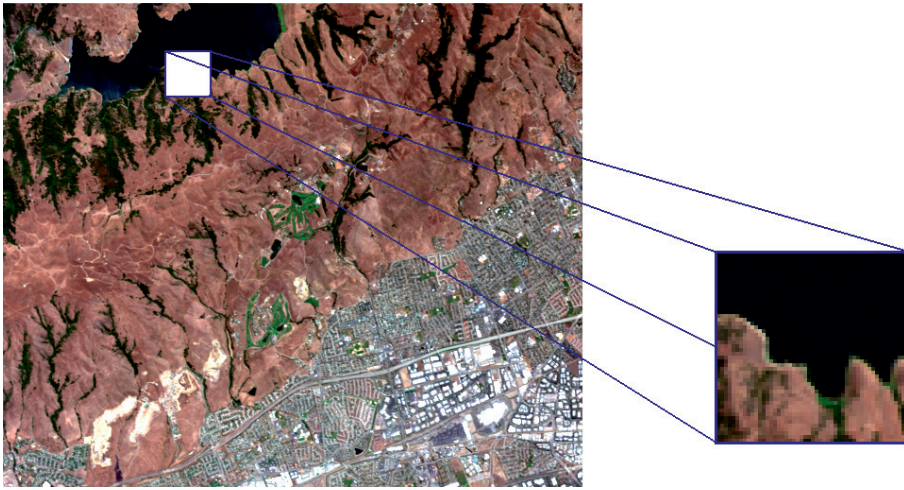
Fig. 3. Real hyperspectral image acquired by AVIRIS over Moffett Field in 1997 (*left*) and the region of interest (*right*).
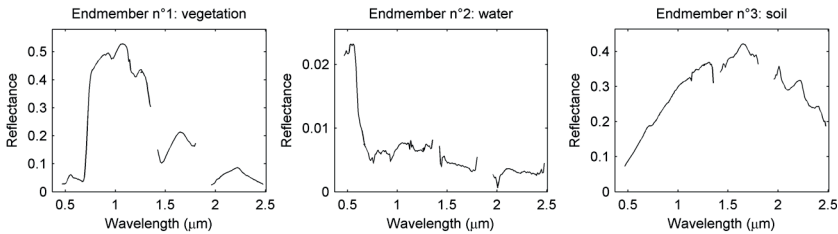

Fig. 4. The $R = 3$ endmember spectra obtained by N-FINDR.

the geometric EEA such as VCA, N-FINDR or PPI provide inadequate endmember estimates. To overcome this issue, we propose to solve the linear unmixing problem in a fully unsupervised Bayesian framework, by estimating the endmember spectra and the corresponding abundances jointly. This approach casts the unmixing problem as an blind source separation (BSS), that as received a huge interest in the signal processing literature. In particular, it is well known that independent component analysis (ICA) (Hyvärinen *et al.* 2001) is a powerful solution of BSS problems. However, as noticed in (Nascimento & Bioucas-Dias 2005b) and (Dobigeon & Achard 2005), ICA-based algorithms fails to solve the unmixing problem, mainly due to the high correlation between the source signals. Other strategies, based on non-negative matrix factorization techniques (Paatero & Tapper 1994), can be used to jointly estimate the endmember spectra and the abundance coefficients. However, these algorithms do not take explicitly into account the sum-to-one constraint on the abundance coefficients. Conversely, the Bayesian framework is a convenient way to ensure all the constraints (positivity
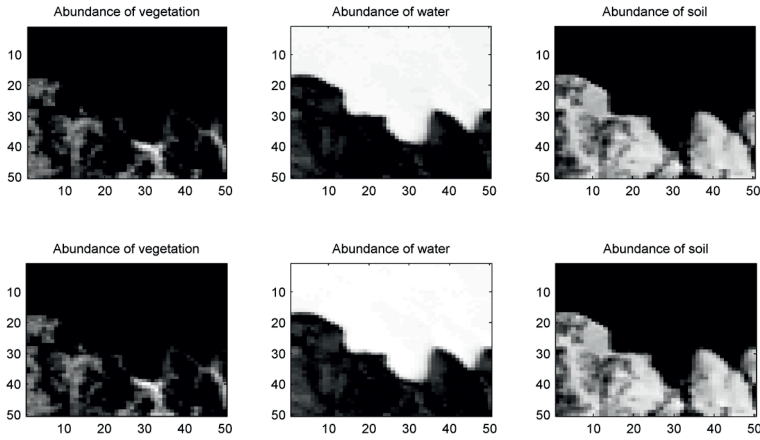
Fig. 5. *Top*: abundance maps estimated by the proposed algorithm. *Bottom*: abundance maps estimated by the unmixing routine provided with the ENVI software.

on the abundance coefficients and endmember spectra, additivity on the abundance coefficients) by defining appropriate prior distributions for the unknown parameters.

Moreover, a geometrical interpretation of the linear unmixing problem allows one to show that the spectral signatures can be estimated in an appropriate lower-dimensional subspace. This estimation in a subspace allows the number of degree of freedom to be significantly reduced for the parameters, while ensuring the physical constraints.

4.1   Bayesian model

Unsupervised spectral unmixing can be formulated as a blind source separation problem. Thus, the joint estimation of the endmembers and the abundances requires to consider all the image pixels $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_P]$ during the analysis. From a pixel-wise analysis in Section 3.1, spectral unmixing is now conducted on a whole hyperspectral image. More specifically, the previous Bayesian model introduced in 3.1 is extended by defining a prior model for the endmember spectra. The new posterior distribution associated with the new set of unknown parameters is finally derived.

4.1.1   Likelihood function

By assuming the independence of the noise vector, the new likelihood function associated with the observed pixel matrix $\mathbf{Y}$ is the product of the marginal likelihood functions (3.1.1)

$$f\left(\mathbf{Y}|\mathbf{M}, \mathbf{C}\right) = \prod_{p=1}^{P} f\left(\mathbf{y}_p|\mathbf{M}, \mathbf{c}_p\right)$$

where $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_P]^T$ is a matrix coming from the reparametrization (3.3) of the abundance vectors and $f(\mathbf{y}_p | \mathbf{M}, \mathbf{c}_p)$ has been defined in (3.1.1).

4.1.2   Prior of the abundance coefficients

The coefficient vectors $\mathbf{c}_1, \ldots, \mathbf{c}_P$ are assumed to be *a priori* independent. Thus, the prior distribution for the coefficient matrix $\mathbf{C}$ can be written as the product of the prior chosen in paragraph 3.1.2

$$f(\mathbf{C}) = \prod_{p=1}^{P} f(\mathbf{c}_p)$$

with

$$f(\mathbf{c}_p) \propto \mathbf{1}_{\mathcal{S}}(\mathbf{c}_p)$$

where $\mathcal{S}$ has been defined in (3.4). This prior allows the constraints inherent to the linear mixing model to be ensured. Moreover, this prior has the great advantage of imposing a constraint on the size of the simplex spanned by the endmembers in the hyperspectral space. Indeed, as demonstrated in (Dobigeon *et al.* 2009), among two admissible solutions for the unmixing problem, this prior will favor the solution that corresponds to the simplex of minimum volume. Note that this property has been exploited also in (Arngren *et al.* 2011; Bowles *et al.* 1995; Craig 1994).

4.1.3   Prior model for the endmembers

Dimensionality reduction.   First, notice that the set

$$\mathcal{S}_{\mathbf{M}} = \left\{ \mathbf{x} \in \mathbb{R}^L; \ \mathbf{x} = \sum_{r=1}^{R} \lambda_r \mathbf{m}_r, \ \sum_{r=1}^{R} \lambda_r = 1, \ \lambda_r \geq 0 \right\}$$

is a convex polytope of $\mathbb{R}^L$ whose vertices $\mathbf{m}_1, \ldots, \mathbf{m}_R$ are the $R \ll L$ spectral signatures to be estimated. As a consequence, the unobserved data $\mathbf{X} = \mathbf{MA} = \mathbf{Y} - \mathbf{N}$ can be represented in a lower-dimensional subspace $\mathcal{V}_K$ of $\mathbb{R}^K$ with $R - 1 \leq K \ll L$ without any loss of information. In this subspace, the noise-free data $\mathbf{X}$ span a $(R-1)$-simplex whose vertices are the projections of the endmembers. As stated in (Keshava & Mustard 2002), dimensional reduction is a classical step while performing spectral unmixing, required by numerous EEAs, such as N-FINDR (Winter 1999) and PPI (Boardman 1993). In this paper, we propose to estimate the projections $\mathbf{t}_r$ $(r = 1, \ldots, R)$ of the spectral signatures $\mathbf{m}_r$ onto the subspace $\mathcal{V}_K$. This approach allows the number of degrees of freedom to be significantly reduced. We assume that this subspace has been previously estimated by a dimensional reduction technique (*e.g.*, principal component analysis, PCA).

PCA-based dimensional reduction. The empirical covariance matrix of $\boldsymbol{\Upsilon}$ of the observed data $\mathbf{Y}$ is

$$\boldsymbol{\Upsilon} = \frac{1}{P} \sum_{p=1}^{P} (\mathbf{y}_p - \bar{\mathbf{y}}) (\mathbf{y}_p - \bar{\mathbf{y}})^T \tag{4.1}$$

where $\bar{\mathbf{y}}$ is the empirical mean

$$\bar{\mathbf{y}} = \frac{1}{P} \sum_{p=1}^{P} \mathbf{y}_p. \tag{4.2}$$

Let

$$\begin{cases} \mathbf{D} = \operatorname{diag}\left(\lambda_1, \ldots, \lambda_K\right), \\ \mathbf{V} = \left[\mathbf{v}_1, \ldots, \mathbf{v}_K\right]^T \end{cases} \tag{4.3}$$

denote the diagonal matrix of the K highest eigenvalues and the corresponding matrix of eigenvectors of $\boldsymbol{\Upsilon}$, respectively. The projected vector $\mathbf{t}_r \in \mathbb{R}^K$ of the endmember spectrum $\mathbf{m}_r \in \mathbb{R}^L$ is then obtained by the affine transformation

$$\mathbf{t}_r = \mathbf{P} \left(\mathbf{m}_r - \bar{\mathbf{y}}\right) \tag{4.4}$$

where $\mathbf{P} = \mathbf{D}^{-\frac{1}{2}} \mathbf{V}$. Equivalently,

$$\mathbf{m}_r = \mathbf{U} \mathbf{t}_r + \bar{\mathbf{y}} \tag{4.5}$$

where $\mathbf{U} = \mathbf{V}^T \mathbf{D}^{\frac{1}{2}}$. Note that in the subspace $\mathcal{V}_{R-1}$ obtained for $K = R - 1$, the vectors $\{\mathbf{t}_r\}_{r=1,\ldots,R}$ span a simplex that the classical EEAs (*e.g.*, N-FINDR Winter 1999, MVT Craig 1994 and ICE Berman *et al.* 2004) try to estimate. We propose to estimate the projected vertices $\mathbf{t}_r$ $(r = 1, \ldots, R)$ of this simplex in a Bayesian setting. The prior distributions assigned to the projections $\mathbf{t}_r$ $(r = 1, \ldots, R)$ are detailed in the following paragraph.

Prior distributions of the projected endmembers. The spectral signature $\mathbf{m}_r \in \mathbb{R}^L$ and its projection $\mathbf{t}_r \in \mathbb{R}^K$ onto $\mathcal{V}_K$ are related by $\mathbf{t}_r = \mathbf{P} \left(\mathbf{m}_r - \bar{\mathbf{y}}\right)$ and $\mathbf{m}_r = \mathbf{U} \mathbf{t}_r + \bar{\mathbf{y}}$, where $\mathbf{P}$ is a projection matrix, $\mathbf{U}$ is the pseudo-inverse of $\mathbf{P}$ and $\bar{\mathbf{y}}$ is the empirical mean of the observations. The prior distributions chosen for the endmember spectra should be chosen such that the endmember spectra satisfy the positivity constraints (2.5). Straightforward computations allows the space $\mathcal{T}_r \subset \mathcal{V}_K$ to be identified such that

$$\{m_{l,r} \geq 0, \ \forall l = 1, \ldots, L\} \quad \Leftrightarrow \quad \{\mathbf{t}_r \in \mathcal{T}_r\} \tag{4.6}$$

thanks to the L following inequalities

$$\mathcal{T}_r = \left\{ \mathbf{t}_r; \bar{y}_l + \sum_{k=1}^{K} u_{l,k} t_{k,r} \geq 0, l = 1, \ldots, L \right\}, \tag{4.7}$$

One of the originality of the proposed blind source separation method consists of defining prior distributions for the endmember projections $\mathbf{t}_r$ onto the subspace $\mathcal{V}_K$ instead of the endmembers $\mathbf{m}_r$ themselves. More precisely, a multivariate Gaussian distribution

$$\mathbf{t}_r \sim \mathcal{N}_{\mathcal{T}_r}\left(\mathbf{e}_r, s_r^2 \mathbf{I}_K\right) \tag{4.8}$$

truncated to the set $\mathcal{T}_r$ is chosen as prior distribution for each vector $\mathbf{t}_r$ ($r = 1, \ldots, R$). The mean vectors $\mathbf{e}_r$ of these prior distributions are fixed to some values corresponding to the solutions provided by fast EEA, such as N-FINDR and VCA. In absence of additional prior information, the variances s_r^2 ($r = 1, \ldots, R$) are fixed to large values $s_1^2 = \ldots = s_R^2 = 50$, which allows some deviations to be modeled between the actual endmember projections $\mathbf{t}_r$ and the crude estimations $\mathbf{e}_r$ provided by N-FINDR or VCA.

4.1.4   Posterior distribution

Following the the Bayes rule, the prior distributions of unknown parameters defined in the paragraphs 4.1.3 and 4.1.2, associated with the likelihood function defined in paragraph 4.1.1, lead to the following joint posterior distribution

$$
\begin{aligned}
f\left(\mathbf{C}, \mathbf{T}, \sigma^2 \big| \mathbf{Y}\right) \propto & \prod_{r=1}^{R} \exp\left[-\frac{\|\mathbf{t}_r - \mathbf{e}_r\|^2}{2s_r^2}\right] \mathbf{1}_{\mathcal{T}_r}\left(\mathbf{t}_r\right) \\
& \times \prod_{p=1}^{P}\left[\left(\frac{1}{\sigma^2}\right)^{\frac{L}{2}+1} \exp\left(-\frac{\|\mathbf{y}_p - (\mathbf{U}\mathbf{T} + \bar{\mathbf{y}}\mathbf{1}_{R-1})\mathbf{a}_p\|^2}{2\sigma^2}\right)\right] \\
& \times \prod_{p=1}^{P} \mathbf{1}_{\mathcal{S}}\left(\mathbf{c}_p\right).
\end{aligned}
\tag{4.9}
$$

Since the standard Bayesian estimators (*e.g.*, minimum mean square error (MMSE) or maximum *a posteriori* (MAP) estimators) are difficult to derive from (4.9), a Gibbs algorithm, detailed in the following paragraph, allows samples $\left\{\mathbf{C}^{(t)}, \mathbf{T}^{(t)}, \sigma^{2(t)}\right\}$ to be generated according to this distribution. These samples are then used to approximate the Bayesian estimators.

4.2   Gibbs sampler

The Gibbs sampler that allows samples to be asymptotically distributed according to the posterior (4.9) is detailed below. This algorithm is similar to the Gibbs sampler introduced in paragraph 3.2 (Algo. 1) with an additional step that consists of sampling according to the conditional distribution $f\left(\mathbf{T}|\mathbf{C}, \sigma^2, \mathbf{Y}\right)$ (see also Algo. 2).

Algorithm 2 Gibbs sampler for **unsupervised** unmixing

1: *% Pre-processing*
2: Computing the empirical mean $\bar{\mathbf{y}}$ following (4.2),
3: Computing the matrices $\mathbf{D}$ and $\mathbf{V}$ following (4.3) thanks to a PCA,
4: Set $\mathbf{U} = \left(\mathbf{V}^T\mathbf{V}\right)^{-1}\mathbf{V}^T\mathbf{D}^{\frac{1}{2}}$,
5: Choose the crude estimations $\mathbf{e}_r \in \mathcal{V}_K$ required in (4.8),
6: *% Initialization*
7: **for** $r = 1, \ldots, R$ **do**
8:     Sampling $\mathbf{t}_r^{(0)}$ according to (4.8),
9:     Set $\mathbf{m}_r^{(0)} = \mathbf{U}\mathbf{t}_r^{(0)} + \bar{\mathbf{y}}$,
10: **end for**
11: Sampling $\sigma^{2(0)}$ according to (3.5),
12: *% Iterations*
13: **for** $t = 1, 2, \ldots,$ **do**
14:     **for** $p = 1, \ldots, P$ **do**
15:         Sampling $\mathbf{c}_p^{(t)}$ according to (4.11),
16:     **end for**
17:     **for** $r = 1, \ldots, R$ **do**
18:         **for** $k = 1, \ldots, K$ **do**
19:             Sampling $t_{k,r}^{(t)}$ according to (4.15),
20:         **end for**
21:         Set $\mathbf{m}_r^{(t)} = \mathbf{U}\mathbf{t}_r^{(0)} + \bar{\mathbf{y}}$,
22:     **end for**
23:     Sampling $\sigma^{2(t)}$ according to (4.17).
24: **end for**

4.2.1   Sampling according to $f\left(\mathbf{C}|\mathbf{T}, \sigma^2, \mathbf{Y}\right)$

For each pixel p, as in paragraph 3.2.1, the conditional distribution of the coefficient vector $\mathbf{c}_p$ is

$$f\left(\mathbf{c}_p \left|\mathbf{T}, \sigma^2, \mathbf{y}_p\right.\right) \exp\left[-\frac{\left(\mathbf{c}_p - \boldsymbol{v}_p\right)^T \boldsymbol{\Sigma}_p^{-1}\left(\mathbf{c}_p - \boldsymbol{v}_p\right)}{2}\right]\mathbf{1}_{\mathcal{S}}\left(\mathbf{c}_p\right), \qquad (4.10)$$

where $\boldsymbol{\Sigma}_p$ and $\boldsymbol{v}_p$ have been defined in 3.2.1. Consequently, the vector $\mathbf{c}_p|\mathbf{T}, \sigma^2, \mathbf{y}_p$ is distributed according to a multivariate Gaussian distribution truncated onto the simplex $\mathcal{S}$ defined by (3.4)

$$\mathbf{c}_p\left|\mathbf{T}, \sigma^2, \mathbf{y}_p \sim \mathcal{N}_{\mathcal{S}}\left(\boldsymbol{v}_p, \boldsymbol{\Sigma}_p\right).\right. \qquad (4.11)$$

4.2.2 Sampling according to $f\left(\mathbf{T}|\mathbf{C}, \sigma^2, \mathbf{Y}\right)$

Let $\mathbf{T}_{-r}$ denote the $\mathbf{T}$ whose rth column has been removed. The conditional posterior distribution of $\mathbf{t}_r$ $(r = 1, \ldots, R)$ is

$$f\left(\mathbf{t}_r|\mathbf{T}_{-r}, \mathbf{c}_r, \sigma^2, \mathbf{Y}\right) \propto \exp\left[-\frac{1}{2}(\mathbf{t}_r - \boldsymbol{\tau}_r)^T \boldsymbol{\Lambda}_r^{-1}(\mathbf{t}_r - \boldsymbol{\tau}_r)\right]\mathbf{1}_{\mathcal{T}_r}(\mathbf{t}_r), \qquad (4.12)$$

with

$$\begin{cases} \boldsymbol{\Lambda}_r = \left[\displaystyle\sum_{p=1}^{P} a_{p,r}^2 \mathbf{U}^T \boldsymbol{\Sigma}_{\mathrm{n}}^{-1}\mathbf{U} + \frac{1}{s_r^2}\mathbf{I}_K\right]^{-1}, \\[2.5em] \boldsymbol{\tau}_r = \boldsymbol{\Lambda}_r\left[\displaystyle\sum_{p=1}^{P} a_{p,r}\mathbf{U}^T \boldsymbol{\Sigma}_{\mathrm{n}}^{-1}\boldsymbol{\epsilon}_{p,r} + \frac{1}{s_r^2}\mathbf{e}_r\right], \end{cases} \qquad (4.13)$$

and

$$\boldsymbol{\epsilon}_{p,r} = \mathbf{y}_p - a_{p,r}\bar{\mathbf{y}} - \sum_{j \neq r} a_{p,j}\mathbf{m}_j. \qquad (4.14)$$

Generating vectors distributed according to this distribution is not straightforward, mainly due to the truncature on the set $\mathcal{T}_r$. One alternative strategy consists of generating each component $t_{k,r}$ of $\mathbf{t}_r$ conditionally upon the others $\mathbf{t}_{-k,r} = \{t_{j,r}\}_{j \neq k}$. By denoting $\mathcal{U}_k^+ = \{l; u_{l,k} > 0\}$, $\mathcal{U}_k^- = \{l; u_{l,k} < 0\}$ and $\varepsilon_{l,k,r} = \bar{y}_l + \sum_{j \neq k} u_{l,j}t_{j,r}$, it follows

$$t_{k,r}|\mathbf{t}_{-k,r}, \mathbf{T}_{-r}, \mathbf{c}_r, \sigma^2, \mathbf{Y} \sim \mathcal{N}_{\left[t_{k,r}^-, t_{k,r}^+\right]}\left(w_{k,r}, z_{k,r}^2\right), \qquad (4.15)$$

with

$$\begin{cases} t_{k,r}^- = \displaystyle\max_{l \in \mathcal{U}_k^+} -\frac{\varepsilon_{l,k,r}}{u_{l,k}}, \\[1.5em] t_{k,r}^+ = \displaystyle\min_{l \in \mathcal{U}_k^-} -\frac{\varepsilon_{l,k,r}}{u_{l,k}}, \end{cases} \qquad (4.16)$$

where $w_{k,r}$ and $z_{k,r}^2$ are the conditional mean and variance computed following (Kay 1988, p. 324) (see also similar computations in Dobigeon & Tourneret 2007). Generating samples according to the truncated Gaussian distribution (4.15) can be performed using various strategies, such as (Robert 1995).

4.2.3 Sampling according to $f\left(\sigma^2|\mathbf{C}, \mathbf{T}, \mathbf{Y}\right)$

The conditional distribution of $\sigma^2|\mathbf{C}, \mathbf{T}, \mathbf{Y}$ is the inverse-gamma distribution

$$\sigma^2|\mathbf{C}, \mathbf{T}, \mathbf{Y} \sim \mathcal{IG}\left(\frac{PL}{2}, \frac{1}{2}\sum_{p=1}^{P}\|\mathbf{y}_p - \mathbf{M}\mathbf{a}_p\|^2\right). \qquad (4.17)$$

4.3 Simulation results on synthetic data

To illustrate the performance of the proposed method, the algorithm has been applied on a 100×100-pixel image, where $R = 3$ spectral signatures have been linearly mixed: construction concrete, green grass, red bare brick. These signatures have been measured in $L = 413$ spectral bands and are depicted in Figure 6 (top, in black). These materials have been linearly mixed with random proportions (ensuring the sum-to-one and positivity constraints), with an i.i.d. noise corresponding to signal-to-noise ratio SNR = 15dB.
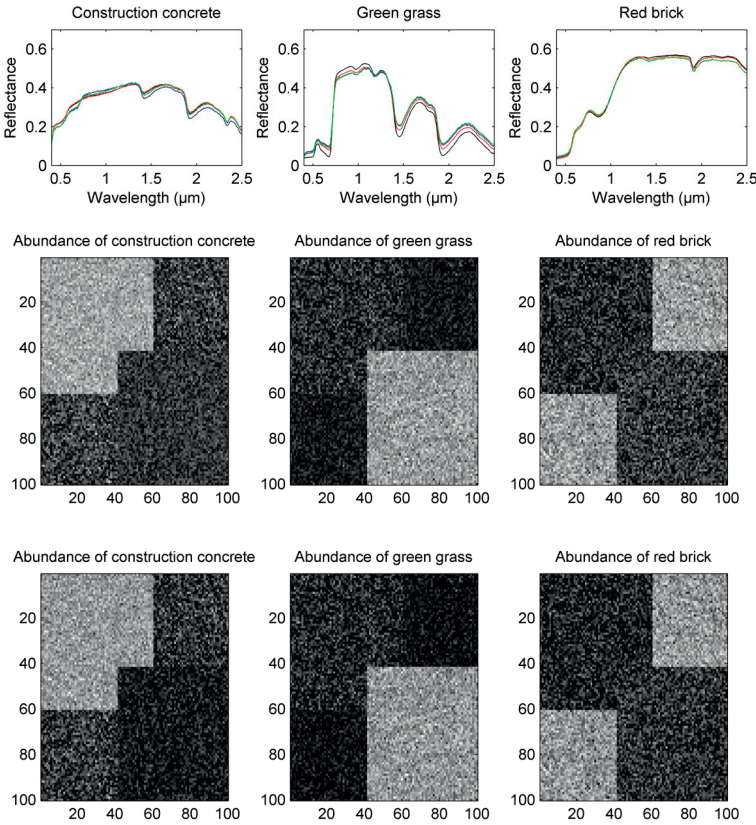
Fig. 6. *Top*: actual spectra (black), spectra estimated by N-FINDR (blue), estimated by VCA (green) and estimated by the proposed approach (red). *Middle* and *bottom*: actual and estimated abundance maps.

The estimation results for the spectral signatures obtained by the proposed algorithm, depicted in Figure 6 (top, in red) have been compared with the results provided by two geometrical EEAs: VCA and N-FINDR. Table 1 (top) reports the mean square errors defined by

$$\mathrm{MSE}_r^2 = \|\hat{\mathbf{m}}_r - \mathbf{m}_r\|^2, \quad r = 1, \dots, R. \tag{4.18}$$

The results regarding the estimation of the 10^4 abundance vectors (see Fig. 6, bottom), are reported in the Table 1 (bottom) in terms of mean square errors for each component:

$$\mathrm{MSE}_r^2 = \sum_{p=1}^{P} (\hat{a}_{p,r} - a_{p,r})^2, \quad r = 1, \ldots, R, \tag{4.19}$$

where $\hat{a}_{p,r}$ is the estimated abundance coefficient for the $\#r$ material in the $\#p$ pixel. These results demonstrate that the proposed algorithm provides better estimation performance than the two other algorithms.

Table 1. Estimation performance comparison between the algorithms VCA, N-FINDR and the proposed Bayesian approach: mean square errors between the $R = 3$ estimated and actual spectra (*top*), global mean square errors between the estimated and actual abundances (*bottom*).

Spectra	Bayesian algo.	VCA	N-FINDR
Endmember #1	0.10	1.29	0.54
Endmember #2	2.68	15.59	5.19
Endmember #3	0.16	4.35	0.57
Abundances	Bayesian algo.	VCA	N-FINDR
Endmember #1	25.68	57.43	30.66
Endmember #2	29.97	74.48	46.45
Endmember #3	3.19	83.02	11.22

4.4 Results on real data

The proposed algorithm is finally applied to the Moffett Field image introduced in Section 3.4. The $R = 3$ endmembers identified by the Bayesian algorithm are depicted in Figure 7 (top). The corresponding estimated abundance maps are represented in Figure 7 (bottom). Both results are in good agreement with those of Figures 4 and 5 obtained using a supervised unmixing approach.

5 Conclusion

This article presented two Bayesian algorithms to solve the problem of linear unmixing of hyperspectral images in supervised and unsupervised frameworks. For each scenario, suitable prior distributions were assigned to the unknown parameters. In particular, these distributions were chosen to ensure constraints inherent to the mixing model: positivity and additivity for the abundance coefficients and positivity for the endmember spectra. MCMC algorithms were designed to generate samples distributed according to the posterior distribution of the unknown parameters. Simulation results, obtained on synthetic and real hyperspectral images, demonstrated the interest of the proposed methods. Both of the strategies
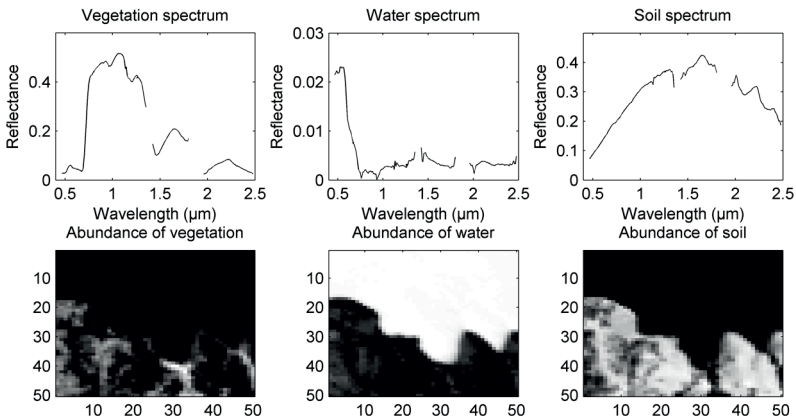
Fig. 7. *Top*: the $R = 3$ endmembers estimated by the unsupervised algorithm in the Moffett Field image. *Bottom*: the corresponding estimated abundance maps.

detailed in this article ignore any spatial correlations between the observed pixels. To improve the unmixing performance, intrinsic dependencies between the parameters of interest, *e.g.*, the abundance vectors, could be exploited. Extending the previous approaches, a hidden Markov model has been introduced in (Eches *et al.* 2011). Conversely, Mittelman *et al.* have proposed a nonparametric Bayesian algorithm to jointly unmix and classify hyperspectral images (Mittelman *et al.* 2012). Future works also include the design of efficient unmixing algorithms to analyze hyperspectral data resulting from non-linear mixtures. Encouraging results have been obtained in (Altmann *et al.* 2012; Halimi *et al.* 2011).

References

Akgun, T., Altunbasak, Y., & Mersereau, R.M., 2005, IEEE Trans. Image Process., 14, 1860

Altmann, Y., Halimi, A., Dobigeon, N., & Tourneret, J.-Y., 2012, IEEE Trans. Image Process., 21, 3017

Arngren, M., Schmidt, M.N., & Larsen, J., 2011, J. Signal Proc. Syst., 65, 479

Berman, M., Kiiveri, H., Lagerstrom, R., Ernst, A., Dunne, R., & Huntington, J.F., 2004, IEEE Trans. Geosci. Remote Sens., 42, 2085

Bioucas-Dias, J.M., Plaza, A., Dobigeon, N., *et al.*, 2012, IEEE J. Sel. Topics Appl. Earth Observations Remote Sens., 5, 354

Blumensath, T., & Davies, M.E., 2007, IEEE Trans. Signal Process., 55, 4474

Boardman, J., 1993, in Summaries 4th Annu. JPL Airborne Geoscience Workshop, Vol. 1 (JPL Pub., Washington, D.C.), 11

Bowles, J.H., Palmadesso, P.J., Antoniades, J.A., Baumback, M.M., & Rickard, L.J., 1995, ed. M. Strojnik & B.F. Andresen, Infrared Spaceborne Remote Sensing III, SPIE, 2553, 148

Chang, C.-I., 2003, Hyperspectral Imaging: Techniques for Spectral detection and classification (Kluwer, New York)

Chang, C.-I., & Ji, B., 2001, IEEE Trans. Geosci. Remote Sensing, 44, 378

Chen, F.W., 2005, IEEE Geosci. Remote Sensing Lett., 2, 64

Chen, M.-H., & Deely, J.J., 1996, J. Agricultural, Biological Environmental Stat., 1, 467

Christophe, E., Léger, D., & Mailhes, C., 2005, IEEE Trans. Geosci. Remote Sensing, 43, 2103

Craig, M., 1994, IEEE Trans. Geosci. Remote Sens., 542

Dobigeon, N., & Achard, V., 2005, ed. L. Bruzzone, Image and Signal Processing for Remote Sensing XI, SPIE, 5982, 335

Dobigeon, N., Moussaoui, S., Coulon, M., Tourneret, J.-Y., & Hero, A.O., 2009, IEEE Trans. Signal Process., 57, 4355

Dobigeon, N., & Tourneret, J.-Y., 2007, Efficient sampling according to a multivariate Gaussian distribution truncated on a simplex, Technical report, IRIT/ENSEEIHT/TéSA, France

Dobigeon, N., Tourneret, J.-Y., & Hero III, A.O., 2008a, in Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP, Las Vegas, USA), 3433

Dobigeon, N., Tourneret, J.-Y., & Chang, C.-I., 2008b, IEEE Trans. Signal Process., 56, 2684

Dobigeon, N., Tourneret, J.-Y., & Davy, M., 2007, IEEE Trans. Signal Process., 55, 1251

Eches, O., Dobigeon, N., & Tourneret, J.Y., 2011, IEEE Trans. Geosci. Remote Sensing, 49, 4239

Févotte, C., & Godsill, S.J., 2006, IEEE Trans. Audio, Speech, Language Process., 14, 2174

Halimi, A., Altmann, Y., Dobigeon, N., & Tourneret, J.-Y., 2011, IEEE Trans. Geosci. Remote Sensing, 49, 4153

Hapke, B.W., 1981, J. Geophys. Res., 86, 3039

Heinz, D.C., & Chang, C.-I., 2001, IEEE Trans. Geosci. Remote Sens., 29, 529

Hyvärinen, A., Karhunen, J., & Oja, E., 2001, Independent Component Analysis (John Wiley, New York)

Jackson, Q., & Landgrebe, D.A., 2002, IEEE Trans. Geosci. Remote Sens., 40, 1082

Jeffreys, H., 1961, Theory of Probability, 3 edition (Oxford University Press, London)

Jet Propulsion Lab. (JPL), 2006, AVIRIS Free Data

Johnson, P.E., Smith, M.O., Taylor-George, S., & Adams, J.B., 1983, J. Geophys. Res., 88, 3557

Kay, S.M., 1988, Modern spectral estimation (Prentice Hall)

Keshava, N., & Mustard, J.F., 2002, IEEE Signal Process. Mag., 19, 44

Manolakis, D., Siracusa, C., & Shaw, G., 2001, IEEE Trans. Geosci. Remote Sens., 39, 1392

Mittelman, R., Dobigeon, N., & Hero III, A.O., 2012, IEEE Trans. Signal Process., 60, 1656

Moussaoui, S., Brie, D., Mohammad-Djafari, A., & Carteret, C., 2006, IEEE Trans. Signal Process., 54, 4133

Nascimento, J.M., & Bioucas-Dias, J.M., 2005a, IEEE Trans. Geosci. Remote Sens., 43, 898

Nascimento, J.M.P., & Bioucas-Dias, J.M., 2005b, IEEE Trans. Geosci. Remote Sens., 43, 175

Paatero, P., & Tapper, U., 1994, Environmetrics, 5, 111

Plaza, J., Pérez, R., Plaza, A., Martínez, P., & Valencia, D., 2005, ed. J.O. Jensen & J.-M. Thériault, Chemical and Biological Standoff Detection III, SPIE, 5995, 79

Punskaya, E., Andrieu, C., Doucet, A., & Fitzgerald, W., 2002, IEEE Trans. Signal Process., 50, 747

Rellier, G., Descombes, X., Falzon, F., & Zerubia, J., 2004, IEEE Trans. Geosci. Remote Sens., 42, 1543

Robert, C.P., 1995, Stat. Comput., 5, 121

Robert, C.P., 2007, The Bayesian Choice: from Decision-Theoretic Motivations to Computational Implementation, 2 edition (Springer Texts in Statistics, Springer-Verlag, New York)

Robert, C.P., & Casella, G., 1999, Monte Carlo Statistical Methods (Springer-Verlag, New York)

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

Tang, X., & Pearlman, W.A., 2004, Proc. IEEE Int. Conf. Image Proc. (ICIP), 5, 3283

Theys, C., Dobigeon, N., Tourneret, J.-Y., & Lantéri, H., 2009, in Proc. IEEE-SP Workshop Stat. and Signal Processing (SSP) (Cardiff, UK), 729

Tu, T.M., Chen, C.H., & Chang, C.-I., 1998, IEEE Trans. Geosci. Remote Sens., 36, 171

Winter, M., 1999, in Proc. 13th Int. Conf. Appl. Geologic Remote Sensing, 2, 337, Vancouver