Subspace-based Bayesian blind source separation for hyperspectral imagery

Nicolas Dobigeon*, Saïd Moussaoui†, Martial Coulon*, Jean-Yves Tourneret* and Alfred O. Hero‡

*University of Toulouse, IRIT/INP-ENSEEIHT, 2 rue Charles Camichel, BP 7122, 31071 Toulouse cedex 7, France
†IRCCyN - CNRS UMR 6597, ECN, 1 rue de la Noë, BP 92101, 44321 Nantes Cedex 3, France
‡University of Michigan, Department of EECS, Ann Arbor, MI 48109-2122, USA

Abstract—In this paper, a fully Bayesian algorithm for endmember extraction and abundance estimation for hyperspectral imagery is introduced. Following the linear mixing model, each pixel spectrum of the hyperspectral image is decomposed as a linear combination of pure endmember spectra. The estimation of the unknown endmember spectra and the corresponding abundances is conducted in a unified manner by generating the posterior distribution of the unknown parameters under a hierarchical Bayesian model. The proposed model accounts for non-negativity and full-additivity constraints, and exploits the fact that the endmember spectra lie on a lower dimensional space. A Gibbs algorithm is proposed to generate samples distributed according to the posterior of interest. Simulation results illustrate the accuracy of the proposed joint Bayesian estimator.

I. INTRODUCTION

Over the last years, the spectral unmixing problem has been considered by many researchers. Spectral unmixing consists of decomposing an observed pixel spectrum into a collection of pure spectra, usually referred to as *endmembers*, and estimating the proportions or *abundances* of each material in the image pixels [1]. To describe the mixture, the most frequently encountered model is the linear mixing model (LMM) which gives a good approximation in the reflective spectral domain ranging from 0.4μm to 2.5μm. It assumes that the observed pixel spectrum is a weighted linear combination of the endmember spectra.

Spectral unmixing has often been handled as a two-step procedure: i) the endmember extraction step dedicated to the identification of the macroscopic materials that are present in the observed scene and ii) the inversion step which consists of estimating the proportions of the materials previously identified. This paper proposes an algorithm that estimates the endmember spectra and their respective abundances jointly. This approach casts spectral unmixing as a blind source separation (BSS) problem. The Bayesian model studied in this paper uses a Gibbs sampling algorithm to efficiently solve the constrained spectral unmixing problem without requiring the presence of pure pixels in the hyperspectral image. In many works, Bayesian estimation approaches have been adopted to solve BSS problems like spectral unmixing. The Bayesian formulation allows one to directly incorporate constraints into the model. These constraints include sparsity [2]; non-negativity [3]; full additivity (sum-to-one constraint) [4]. In this paper, prior distributions are proposed for the abundances and endmember spectra to enforce the constraints inherent to the hyperspectral mixing model. These constraints include non-negativity and full-additivity of the abundance coefficients (as in [4]) and non-negativity of the endmember spectra.

Moreover, the proposed joint spectral unmixing approach is able to solve the endmember spectrum estimation problem directly on a lower dimensional space within a Bayesian framework. We believe that this is one of the principal factors leading to performance improvements that we show in Section V. The problem of hyperparameter selection in our Bayesian model is circumvented by adopting the hierarchical Bayesian approach of [4] that produces a parameter-independent Bayesian posterior distribution for the endmember spectra and abundances. To overcome the complexity of the full posterior distribution, a Gibbs sampling strategy is derived to approximate standard Bayesian estimators, e.g., the minimum mean squared error (MMSE) estimator. Moreover, as the full posterior distribution of all the unknown parameters is available, confidence intervals can be easily computed. These measures allow one to quantify the accuracy of the different estimates.

II. LINEAR MIXING MODEL AND PROBLEM STATEMENT

Consider P pixels of an hyperspectral image acquired in L spectral bands. According to the linear mixing model (LMM), described for instance in [1], the L-spectrum $\boldsymbol{y}_p = [y_{p,1}, \ldots, y_{p,L}]^T$ of the pth pixel ($p = 1, \ldots, P$) is assumed to be a linear combination of R spectra $\boldsymbol{m}_r$ corrupted by an additive Gaussian noise

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

where $\boldsymbol{m}_r = [m_{r,1}, \ldots, m_{r,L}]^T$ denotes the spectrum of the rth material, $a_{p,r}$ is the fraction of the rth material in the pth observation, R is the number of materials, L is the number of available spectral bands and P is the number of observations (pixels). Moreover, in (1), $\boldsymbol{n}_p = [n_{p,1}, \ldots, n_{p,L}]^T$ is an additive noise sequence which is assumed to be an independent and identically distributed (i.i.d.) zero-mean Gaussian sequence with covariance matrix $\boldsymbol{\Sigma}_\mathrm{n} = \sigma^2 \mathbf{I}_L$, where $\mathbf{I}_L$ is the identity matrix of dimension $L \times L$, i.e.,

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

Finally, note that the model in (1) can be easily modified (see [5] and [4]).

Due to physical considerations [1], the fraction vectors $\boldsymbol{a}_p = [a_{p,1}, \ldots, a_{p,R}]^T$ in (1) satisfy the following non-negativity and full-additivity (or *sum-to-one*) constraints

$$\left\{ \begin{array}{l} a_{p,r} \geq 0, \ \forall r = 1, \ldots, R, \\ \sum_{r=1}^{R} a_{p,r} = 1. \end{array} \right. \tag{3}$$

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

$$\mathcal{A} = \left\{ \boldsymbol{a} : \ \|\boldsymbol{a}\|_1 = 1 \text{ and } \boldsymbol{a} \succeq \mathbf{0} \right\} \tag{4}$$

where $\|\cdot\|_1$ is the ℓ_1 norm defined as $\|\boldsymbol{x}\|_1 = \sum_i |x_i|$, and $\boldsymbol{a} \succeq \mathbf{0}$ stands for the set of inequalities $\{a_r \geq 0\}_{r=1,\ldots,R}$. Moreover, the endmember spectra component $m_{r,l}$ must satisfy the following non-negativity constraints

$$m_{r,l} \geq 0, \ \forall r = 1, \ldots, R, \ \forall l = 1, \ldots, L. \tag{5}$$

Considering all pixels, standard matrix notation yields $\boldsymbol{Y} = \boldsymbol{M}\boldsymbol{A} + \boldsymbol{N}$ where $\boldsymbol{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_P]$, $\boldsymbol{M} = [\boldsymbol{m}_1, \ldots, \boldsymbol{m}_R]$,

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
DEC 2009		**00-00-2009 to 00-00-2009**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Subspace-based Bayesian blind source separation for hyperspectral imagery	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
University of Michigan,Department of Electrical and Computer Engineering,Ann Arbor,MI,48109-2122	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Third International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (IEEE CAMSAP, 2009), 13-16 Dec, Aruba, Dutch Antilles. U.S. Government or Federal Rights License

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE	**Same as Report (SAR)**	**4**	
unclassified	**unclassified**	**unclassified**			

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

$\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_P]$ and $\boldsymbol{N} = [\boldsymbol{n}_1, \ldots, \boldsymbol{n}_P]$. In this work, we propose to estimate $\boldsymbol{A}$ and $\boldsymbol{M}$ from the noisy observations $\boldsymbol{Y}$ under the constraints in (3) and (5).

III. BAYESIAN MODEL

A. Likelihood

The linear mixing model defined in (1) and the statistical properties in (2) of the noise vector $\boldsymbol{n}_p$ result in a conditionally Gaussian distribution for the observation of the pth pixel: $\boldsymbol{y}_p | \boldsymbol{M}, \boldsymbol{a}_p, \sigma^2 \sim \mathcal{N}\left(\boldsymbol{M}\boldsymbol{a}_p, \sigma^2 \mathbf{I}_L\right)$. Assuming independence between the noise sequences $\boldsymbol{n}_p$ ($p = 1, \ldots, P$), the likelihood function of all the observations $\boldsymbol{Y}$ is

$$f\left(\boldsymbol{Y} | \boldsymbol{M}, \boldsymbol{A}, \sigma^2\right) = \prod_{p=1}^{P} \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{L}{2}} \exp\left[-\frac{\|\boldsymbol{y}_p - \boldsymbol{M}\boldsymbol{a}_p\|^2}{2\sigma^2}\right],$$
(6)

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

B. Prior model for the endmember spectra

1) Dimensionality reduction: It is interesting to note that the unobserved matrix $\boldsymbol{X} = \boldsymbol{M}\boldsymbol{A} = \boldsymbol{Y} - \boldsymbol{N}$ is rank deficient under the linear model (1). Consequently, in the noise-free case, $\boldsymbol{X}$ can be represented in a suitable lower-dimensional subset $\mathcal{V}_K$ of $\mathbb{R}^K$ ($R - 1 \leq K \leq L$) without loss of information. As noted in [1], dimensionality reduction is a common step of the spectral unmixing, adopted by numerous endmember extraction algorithms (EEAs), such as N-FINDR [6] or PPI [7]. Similarly, we propose to estimate the projection $\boldsymbol{t}_r$ ($r = 1, \ldots, R$) of the endmember spectra $\boldsymbol{m}_r$ in the subspace $\mathcal{V}_K$. The identification of this subspace can be achieved via a standard dimension reduction procedure. In the sequel, we propose to define $\mathcal{V}_K$ as the subspace spanned by K orthogonal axes $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K$ identified by a principal component analysis (PCA) on the observations $\boldsymbol{Y}$

$$\mathcal{V}_K = \text{span}\left(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K\right).$$
(7)

2) PCA projection: If $\boldsymbol{D}$ and $\boldsymbol{V}$ denote the diagonal matrix of the K highest eigenvalues of the empirical covariance matrix and the corresponding eigenvector matrix, respectively, the PCA projection $\boldsymbol{t}_r \in \mathbb{R}^K$ of the endmember spectrum $\boldsymbol{m}_r \in \mathbb{R}^L$ is

$$\boldsymbol{t}_r = \boldsymbol{P}\left(\boldsymbol{m}_r - \bar{\boldsymbol{y}}\right)$$
(8)

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

$$\boldsymbol{m}_r = \boldsymbol{U}\boldsymbol{t}_r + \bar{\boldsymbol{y}}$$
(9)

with $\boldsymbol{U} = \boldsymbol{V}^T \boldsymbol{D}^{\frac{1}{2}}$. Note that in the subspace $\mathcal{V}_{R-1}$ obtained for $K = R - 1$, the vectors $\{\boldsymbol{t}_r\}_{r=1,\ldots,R}$ form a simplex that standard EEAs try to recover. In this paper, we estimate the vertices $\boldsymbol{t}_r$ ($r = 1, \ldots, R$) of this simplex using a Bayesian approach. The Bayesian prior distributions for the projections $\boldsymbol{t}_r$ ($r = 1, \ldots, R$) are introduced in the following paragraph.

3) Prior distribution for the projected spectra: To ensure non-negativity constraints (5) of the corresponding reconstructed $L \times 1$ spectra $\boldsymbol{m}_r$, a conjugate multivariate Gaussian distribution (MGD) $\mathcal{N}_{\mathcal{T}_r}\left(\boldsymbol{e}_r, s_r^2 \mathbf{I}_K\right)$ truncated on the set $\mathcal{T}_r$ is chosen as prior distribution for $\boldsymbol{t}_r$, assumed to be *a priori* independent. The set $\mathcal{T}_r \subset \mathcal{V}_K$ is explicitly defined in [8] and has the following property

$$\{m_{l,r} \geq 0, \ \forall l = 1, \ldots, L\} \quad \Leftrightarrow \quad \{\boldsymbol{t}_r \in \mathcal{T}_r\}.$$
(10)

This paper proposes to select the *a priori* mean vectors $\boldsymbol{e}_r$ ($r = 1, \ldots, R$) as the projected spectra of pure components previously

identified by an EEA, e.g., N-FINDR. The variances s_r^2 ($r = 1, \ldots, R$) reflect the degree of confidence given to this prior information. When no additional knowledge is available, these variances are fixed to large values.

C. Abundance prior

For each observed pixel p, with the full additivity constraint in (3), the abundance vectors $\boldsymbol{a}_p$ ($p = 1, \ldots, P$) can be rewritten as

$$\boldsymbol{a}_p = \begin{bmatrix} \boldsymbol{c}_p \\ a_{p,R} \end{bmatrix} \quad \text{with} \quad \boldsymbol{c}_p = \begin{bmatrix} a_{p,1} \\ \vdots \\ a_{p,R-1} \end{bmatrix},$$

and $a_{p,R} = 1 - \sum_{r=1}^{R-1} a_{p,r}$. Following the model in [4], the priors chosen for $\boldsymbol{c}_p$ ($p = 1, \ldots, P$) are uniform distributions on the simplex $\mathcal{S}$ defined by

$$\mathcal{S} = \left\{\boldsymbol{c}_p; \ \|\boldsymbol{c}_p\|_1 \leq 1 \text{ and } \boldsymbol{c}_p \succeq \boldsymbol{0}\right\}.$$
(11)

Under the assumption of statistical independence between the abundance vectors $\boldsymbol{c}_p$ ($p = 1, \ldots, P$), the full prior distribution for partial abundance matrix $\boldsymbol{C} = [\boldsymbol{c}_1, \ldots, \boldsymbol{c}_P]^T$ can be written

$$f\left(\boldsymbol{C}\right) \propto \prod_{p=1}^{P} \mathbf{1}_{\mathcal{S}}\left(\boldsymbol{c}_p\right).$$
(12)

As noted in [4], the uniform prior distribution reflects a lack of *a priori* knowledge about the abundance vector. However, as demonstrated in [8], among two *a priori* equiprobable solutions of the BSS problem, the uniform prior allows one to favor *a posteriori* the solution corresponding to the polytope in the projection subset $\mathcal{V}_K$ having smallest volume.

D. Noise variance prior

A conjugate inverse-gamma distribution is chosen as prior for σ^2

$$\sigma^2 | \nu, \gamma \sim \mathcal{IG}\left(\frac{\nu}{2}, \frac{\gamma}{2}\right),$$
(13)

where the hyperparameter ν will be fixed to $\nu = 2$ and γ will be a random and adjustable hyperparameter, whose prior distribution is defined below.

E. Prior distribution for hyperparameter γ

The prior for γ is a non-informative Jeffreys' prior which reflects the lack of knowledge regarding this hyperparameter

$$f\left(\gamma\right) \propto \frac{1}{\gamma} \mathbf{1}_{\mathbb{R}^+}\left(\gamma\right).$$
(14)

F. Posterior distribution

The posterior distribution of the unknown parameter vector $\boldsymbol{\theta} = \left\{\boldsymbol{C}, \boldsymbol{T}, \sigma^2\right\}$ can be computed from marginalization using the following hierarchical structure

$$f(\boldsymbol{\theta}|\boldsymbol{Y}) = \int f(\boldsymbol{\theta}, \gamma | \boldsymbol{Y}) d\gamma \propto \int f(\boldsymbol{Y}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\gamma) f(\gamma) d\gamma \quad (15)$$

where $f\left(\boldsymbol{Y}|\boldsymbol{\theta}\right)$ and $f\left(\gamma\right)$ are defined in (6) and (14) respectively. Moreover, under the assumption of *a priori* independence between $\boldsymbol{C}, \boldsymbol{T}$ and σ^2, the following result can be obtained

$$f\left(\boldsymbol{\theta}|\gamma\right) = f\left(\boldsymbol{C}\right) f\left(\boldsymbol{T} \mid \boldsymbol{e}, \mathbf{s}^2\right) f\left(\sigma^2 \mid \nu, \gamma\right)$$
(16)

where $f(C)$, $f(T \mid e, \mathbf{s}^2)$ and $f(\sigma^2 \mid \nu, \gamma)$ have been previously defined. This hierarchical structure allows one to integrate out the hyperparameter γ from the joint distribution $f(\boldsymbol{\theta}, \gamma | \boldsymbol{Y})$, yielding

$$
\begin{aligned}
f\left(\boldsymbol{C}, \boldsymbol{T}, \sigma^2 \big| \boldsymbol{Y}\right) &\propto \prod_{p=1}^{P} \mathbf{1}_{\mathcal{S}}\left(\boldsymbol{c}_p\right) \\
&\times \prod_{r=1}^{R} \exp\left[-\frac{\|\boldsymbol{t}_r - \boldsymbol{e}_r\|^2}{2s_r^2}\right] \mathbf{1}_{\mathcal{T}_r}\left(\boldsymbol{t}_r\right) \\
&\times \prod_{p=1}^{P}\left[\left(\frac{1}{\sigma^2}\right)^{\frac{L}{2}+1} \exp\left(-\frac{\left\|\boldsymbol{y}_p - \left(\boldsymbol{U}\boldsymbol{T} + \bar{y}\mathbf{1}_R^T\right)\boldsymbol{a}_p\right\|^2}{2\sigma^2}\right)\right]
\end{aligned}
\tag{17}
$$

where $\mathbf{1}_R = [1, \ldots, 1]^T \in \mathbb{R}^R$. Deriving the Bayesian estimators (e.g., MMSE or MAP) from the posterior distribution in (17) remains intractable. In such case, it is very common to use Markov chain Monte Carlo (MCMC) methods to generate samples asymptotically distributed according to the posterior distribution. The Bayesian estimators can then be approximated using these samples. The next section studies a Gibbs sampling strategy allowing one to generate samples distributed according to (17).

IV. GIBBS SAMPLER

Random samples (denoted by $\cdot^{(t)}$ where t is the iteration index) can be drawn from $f\left(\boldsymbol{C}, \boldsymbol{T}, \sigma^2 \mid \boldsymbol{Y}\right)$ using a Gibbs sampler [9]. This MCMC technique consists of generating samples $\left\{\boldsymbol{C}^{(t)}, \boldsymbol{T}^{(t)}, \boldsymbol{\sigma}^{2(t)}\right\}$ distributed according to the conditional posterior distributions of each parameter.

A. Sampling from $f\left(\boldsymbol{C}|\boldsymbol{T}, \sigma^2, \boldsymbol{Y}\right)$

Straightforward computations yield for each observation

$$
\begin{aligned}
&f\left(\boldsymbol{c}_p \big| \boldsymbol{T}, \sigma^2, \boldsymbol{y}_p\right) \\
&\propto \exp\left[-\frac{\left(\boldsymbol{c}_p - \boldsymbol{v}_p\right)^T \boldsymbol{\Sigma}_p^{-1}\left(\boldsymbol{c}_p - \boldsymbol{v}_p\right)}{2}\right] \mathbf{1}_{\mathcal{S}}\left(\boldsymbol{c}_p\right), \quad (18)
\end{aligned}
$$

where

$$
\begin{cases}
\boldsymbol{\Sigma}_p = \left[\left(\boldsymbol{M}_{\text{-}R} - \boldsymbol{m}_R\mathbf{1}_{R-1}^T\right)^T \boldsymbol{\Sigma}_{\text{n}}^{-1}\left(\boldsymbol{M}_{\text{-}R} - \boldsymbol{m}_R\mathbf{1}_{R-1}^T\right)\right]^{-1}, \\
\boldsymbol{v}_p = \boldsymbol{\Sigma}_p\left[\left(\boldsymbol{M}_{\text{-}R} - \boldsymbol{m}_R\mathbf{1}_{R-1}^T\right)^T \boldsymbol{\Sigma}_{\text{n}}^{-1}\left(\boldsymbol{y}_p - \boldsymbol{m}_R\right)\right],
\end{cases}
\tag{19}
$$

with $\boldsymbol{\Sigma}_{\text{n}}^{-1} = \frac{1}{\sigma^2}\mathbf{I}_L$ and where $\boldsymbol{M}_{\text{-}R}$ denotes the matrix $\boldsymbol{M}$ whose Rth column has been removed. As a consequence, $\boldsymbol{c}_p\big|\boldsymbol{T}, \sigma^2, \boldsymbol{y}_p$ is distributed according to an MGD truncated on the simplex $\mathcal{S}$ in (11)

$$
\boldsymbol{c}_p \big| \boldsymbol{T}, \sigma^2, \boldsymbol{y}_p \sim \mathcal{N}_{\mathcal{S}}\left(\boldsymbol{v}_p, \boldsymbol{\Sigma}_p\right). \tag{20}
$$

Note that samples can be drawn from an MGD truncated on a simplex using efficient Monte Carlo simulation strategies described in [10].

B. Sampling from $f\left(\boldsymbol{T}|\boldsymbol{C}, \sigma^2, \boldsymbol{Y}\right)$

Define $\boldsymbol{T}_{\text{-}r}$ as the matrix $\boldsymbol{T}$ whose rth column has been removed. Then the conditional posterior distribution of $\boldsymbol{t}_r$ $(r = 1, \ldots, R)$ is

$$
\begin{aligned}
&f\left(\boldsymbol{t}_r|\boldsymbol{T}_{\text{-}r}, \boldsymbol{c}_r, \sigma^2, \boldsymbol{Y}\right) \propto \\
&\exp\left[-\frac{1}{2}\left(\boldsymbol{t}_r - \boldsymbol{\tau}_r\right)^T \boldsymbol{\Lambda}_r^{-1}\left(\boldsymbol{t}_r - \boldsymbol{\tau}_r\right)\right] \mathbf{1}_{\mathcal{T}_r}\left(\boldsymbol{t}_r\right), \quad (21)
\end{aligned}
$$

Endm.	Region #1		Region #2		Region #3	
	mean	var.	mean	var.	mean	var.
#1	0.60	0.01	0.25	0.01	0.25	0.02
#2	0.20	0.02	0.50	0.01	0.15	0.005
#3	0.20	0.01	0.25	0.02	0.60	0.02

with

$$
\begin{cases}
\boldsymbol{\Lambda}_r = \left[\sum_{p=1}^{P} a_{p,r}^2 \boldsymbol{U}^T \boldsymbol{\Sigma}_{\text{n}}^{-1}\boldsymbol{U} + \frac{1}{s_r^2}\mathbf{I}_K\right]^{-1}, \\
\boldsymbol{\tau}_r = \boldsymbol{\Lambda}_r\left[\sum_{p=1}^{P} a_{p,r}\boldsymbol{U}^T\boldsymbol{\Sigma}_{\text{n}}^{-1}\boldsymbol{\epsilon}_{p,r} + \frac{1}{s_r^2}\boldsymbol{e}_r\right],
\end{cases}
\tag{22}
$$

and

$$
\boldsymbol{\epsilon}_{p,r} = \boldsymbol{y}_p - a_{p,r}\bar{\boldsymbol{y}} - \sum_{j\neq r} a_{p,j}\boldsymbol{m}_j. \tag{23}
$$

Note that $\boldsymbol{m}_j = \boldsymbol{U}\boldsymbol{t}_j + \bar{\boldsymbol{y}}$. As a consequence, the posterior distribution of $\boldsymbol{t}_r$ is the following truncated MGD

$$
\boldsymbol{t}_r \mid \boldsymbol{T}_{\text{-}r}, \boldsymbol{c}_r, \sigma^2, \boldsymbol{Y} \sim \mathcal{N}_{\mathcal{T}_r}\left(\boldsymbol{\tau}_r, \boldsymbol{\Lambda}_r\right). \tag{24}
$$

C. Sampling from $f\left(\sigma^2|\boldsymbol{C}, \boldsymbol{T}, \boldsymbol{Y}\right)$

The conditional distribution of $\sigma^2|\boldsymbol{C}, \boldsymbol{T}, \boldsymbol{Y}$ is the following inverse Gamma distribution:

$$
\sigma^2|\boldsymbol{C}, \boldsymbol{T}, \boldsymbol{Y} \sim \mathcal{IG}\left(\frac{PL}{2}, \frac{1}{2}\sum_{p=1}^{P}\left\|\boldsymbol{y}_p - \boldsymbol{M}\boldsymbol{a}_p\right\|^2\right). \tag{25}
$$

V. SIMULATIONS ON SYNTHETIC DATA

To illustrate the accuracy of the proposed algorithm, simulations are conducted on a 100×100 synthetic image. This hyperspectral image is composed of three different regions with $R = 3$ pure materials representative of a suburban scene: construction concrete, green grass and red brick. The spectra of these endmembers have been extracted from the spectral libraries distributed with the ENVI software [11] and are represented in Fig. 1 (top, black lines). The reflectances are observed in $L = 413$ spectral bands ranging from $0.4\mu m$ to $2.5\mu m$. These $R = 3$ components have been mixed with proportions that have been randomly generated according to MGDs truncated on the simplex $\mathcal{S}$ with means and variances reported in Table I. The generated abundance maps have been depicted in Fig. 2 (top) in gray scale where a white (resp. black) pixel stands for the presence (resp. absence) of the material. The signal-to-noise ratio has been tuned to $\text{SNR}_{\text{dB}} = 15\text{dB}$.

The resulting hyperspectral data have been unmixed by the proposed algorithm. First, the space $\mathcal{V}_K$ in (7) has been identified by PCA as discussed in paragraph III-B2. The hidden mean vectors $\boldsymbol{e}_r$ $(r = 1, \ldots, R)$ of the normal distributions introduced in paragraph (III-B) have been chosen as the PCA projections of endmembers previously identified by N-FINDR. The hidden variances s_r^2 have all been chosen equal to $s_1^2 = \ldots = s_R^2 = 50$ to obtain vague priors (i.e. large variances). The Gibbs sampler has been run with $N_{\text{MC}} = 1300$ iterations, including $N_{\text{bi}} = 300$ burn-in iterations. The MMSE estimates of the abundance vectors $\boldsymbol{a}_p$ $(p = 1, \ldots, P)$ and the projected spectra $\boldsymbol{t}_r$ $(r = 1, \ldots, R)$ have been approximated by computing empirical averages over the last computed outputs of the sampler. The corresponding endmember spectra estimated
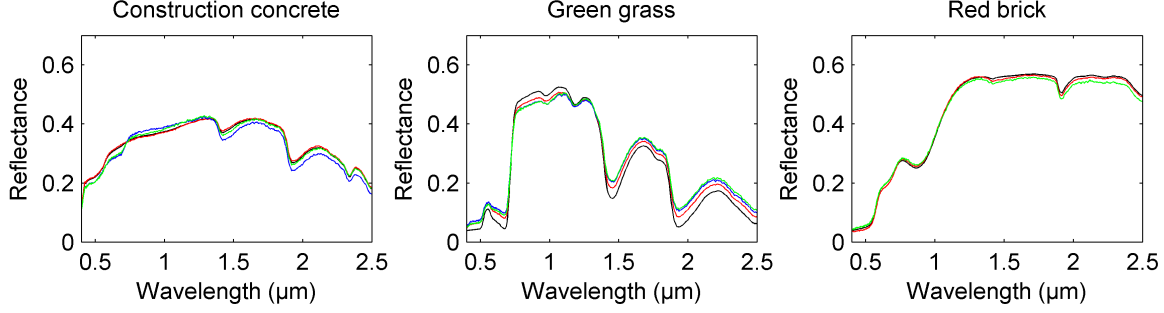
Fig. 1. Actual endmembers (black lines), endmembers estimated by N-FINDR (blue lines), endmembers estimated by VCA (green lines) and endmembers estimated by proposed approach (red lines).

by the proposed algorithm are depicted in Fig. 1 (top, red lines). The proposed algorithm clearly outperforms N-FINDR and VCA, as shown in Fig. 1.

Moreover, the MMSE estimated abundance maps are depicted in Fig. 2 (bottom) and are clearly in good agreement with the simulated maps (top). Note that the proposed Bayesian estimation provides the joint posterior distribution of the unknown parameters. Specifically, this posterior distribution allows one to derive confidence intervals regarding the parameters of interest.
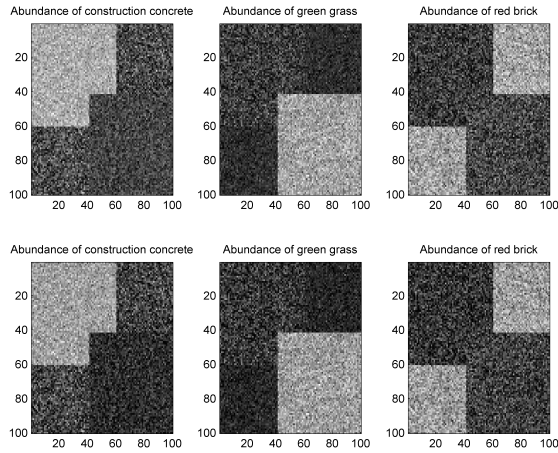
Fig. 2. Top: actual endmember abundance maps. Bottom: estimated end-member abundance maps.

VI. CONCLUSIONS

This paper addressed the unsupervised unmixing problem of hyperspectral images, i.e. estimating the endmember spectra in the observed scene and their respective abundances for each pixel. A Bayesian model as well as an MCMC algorithm was introduced, based on appropriate priors for the abundance vectors to ensure non-negativity and sum-to-one constraints inherent to the linear mixing model. Instead of estimating the endmember spectral signatures in the observation space, we proposed to estimate their projections onto a suitable subspace. In this subspace, these projections were assigned priors that satisfy positivity constraints on the reconstructed endmember spectra. A Gibbs sampling scheme was proposed to generate samples asymptotically distributed according to this posterior. The available samples were then used to approximate the Bayesian estimators for the different parameters of interest. Results of simulations conducted on synthetic hyperspectral images illustrated the accuracy of the proposed Bayesian method when compared with other algorithms from the literature.

REFERENCES

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

[2] N. Dobigeon, S. Moussaoui, J.-Y. Tourneret, and C. Carteret, "Bayesian separation of spectral sources under non-negativity and full additivity constraints," *Sig. Proc.*, vol. 89, no. 12, pp. 2657–2669, Dec. 2009.

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

[4] N. Dobigeon, J.-Y. Tourneret, and C.-I Chang, "Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery," *IEEE Trans. Signal Processing*, vol. 56, no. 7, pp. 2684–2695, July 2008.

[5] N. Dobigeon and J.-Y. Tourneret, "Bayesian sampling of structured noise covariance matrix for hyperspectral imagery," University of Toulouse, Tech. Rep., Dec. 2008. [Online]. Available: http://dobigeon.perso.enseeiht.fr/publis.html

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

[7] J. Boardman, "Automating spectral unmixing of AVIRIS data using convex geometry concepts," in *Summaries 4th Annu. JPL Airborne Geoscience Workshop*, vol. 1. Washington, D.C.: JPL Pub., 1993, pp. 11–14.

[8] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A. O. Hero, "Joint bayesian endmember extraction and linear unmixing for hyperspectral imagery," *IEEE Trans. Signal Processing*, vol. 57, no. 11, Nov. 2009.

[9] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. New York: Springer-Verlag, 1999.

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

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