Joint Bayesian Endmember Extraction and Linear Unmixing for Hyperspectral Imagery

Nicolas Dobigeon, *Member, IEEE*, Saïd Moussaoui, Martial Coulon, Jean-Yves Tourneret, *Senior Member, IEEE*, and Alfred O. Hero, *Fellow, IEEE*

***Abstract*—This paper studies a fully Bayesian algorithm for endmember extraction and abundance estimation for hyperspectral imagery. Each pixel of the hyperspectral image is decomposed as a linear combination of pure endmember spectra following the linear mixing model. The estimation of the unknown endmember spectra is conducted in a unified manner by generating the posterior distribution of abundances and endmember parameters under a hierarchical Bayesian model. This model assumes conjugate prior distributions for these parameters, accounts for nonnegativity and full-additivity constraints, and exploits the fact that the endmember proportions lie on a lower dimensional simplex. A Gibbs sampler is proposed to overcome the complexity of evaluating the resulting posterior distribution. This sampler generates samples distributed according to the posterior distribution and estimates the unknown parameters using these generated samples. The accuracy of the joint Bayesian estimator is illustrated by simulations conducted on synthetic and real AVIRIS images.**

***Index Terms*—Bayesian inference, endmember extraction, hyperspectral imagery, linear spectral unmixing, MCMC methods.**

I. Introduction

OVER the last several decades, much research has been devoted to the spectral unmixing problem. Spectral unmixing is an efficient way to solve standard problems encountered in hyperspectral imagery. These problems include pixel classification [1], material quantification [2] and subpixel detection [3]. Spectral unmixing consists of decomposing a pixel spectrum into a collection of material spectra, usually referred to as *endmembers*, and estimating the corresponding proportions or *abundances* [4]. To describe the mixture, the most frequently encountered model is the macroscopic model which gives a good approximation in the reflective spectral domain ranging from 0.4 μm to 2.5 μm [5]. The linearization of the nonlinear intimate model proposed by Hapke in [6] results in this macroscopic model [7]. The macroscopic model assumes that the observed pixel spectrum is a weighted linear combination of the endmember spectra.

As reported in [4], linear spectral mixture analysis (LSMA) has often been handled as a two-step procedure: the endmember extraction step and the inversion step, respectively. In the first step of 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) [8] and N-FINDR [9], that apply a linear model for the observations with nonnegativity and full-additivity[1] constraints. This model results in endmember spectra located on the vertices of a lower dimensional simplex. PPI and N-FINDR estimate this simplex by identifying the largest simplex contained in the data. Another popular alternative, called vertex component analysis (VCA) has been proposed in [10]. A common assumption in VCA, PPI and N-FINDR is that they require pure pixels to be present in the observed scene, where pure pixels are pixels composed of a single endmember. Alternatively, Craig [11] and Bowles [12] have proposed minimum volume transforms (MVT) to find the smallest simplex that contains all the pixels [11]. However, these MVT-based methods are not fully automated techniques: they provide results that strongly depend on i) the algorithm initialization, ii) some *ad hoc* parameters that have to be selected by the user. More recently, a new MVT algorithm has been introduced in [13]. This minimum volume simplex analysis provides a suboptimal solution of the nonconvex optimization problem. More generally, the MVT approaches avoids the difficult problem of direct parameter estimation on the simplex. Furthermore, as mentioned in [13], the minimum volume simplex analysis (MVSA) provides a suboptimal solution of the nonconvex optimization problem. Note also that the performance of these approaches may be negatively affected by the presence of outliers and noise. The interested reader is invited to consult [14] and [15] for a recent performance comparison of some standard EEAs. The second step in LSMA, called the *inversion* step, consists of estimating the proportions of the materials identified by EEA [16]. The inversion step can use various strategies such as least square estimation [17], maximum likelihood estimation [18] and Bayesian estimation [19].

The central premise of this paper is to propose an algorithm that estimates the endmember spectra and their respective abundances jointly in a single step. This approach casts LSMA as a

Manuscript received January 15, 2009; accepted May 12, 2009. First published September 22, 2009; current version published October 14, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Mark J. Coates. Part of this work was supported by the GdR-ISIS/CNRS, a DGA fellowship from French Ministry of Defence, and by an AFOSR Grant FA9550-06-1-0324.

N. Dobigeon was with the Department of EECS, University of Michigan, Ann Arbor, MI 48109 USA. He is now with IRIT/INP-ENSEEIHT, University of Toulouse, BP 7122, 31071 Toulouse Cedex 7, France (e-mail: Nicolas.Dobigeon@enseeiht.fr).

S. Moussaoui is with the IRCCyN—CNRS UMR 6597, ECN, BP 92101, 44321 Nantes Cedex 3, France (e-mail: said.moussaoui@irccyn.ec-nantes.fr).

M. Coulon and J.-Y. Tourneret are with the IRIT/INP-ENSEEIHT, University of Toulouse, BP 7122, 31071 Toulouse Cedex 7, France (e-mail: Martial.Coulon@enseeiht.fr; Jean-Yves.Tourneret@enseeiht.fr).

A. O. Hero is with the Department of EECS, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: hero@umich.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2009.2025797

[1]The full-additivity constraint, that will be detailed in the following section, refers to a unit ℓ_1-norm.

blind source separation (BSS) problem [20]. In numerous fields, independent component analysis (ICA) [21] has been a mainstay approach to solve BSS problems. In hyperspectral imagery, ICA has also been envisaged [22]. However, as illustrated in [16] and [23], ICA may perform poorly for LSMA due to the strong dependence between the different abundances [24]. Inspired by ICA, dependent component analysis has been introduced in [25] to exploit this dependence. However, this approach assumes that the hyperspectral observations are noise-free. Alternatively, nonnegative matrix factorization (NMF) [26] can also be used to solve BSS problem under nonnegativity constraints. In [27], an NMF algorithm that consists of alternately updating the signature and abundance matrices has been successfully applied to identify constituent in chemical shift imaging. In this work, the additivity constraint has not been taken into account. Basic simulations conducted on synthetic images show that such MNF strategies lead to weak estimation performances. In [28], an iterative algorithm called ICE (iterated constrained endmembers) is proposed to minimize a penalized criterion. As noted in [25], results provided by ICE strongly depend on the choice of the algorithm parameters. More recently, Miao *et al.* have proposed in [29] another iterated optimization scheme performing NMF with an additivity penalty on the abundance coefficients. However, as this is not a hard constraint, it is not necessarily ensured. In addition the performance of the algorithm in [29] degrades significantly when the noise level increases.

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 (see for example [30]) like LSMA. The Bayesian formulation allows one to directly incorporate constraints into the model. These constraints include sparsity [31]; nonnegativity [32]; full additivity (sum-to-one constraint) [33]. 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 nonnegativity and full-additivity of the abundance coefficients (as in [19] and [34]) and nonnegativity of the endmember spectra. To our knowledge, this is the first time that nonnegativity constraints for endmember spectra as well as additivity and nonnegativity constraints for the abundances are jointly considered in a Bayesian model for hyperspectral imagery. In [34], Parra *et al.* propose a Bayesian formulation of the endmember and abundance estimation problem. However, this approach relies on an *ad hoc* autoregressive model of the endmember spectra, which does not necessarily ensure the required positivity constraints. In addition, the MAP estimator proposed in [34] requires an optimization scheme whose convergence is difficult to assess.

Moreover, the proposed joint LSMA 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 on simulated and real data in Sections V and VI. By estimating the parameters on the lower dimensional space we effectively reduce the number of degrees of freedom of the parameters relative to other methods (e.g., [32], [34], and [35]), translating into lower estimator bias and variance. The problem of hyperparameter selection in our Bayesian model is circumvented by adopting the hierarchical Bayesian approach of [19] 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.

The paper is organized as follows. The observation model is described in Section II. The different quantities necessary for the Bayesian formulation are enumerated in Section III. Section IV presents the proposed Gibbs sampler for joint abundance and endmember estimation. Simulation results obtained with synthetic and real AVIRIS data are reported in Sections V and VI respectively. Section VII concludes the paper. An appendix provides details on our parameterization of the simplex and selecting relevant and tractable priors.

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 [4], the L-spectrum $\mathbf{y}_p = [y_{p,1}, \ldots, y_{p,L}]^T$ of the pth pixel ($p = 1, \ldots, P$) is assumed to be a linear combination of R spectra $\mathbf{m}_r$ corrupted by an additive Gaussian noise

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

where $\mathbf{m}_r = [m_{r,1}, \ldots, m_{r,L}]^T$ 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), $\mathbf{n}_p = [n_{p,1}, \ldots, n_{p,L}]^T$ 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.,

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

The proposed model in (2) does not account for any possible correlation in the noise sequences but has been widely adopted in [35]–[37]. However, simulation results reported in paragraph V-D will show that the proposed algorithm is robust to the violation of the i.i.d. noise assumption. Note finally that the model in (1) can be easily modified (see [38]) to handle more complicated noise models with different variances in each spectral band as in [39], or by taking into account correlations between spectral bands as in [19].

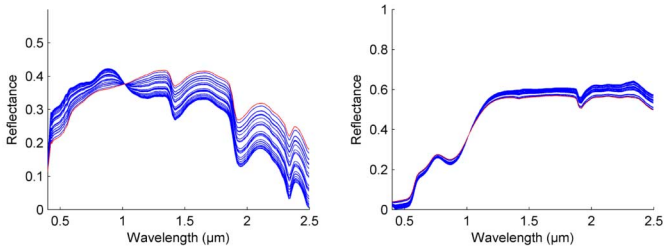


Fig. 1. Range of admissible solution for two endmember spectra: construction concrete (left) and red brick (right). The actual endmember (red lines) are mixed according (1) under the constraints in (3) with random proportions to obtain $P = 2500$ pixels. 50 admissible solutions (blue lines) of the BSS problems in (6) are generated using [42].

Due to physical considerations, described in [3], [19] or [40], the fraction vectors $\mathbf{a}_p = [a_{p,1}, \ldots, a_{p,R}]^T$ in (1) satisfy the following nonnegativity and full-additivity (or *sum-to-one*) constraints

$$\begin{cases} a_{p,r} \geq 0, & \forall r = 1, \ldots, R, \\ \sum_{r=1}^{R} a_{p,r} = 1. \end{cases} \tag{3}$$

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

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

where $\|\cdot\|_1$ is the ℓ_1 norm defined as $\|\mathbf{x}\|_1 = \sum_i |x_i|$, and $\mathbf{a} \succeq \mathbf{0}$ 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 nonnegativity constraints

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

Considering all pixels, standard matrix notation yields

$$\mathbf{Y} = \mathbf{MA} + \mathbf{N} \tag{6}$$

where

$$\begin{aligned} \mathbf{Y} &= [\mathbf{y}_1, \ldots, \mathbf{y}_P], \\ \mathbf{M} &= [\mathbf{m}_1, \ldots, \mathbf{m}_R], \\ \mathbf{A} &= [\mathbf{a}_1, \ldots, \mathbf{a}_P], \\ \mathbf{N} &= [\mathbf{n}_1, \ldots, \mathbf{n}_P]. \end{aligned} \tag{7}$$

In this work, we propose to estimate $\mathbf{A}$ and $\mathbf{M}$ from the noisy observations $\mathbf{Y}$ under the constraints in (3) and (5). Note that the unconstrained BSS problem for estimating M and A from Y is ill-posed: if $\{\mathbf{Y}, \mathbf{A}\}$ is an admissible estimate then $\{\mathbf{YH}, \mathbf{H}^T\mathbf{A}\}$ is also admissible for any unitary matrix H. In the LSMA problem, this nonuniqueness can be partially circumvented by additional constraints such as full-additivity, which enables one to handle the scale indeterminacy. Consequently, these unit ℓ_1-norm constraints on the abundance vectors avoid using more complex strategies for direct estimation of the scale [41]. Despite the constraints in (3) and (5), uniqueness of the couple $\{\mathbf{M}, \mathbf{A}\}$ solution of the LSMA (6) is not systematically ensured. To illustrate this problem, 50 admissible solutions[2] have been depicted in Fig. 1 for $R = 2$ endmembers involved in the mixing of $P = 2500$ pixels [42]. In the following section, the Bayesian model used for the LSMA is presented.

[2]Admissible solutions refer to couples $\{\mathbf{M}, \mathbf{A}\}$ that satisfy (3) and (5) and that follow the model (1) in the noise-free case.

III. Bayesian Model

A. Likelihood

The linear mixing model defined in (1) and the statistical properties in (2) of the noise vector $\mathbf{n}_p$ result in a conditionally Gaussian distribution for the observation of the pth pixel: $\mathbf{y}_p|\mathbf{M}, \mathbf{a}_p, \sigma^2 \sim \mathcal{N}(\mathbf{Ma}_p, \sigma^2\mathbf{I}_L)$. Therefore, the likelihood function of $\mathbf{y}_p$ can be expressed as

$$f(\mathbf{y}_p|\mathbf{M}, \mathbf{a}_p, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{L/2} \exp\left[-\frac{\|\mathbf{y}_p - \mathbf{Ma}_p\|^2}{2\sigma^2}\right] \tag{8}$$

where $\|\mathbf{x}\| = (\mathbf{x}^T\mathbf{x})^{1/2}$ is the ℓ_2 norm. Assuming independence between the noise sequences $\mathbf{n}_p$ $(p = 1, \ldots, P)$, the likelihood function of all the observations $\mathbf{Y}$ is

$$f(\mathbf{Y}|\mathbf{M}, \mathbf{A}, \sigma^2) = \prod_{p=1}^{P} f(\mathbf{y}_p|\mathbf{M}, \mathbf{a}_p, \sigma^2). \tag{9}$$

B. Prior Model for the Endmember Spectra

1) Dimensionality Reduction: It is interesting to note that the unobserved matrix $\mathbf{X} = \mathbf{MA} = \mathbf{Y} - \mathbf{N}$ is rank deficient under the linear model (1). More precisely, the set

$$\mathcal{S}_{\mathbf{M}} = \left\{\mathbf{x} \in \mathbb{R}^L;\ \mathbf{x} = \sum_{r=1}^{R} \lambda_r \mathbf{m}_r,\ \sum_{r=1}^{R} \lambda_r = 1,\ \lambda_r \geq 0\right\} \tag{10}$$

is a $(R-1)$-dimensional convex polytope of $\mathbb{R}^L$ whose vertices are the R endmember spectra $\mathbf{m}_r$ $(r = 1, \ldots, R)$ to be recovered. Consequently, in the noise-free case, $\mathbf{X}$ 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. To illustrate this property, $P = 1000$ pixels resulting from a noise-free mixture of $R = 3$ endmembers are represented in Fig. 2. As noted in [4], dimensionality reduction is a common step of the LSMA, adopted by numerous EEAs, such as N-FINDR [9] or PPI [8]. Similarly, we propose to estimate the projection $\mathbf{t}_r$ $(r = 1, \ldots, R)$ of the endmember spectra $\mathbf{m}_r$ 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 $\mathbf{v}_1, \ldots, \mathbf{v}_K$ identified by a principal component analysis (PCA) on the observations $\mathbf{Y}$ [43]

$$\mathcal{V}_K = \text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_K). \tag{11}$$

The first two principal axes are identified in Fig. 2 for the synthetic hyperspectral data. In the following paragraph, PCA is described. While we do not give details here, this PCA-based

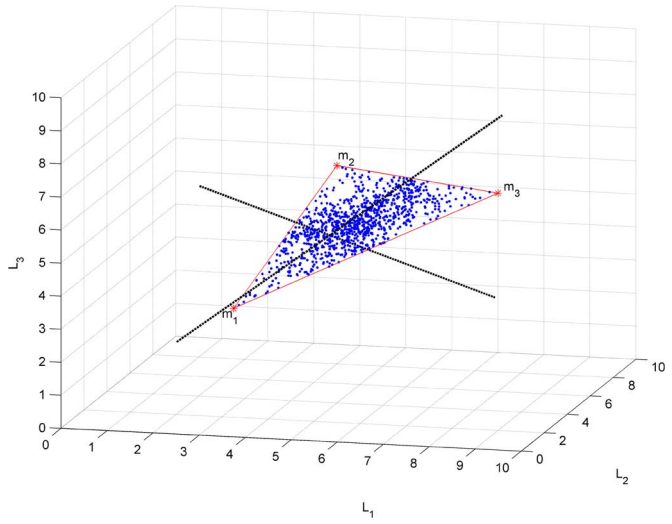


Fig. 2. Example of hyperspectral data observed in 3 spectral bands. The mixed pixels (blue points) belong to the R-dimensional convex polytope $\mathcal{S}_{\mathbf{M}}$ (red lines) whose vertices are the endmembers spectra $\mathbf{m}_1, \ldots, \mathbf{m}_R$ (red stars). The first two principal axes estimated by a PCA appear in dashed lines and define the projection subset $\mathcal{V}_K$.

dimension reduction step can be easily replaced by other projection techniques, such as the maximum noise fraction (MNF) transform [44] that has been considered in paragraph V-D.

2) PCA Projection: The $L \times L$ empirical covariance matrix $\mathbf{\Upsilon}$ of the data $\mathbf{Y}$ is given by

$$\mathbf{\Upsilon} = \frac{1}{P}\sum_{p=1}^{P}(\mathbf{y}_p - \bar{\mathbf{y}})(\mathbf{y}_p - \bar{\mathbf{y}})^T \tag{12}$$

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

$$\bar{\mathbf{y}} = \frac{1}{P}\sum_{p=1}^{P}\mathbf{y}_p. \tag{13}$$

Let

$$\begin{cases} \mathbf{D} = \operatorname{diag}(\lambda_1, \ldots, \lambda_K), \\ \mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_K]^T \end{cases} \tag{14}$$

denote, respectively, the diagonal matrix of the K highest eigenvalues and the corresponding eigenvector matrix of $\mathbf{\Upsilon}$. The PCA projection $\mathbf{t}_r \in \mathbb{R}^K$ of the endmember spectrum $\mathbf{m}_r \in \mathbb{R}^L$ is obtained as follows:

$$\mathbf{t}_r = \mathbf{P}(\mathbf{m}_r - \bar{\mathbf{y}}) \tag{15}$$

with $\mathbf{P} = \mathbf{D}^{-1/2}\mathbf{V}$. Equivalently

$$\mathbf{m}_r = \mathbf{U}\mathbf{t}_r + \bar{\mathbf{y}} \tag{16}$$

with $\mathbf{U} = \mathbf{V}^T\mathbf{D}^{1/2}$. Note that in the subspace $\mathcal{V}_{R-1}$ obtained for $K = R - 1$, the vectors $\{\mathbf{t}_r\}_{r=1,\ldots,R}$ form a simplex that standard EEAs such as N-FINDR [9], MVT [11] and ICE [28] try to recover. In this paper, we estimate the vertices $\mathbf{t}_r$ $(r = 1, \ldots, R)$ of this simplex using a Bayesian approach. The Bayesian prior distributions for the projections $\mathbf{t}_r$ $(r = 1, \ldots, R)$ are introduced in the following paragraph.

3) Prior Distribution for the Projected Spectra: All the elements of the subspace $\mathcal{V}_K$ may not be appropriate projected spectra according to (15). Indeed, the $K \times 1$ vector $\mathbf{t}_r$ has to ensure nonnegativity constraints (5) of the corresponding reconstructed $L \times 1$ spectra $\mathbf{m}_r$. For each endmember $\mathbf{m}_r$, straightforward computations establish that for any $r = 1, \ldots, R$

$$\{m_{l,r} \geq 0, \quad \forall\, l = 1, \ldots, L\} \Leftrightarrow \{\mathbf{t}_r \in \mathcal{T}_r\} \tag{17}$$

where the set $\mathcal{T}_r \subset \mathcal{V}_K$ is defined by the following L inequalities:

$$\mathcal{T}_r = \left\{ \mathbf{t}_r; \bar{y}_l + \sum_{k=1}^{K} u_{l,k}t_{k,r} \geq 0, l = 1, \ldots, L \right\} \tag{18}$$

with $\bar{\mathbf{y}} = [\bar{y}_1, \ldots, \bar{y}_L]^T$ and $\mathbf{U} = [u_{l,k}]$. A conjugate[3] multivariate Gaussian distribution (MGD) $\mathcal{N}_{\mathcal{T}_r}\left(\mathbf{e}_r, s_r^2\mathbf{I}_K\right)$ truncated on the set $\mathcal{T}_r$ is chosen as prior distribution for $\mathbf{t}_r$. The probability density function (pdf) $\phi_{\mathcal{T}_r}(\cdot)$ of this truncated MGD is defined by

$$\phi_{\mathcal{T}_r}\left(\mathbf{t}_r \middle| \mathbf{e}_r, s_r^2\mathbf{I}_K\right) \propto \phi\left(\mathbf{t}_r \middle| \mathbf{e}_r, s_r^2\mathbf{I}_K\right)\mathbf{1}_{\mathcal{T}_r}(\mathbf{t}_r) \tag{19}$$

where $\propto$ stands for "proportional to", $\phi(\cdot|\mathbf{u}, \mathbf{W})$ is the pdf of the standard MGD $\mathcal{N}(\mathbf{u}, \mathbf{W})$ with mean vector $\mathbf{u}$ and covariance matrix $\mathbf{W}$, and $\mathbf{1}_{\mathcal{T}_r}(\cdot)$ is the indicator function on the set $\mathcal{T}_r$

$$\mathbf{1}_{\mathcal{T}_r}(\mathbf{x}) = \begin{cases} 1, & \text{if } \mathbf{x} \in \mathcal{T}_r; \\ 0, & \text{overwise.} \end{cases} \tag{20}$$

The normalizing constant $K_{\mathcal{T}_r}\left(\mathbf{e}_r, s_r^2\mathbf{I}_K\right)$ in (19) is defined as follows:

$$K_{\mathcal{T}_r}\left(\mathbf{e}_r, s_r^2\mathbf{I}_K\right) = \int_{\mathcal{T}_r} \phi\left(\mathbf{x}|\mathbf{e}_r, s_r^2\mathbf{I}_K\right) d\mathbf{x}. \tag{21}$$

This paper proposes to select the mean vectors $\mathbf{e}_r$ $(r = 1, \ldots, R)$ in (19) as the projected spectra of pure components previously identified by 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 ($s_1^2 = \cdots = s_R^2 = 50$ in our simulations).

By assuming *a priori* independence of the vectors $\mathbf{t}_r$ $(r = 1, \ldots, R)$, the prior distribution for the projected endmember matrix $\mathbf{T} = [\mathbf{t}_1, \ldots, \mathbf{t}_R]$ is

$$f(\mathbf{T} \mid \mathbf{E}, \mathbf{s}^2) = \prod_{r=1}^{R} \phi_{\mathcal{T}_r}\left(\mathbf{t}_r \middle| \mathbf{e}_r, s_r^2\mathbf{I}_K\right) \tag{22}$$

where $\mathbf{E} = [\mathbf{e}_1, \ldots, \mathbf{e}_R]$ and $\mathbf{s}^2 = \left[s_1^2, \ldots, s_R^2\right]$.

[3]For the main motivations of choosing conjugate priors, see for instance [45, Ch. 3].

C. Abundance Prior

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

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

and $a_{p,R} = 1 - \sum_{r=1}^{R-1} a_{p,r}$. Following the model in [19], the priors chosen for $\mathbf{c}_p\,(p=1,\ldots,P)$ are uniform distributions on the simplex $\mathcal{S}$ defined by

$$\mathcal{S} = \left\{\mathbf{c}_p;\ \|\mathbf{c}_p\|_1 \leq 1 \text{ and } \mathbf{c}_p \succeq \mathbf{0}\right\}. \tag{23}$$

Choosing this prior distribution for $\mathbf{c}_p\,(p=1,\ldots,P)$ is equivalent to electing a Dirichlet distribution $\mathcal{D}\,(1,\ldots,1)$, i.e., a uniform distribution on $\mathcal{A}$ defined in (4), as prior distribution for the full abundance vector $\mathbf{a}_p$ [45, Appendix A]. However, the proposed reparametrization will prove to be well adapted to the Gibbs sampling strategy introduced in Section IV. Under the assumption of statistical independence between the abundance vectors $\mathbf{c}_p\,(p=1,\ldots,P)$, the full prior distribution for partial abundance matrix $\mathbf{C} = [\mathbf{c}_1,\ldots,\mathbf{c}_P]^T$ can be written

$$f\,(\mathbf{C}) \propto \prod_{p=1}^{P} \mathbf{1}_{\mathcal{S}}\,(\mathbf{c}_p). \tag{24}$$

As noted in [19], the uniform prior distribution reflects a lack of *a priori* knowledge about the abundance vector. As mentioned in [25], it might be interesting to consider nonuniform priors for the abundances when considering images in which there are not spectral vectors in some or in all the simplex facets. For instance, following the approach in [25], Dirichlet distributions $\mathcal{D}\,(\delta_1,\ldots,\delta_R)$ could be proposed as priors for the abundance vectors. However, as direct estimation of the hyperparameters $\delta_1,\ldots,\delta_R$ remains difficult, this modification might significantly increase the complexity of the Bayesian model and the computational cost of the algorithm. Moreover, for the BSS problem addressed in this work, the uniform prior in (24) imposes a strong constraint on the size of the simplex to be recovered. As demonstrated in the Appendix, 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. This property has also been exploited in [11].

D. Noise Variance Prior

A conjugate prior is chosen for σ^2

$$\sigma^2\,|\nu,\gamma \sim \mathcal{IG}\left(\frac{\nu}{2},\frac{\gamma}{2}\right) \tag{25}$$

where $\mathcal{IG}\,(\nu/2,\gamma/2)$ denotes the inverse-gamma distribution with parameters $\nu/2$ and $\gamma/2$. As in previous works [46], [47], the hyperparameter ν will be fixed to $\nu = 2$. On the other hand, 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 noninformative Jeffreys' prior [48] which reflects the lack of knowledge regarding this hyperparameter

$$f\,(\gamma) \propto \frac{1}{\gamma}\mathbf{1}_{\mathbb{R}^+}\,(\gamma). \tag{26}$$

F. Posterior Distribution

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

$$f(\boldsymbol{\theta}|\mathbf{Y}) = \int f(\boldsymbol{\theta},\gamma|\mathbf{Y})d\gamma \propto \int f(\mathbf{Y}|\boldsymbol{\theta})f(\boldsymbol{\theta}|\gamma)f(\gamma)d\gamma \tag{27}$$

where $f\left(\mathbf{Y}\middle|\boldsymbol{\theta}\right)$ and $f\,(\gamma)$ are defined in (9) and (26), respectively. Moreover, under the assumption of *a priori* independence between $\mathbf{C}$, $\mathbf{T}$ and σ^2, the following result can be obtained:

$$f\left(\boldsymbol{\theta}\middle|\gamma\right) = f\,(\mathbf{C})\,f\left(\mathbf{T}\mid\mathbf{E},\mathbf{s}^2\right)\,f\left(\sigma^2\mid\nu,\gamma\right) \tag{28}$$

where $f\,(\mathbf{C})$, $f\left(\mathbf{T}\mid\mathbf{E},\mathbf{s}^2\right)$ and $f\left(\sigma^2\mid\nu,\gamma\right)$ have been defined in (24), (22) and (25), respectively. This hierarchical structure allows one to integrate out the hyperparameter γ from the joint distribution $f\,(\boldsymbol{\theta},\gamma|\mathbf{Y})$, yielding

$$\begin{aligned} &f\left(\mathbf{C},\mathbf{T},\sigma^2\middle|\mathbf{Y}\right) \\ &\propto \prod_{p=1}^{P}\mathbf{1}_{\mathcal{S}}\,(\mathbf{c}_p) \\ &\quad\times\prod_{r=1}^{R}\exp\left[-\frac{\|\mathbf{t}_r-\mathbf{e}_r\|^2}{2s_r^2}\right]\mathbf{1}_{\mathcal{T}_r}\,(\mathbf{t}_r) \\ &\quad\times\prod_{p=1}^{P}\left[\left(\frac{1}{\sigma^2}\right)^{L/2+1}\exp\left(-\frac{\left\|\mathbf{y}_p-\left(\mathbf{U}\mathbf{T}+\bar{\mathbf{y}}\mathbf{1}_R^T\right)\mathbf{a}_p\right\|^2}{2\sigma^2}\right)\right]. \end{aligned} \tag{29}$$

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 (29) 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 (29).

IV. Gibbs Sampler

Random samples (denoted by $\cdot^{(t)}$ where t is the iteration index) can be drawn from $f\left(\mathbf{C},\mathbf{T},\sigma^2\mid\mathbf{Y}\right)$ using a Gibbs sampler [49]. This MCMC technique consists of generating samples $\left\{\mathbf{C}^{(t)},\mathbf{T}^{(t)},\boldsymbol{\sigma}^{2(t)}\right\}$ distributed according to the conditional posterior distributions of each parameter.

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

Straightforward computations yield for each observation

$$f\left(\mathbf{c}_p\left|\mathbf{T},\sigma^2,\mathbf{y}_p\right.\right) \propto \exp\left[-\frac{\left(\mathbf{c}_p-\boldsymbol{\upsilon}_p\right)^T\boldsymbol{\Sigma}_p^{-1}\left(\mathbf{c}_p-\boldsymbol{\upsilon}_p\right)}{2}\right]\mathbf{1}_{\mathcal{S}}\left(\mathbf{c}_p\right) \quad (30)$$

where

$$\begin{cases}\boldsymbol{\Sigma}_p=\left[\left(\mathbf{M}_{-R}-\mathbf{m}_R\mathbf{1}_{R-1}^T\right)^T\boldsymbol{\Sigma}_{\mathrm{n}}^{-1}\left(\mathbf{M}_{-R}-\mathbf{m}_R\mathbf{1}_{R-1}^T\right)\right]^{-1}\\ \boldsymbol{\upsilon}_p=\boldsymbol{\Sigma}_p\left[\left(\mathbf{M}_{-R}-\mathbf{m}_R\mathbf{1}_{R-1}^T\right)^T\boldsymbol{\Sigma}_{\mathrm{n}}^{-1}\left(\mathbf{y}_p-\mathbf{m}_R\right)\right]\end{cases} \quad (31)$$

with $\boldsymbol{\Sigma}_{\mathrm{n}}^{-1}=1/\sigma^2\mathbf{I}_L$ and where $\mathbf{M}_{-R}$ denotes the matrix $\mathbf{M}$ whose Rth column has been removed. As a consequence, $\mathbf{c}_p\left|\mathbf{T}\right.$, $\sigma^2,\mathbf{y}_p$ is distributed according to an MGD truncated on the simplex $\mathcal{S}$ in (23)

$$\mathbf{c}_p\left|\mathbf{T},\sigma^2,\mathbf{y}_p\right.\sim\mathcal{N}_{\mathcal{S}}\left(\boldsymbol{\upsilon}_p,\boldsymbol{\Sigma}_p\right). \quad (32)$$

Note that samples can be drawn from an MGD truncated on a simplex using efficient Monte Carlo simulation strategies such as described in [50].

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

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

$$f\left(\mathbf{t}_r|\mathbf{T}_{-r},\mathbf{c}_r,\sigma^2,\mathbf{Y}\right) \propto \exp\left[-\frac{1}{2}\left(\mathbf{t}_r-\boldsymbol{\tau}_r\right)^T\boldsymbol{\Lambda}_r^{-1}\left(\mathbf{t}_r-\boldsymbol{\tau}_r\right)\right]\mathbf{1}_{\mathcal{T}_r}\left(\mathbf{t}_r\right) \quad (33)$$

with

$$\begin{cases}\boldsymbol{\Lambda}_r=\left[\sum_{p=1}^P a_{p,r}^2\mathbf{U}^T\boldsymbol{\Sigma}_{\mathrm{n}}^{-1}\mathbf{U}+\frac{1}{s_r^2}\mathbf{I}_K\right]^{-1},\\ \boldsymbol{\tau}_r=\boldsymbol{\Lambda}_r\left[\sum_{p=1}^P a_{p,r}\mathbf{U}^T\boldsymbol{\Sigma}_{\mathrm{n}}^{-1}\boldsymbol{\epsilon}_{p,r}+\frac{1}{s_r^2}\mathbf{e}_r\right]\end{cases} \quad (34)$$

and

$$\boldsymbol{\epsilon}_{p,r}=\mathbf{y}_p-a_{p,r}\bar{\mathbf{y}}-\sum_{j\neq r}a_{p,j}\mathbf{m}_j. \quad (35)$$

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

$$\mathbf{t}_r\mid\mathbf{T}_{-r},\mathbf{c}_r,\sigma^2,\mathbf{Y}\sim\mathcal{N}_{\mathcal{T}_r}\left(\boldsymbol{\tau}_r,\boldsymbol{\Lambda}_r\right). \quad (36)$$

Generating vectors distributed according to this distribution is a difficult task, mainly due to the truncation on the subset $\mathcal{T}_r$. An alternative consists of generating each component $t_{k,r}$ of $\mathbf{t}_r$ conditionally upon the others $\mathbf{t}_{-k,r}=\{t_{j,r}\}_{j\neq k}$. More precisely, by denoting $\mathcal{U}_k^+=\{l;u_{l,k}>0\}$, $\mathcal{U}_k^-=\{l;u_{l,k}<0\}$ and $\varepsilon_{l,k,r}=\bar{y}_l+\sum_{j\neq k}u_{l,j}t_{j,r}$, one can write

$$t_{k,r}|\mathbf{t}_{-k,r},\mathbf{T}_{-r},\mathbf{c}_r,\sigma^2,\mathbf{Y}\sim\mathcal{N}_{\left[t_{k,r}^-,t_{k,r}^+\right]}\left(w_{k,r},z_{k,r}^2\right) \quad (37)$$

with

$$\begin{cases}t_{k,r}^-=\max_{l\in\mathcal{U}_k^+}-\frac{\varepsilon_{l,k,r}}{u_{l,k}}\\ t_{k,r}^+=\min_{l\in\mathcal{U}_k^-}-\frac{\varepsilon_{l,k,r}}{u_{l,k}}\end{cases} \quad (38)$$

and where $w_{k,r}$ and $z_{k,r}^2$ are the conditional mean and variance, respectively, derived from the partitioned mean vector and covariance matrix [51, p. 324] (see [50] for similar computations). Generating samples distributed according to the two-sided truncated Gaussian distribution in (37) can be easily achieved with the algorithm described in [52].

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

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

$$\sigma^2|\mathbf{C},\mathbf{T},\mathbf{Y}\sim\mathcal{IG}\left(\frac{PL}{2},\frac{1}{2}\sum_{p=1}^P\|\mathbf{y}_p-\mathbf{M}\mathbf{a}_p\|^2\right). \quad (39)$$

Simulating according to this inverse Gamma distribution can be achieved using a Gamma variate generator (see [53, Ch. 9] and [45, Appendix A]).

To summarize, the hyperparameters that have to be fixed at the beginning of the algorithm are chosen as follows: $\nu=2$, $s_1^2=\cdots=s_R^2=50$ and $\{\mathbf{e}_r\}_{r=1,\ldots,R}$ are set to projected spectra identified by a standard EEA (e.g., N-FINDR).

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 [54] and are represented in Fig. 3 (top, black lines). The reflectances are observed in $L=413$ spectral bands ranging from 0.4 to 2.5 μm. 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. 4 (top) in gray scale where a white (respectively, black) pixel stands for the presence (resp. absence) of the material. The signal-to-noise ratio has been tuned to $\mathrm{SNR}_{\mathrm{dB}}=15$ dB.

A. Endmember Spectrum Estimation

The resulting hyperspectral data have been unmixed by the proposed algorithm. First, the space $\mathcal{V}_K$ in (11) has been identified by PCA as discussed in Section III-B-2. The hidden mean vectors $\mathbf{e}_r$ $(r=1,\ldots,R)$ of the normal distributions in (19) 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 $s_1^2=\cdots=s_R^2=50$ to obtain vague priors (i.e., large variances). The Gibbs sampler has been run with $N_{\mathrm{MC}}=1300$ iterations, including $N_{\mathrm{bi}}=300$ burn-in iterations. The MMSE estimates of the abundance vectors $\mathbf{a}_p$ $(p=1,\ldots,P)$ and the projected spectra $\mathbf{t}_r$ $(r=1,\ldots,R)$ have been approximated by computing empirical averages over the last computed outputs of the sampler $\left\{\mathbf{a}_r^{(t)}\right\}_{t=1,\ldots,N_{\mathrm{MC}}}$ and $\left\{\mathbf{t}_r^{(t)}\right\}_{t=1,\ldots,N_{\mathrm{MC}}}$, following the MMSE principle

$$\hat{\mathbf{x}}_{\mathrm{MMSE}}=\mathrm{E}\left[\mathbf{x}|\mathbf{y}\right]\approx\frac{1}{N_{\mathrm{MC}}-N_{\mathrm{bi}}}\sum_{t=N_{\mathrm{bi}}+1}^{N_{\mathrm{MC}}}\mathbf{x}^{(t)}. \quad (40)$$

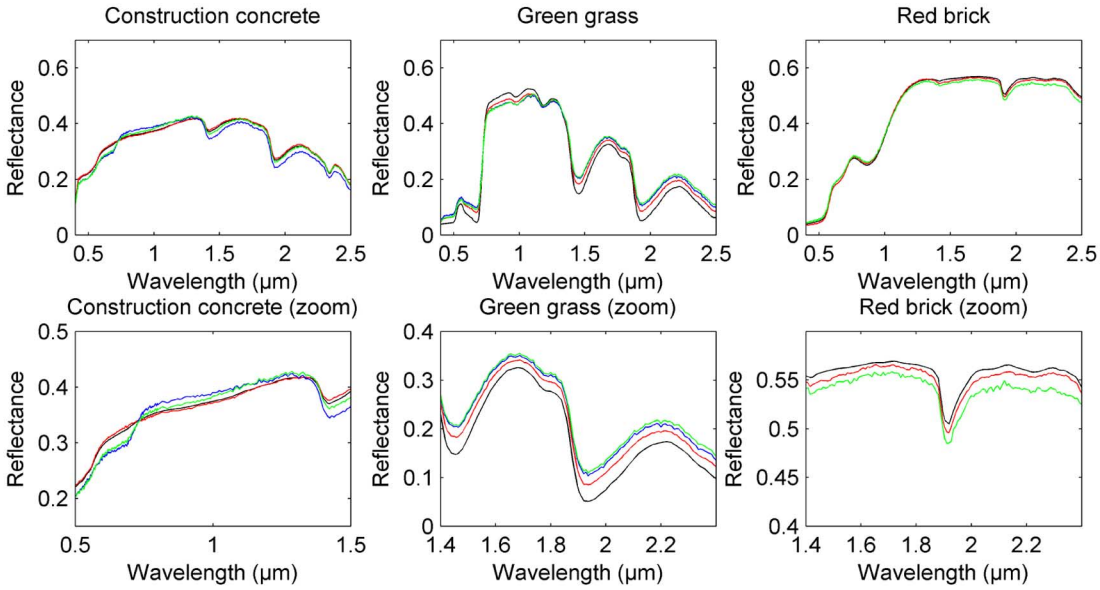


Fig. 3. Actual endmembers (black lines), endmembers estimated by N-FINDR (blue lines), endmembers estimated by VCA (green lines) and endmembers estimated by proposed approach (red lines).

TABLE I
ABUNDANCE MEANS AND VARIANCES OF EACH ENDMEMBER IN EACH REGION OF THE 100×100 HYPERSPECTRAL IMAGE

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

The corresponding endmember spectra estimated by the proposed algorithm are depicted in Fig. 3 (top, red lines). The proposed algorithm clearly outperforms N-FINDR and VCA, as shown in Fig. 3. The scatter plot in Fig. 5 provides additional insight. The N-FINDR and VCA algorithms assume the presence of pure pixels in the data. However, as none of these pixels are pure, N-FINDR and VCA provide poorer results than the proposed joint Bayesian algorithm. To illustrate this point, the performances of the different algorithms have been compared via two criteria. First, the mean square errors (MSEs)

$$\mathrm{MSE}_r^2 = \|\hat{\mathbf{m}}_r - \mathbf{m}_r\|^2, \quad r = 1, \ldots, R \tag{41}$$

are good quality indicators for the estimates. In addition, another metric frequently encountered in hyperspectral imagery literature, known as the spectral angle distance (SAD), has been considered. The SAD measures the angle between the actual and the corresponding estimated spectrum

$$\mathrm{SAD}_r = \arccos\left(\frac{\langle \hat{\mathbf{m}}_r, \mathbf{m}_r \rangle}{\|\hat{\mathbf{m}}_r\| \, \|\mathbf{m}_r\|}\right) \tag{42}$$

where $\langle \cdot, \cdot \rangle$ stands for the scalar product. These performance criteria computed for the endmember spectra estimated by the different algorithm are reported in Table II. They show that the proposed method performs significantly better than the others. The computation times required by each of these algorithms are reported in Table III for an unoptimized MATLAB 2007b 32-bit implementation on a 2.2-GHz Intel Core 2. Obviously, the complexity of the VCA and N-FINDR methods are lower than the proposed approach. Note however that, unlike to the joint Bayesian procedure, these standard EEA must be coupled with an abundance estimation algorithm. Moreover they only provide point estimates of the endmember spectra. Note finally that the computational complexity of N-FINDR, because it is combinatorial, increases drastically with the number of pixels and endmembers.

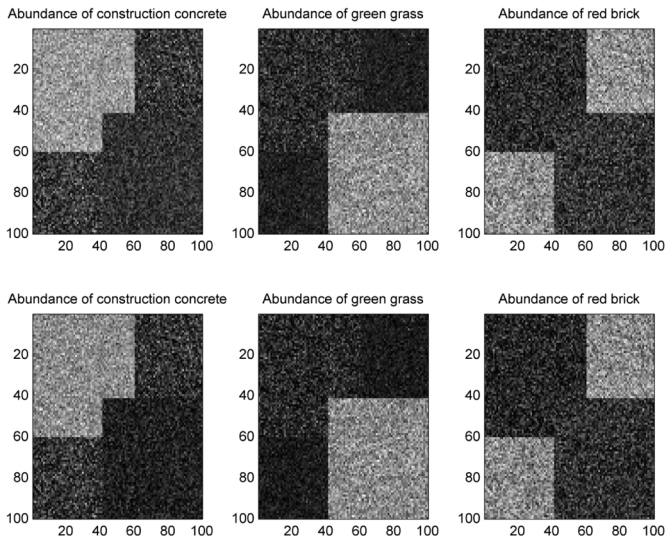


Fig. 4. Top: actual endmember abundance maps. Bottom: estimated endmember abundance maps.

TABLE II
PERFORMANCE COMPARISON BETWEEN VCA, N-FINDR AND THE PROPOSED BAYESIAN METHOD: MSE^2 AND SAD ($\times 10^{-1}$) BETWEEN THE ACTUAL AND THE ESTIMATED ENDMEMBER SPECTRA

		End.	Bayesian		VCA		N-FINDR		cNMF		MVC-NMF	
			MSE^2	SAD	MSE^2	SAD	MSE^2	SAD	MSE^2	SAD	MSE^2	SAD
SNR = 5dB	$R=3$	#1	1.70	0.63	10.29	1.36	2.69	0.75	16.78	1.59	16.61	1.63
		#2	6.56	1.49	7.37	1.51	10.87	1.80	5.15	1.29	30.97	1.10
		#3	2.70	0.59	2.94	0.60	2.94	0.60	4.90	0.78	5.80	0.73
	$R=5$	#1	0.70	0.01	9.72	1.48	6.32	1.21	14.82	1.86	N/A	N/A
		#2	1.05	0.49	12.46	1.73	7.44	1.05	10.48	1.79	N/A	N/A
		#3	1.04	0.49	7.26	1.17	1.05	5.61	25.51	2.00	N/A	N/A
		#4	1.05	0.49	6.76	0.90	6.76	0.90	25.10	1.71	N/A	N/A
		#5	1.04	0.49	2.32	0.60	10.26	0.70	10.67	1.13	N/A	N/A
SNR = 15dB	$R=3$	#1	0.10	0.15	1.29	0.48	0.54	0.33	16.77	1.58	8.03	1.27
		#2	2.68	0.92	5.18	1.27	5.19	1.26	5.15	1.28	2.43	0.49
		#3	0.16	0.12	0.57	0.22	0.57	0.22	4.90	0.78	26.01	0.94
	$R=5$	#1	0.15	0.18	2.61	0.78	4.63	1.04	1.83	0.64	73.07	4.27
		#2	0.55	0.40	7.62	1.36	7.62	1.36	5.01	0.93	140.12	3.77
		#3	0.31	0.22	2.23	0.54	2.32	0.54	6.65	1.02	27.72	2.03
		#4	0.39	0.15	5.81	0.74	5.81	0.74	10.56	0.73	129.61	3.58
		#5	0.62	0.25	0.80	0.30	6.52	0.77	2.76	0.58	50.47	2.73
SNR = 25dB	$R=3$	#1	0.05	0.09	1.14	0.52	1.14	0.52	1.75	0.52	17.94	1.98
		#2	2.19	0.83	5.65	1.33	5.65	1.33	2.88	0.95	7.98	1.38
		#3	0.17	0.14	0.66	0.22	0.66	0.22	0.82	0.34	30.91	1.04
	$R=5$	#1	0.42	0.29	0.70	0.40	0.70	0.40	1.74	0.59	98.49	4.10
		#2	0.37	0.34	11.34	1.44	11.11	1.80	3.86	0.77	11.34	1.44
		#3	0.46	0.29	1.44	0.48	1.44	0.47	16.25	1.71	10.93	1.36
		#4	0.07	0.09	2.36	0.44	5.68	0.52	6.86	0.62	74.02	2.84
		#5	0.35	0.20	1.54	0.42	2.92	0.59	1.34	0.46	97.27	3.73

TABLE III
COMPUTATIONAL TIMES OF VCA, N-FINDR AND THE PROPOSED BAYESIAN METHOD FOR UNMIXING $P = 32 \times 32$ PIXELS WITH $R = 3$ ENDMEMBERS

	Bayesian	VCA	N-FINDR
Times (s)	1511	1	23

B. Abundance Estimation

The MMSE estimates of the abundance vectors for the $P = 10^4$ pixels of the image have been computed following the MMSE principle in (40)

$$\hat{\mathbf{a}}_p = \frac{1}{N_{\mathrm{MC}} - N_{\mathrm{bi}}} \sum_{t=N_{\mathrm{bi}}+1}^{N_{\mathrm{MC}}} \mathbf{a}_p^{(t)}. \tag{43}$$

The corresponding estimated abundance maps are depicted in Fig. 4 (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, these posteriors allow one to derive confidence intervals regarding the parameters of interest. For instance, the posterior distributions of the abundance coefficients are depicted in Fig. 6 for pixel number 100. Note that these estimated posteriors are in good agreement with the actual values of $\mathbf{a}_{100}$ depicted in red dotted lines.

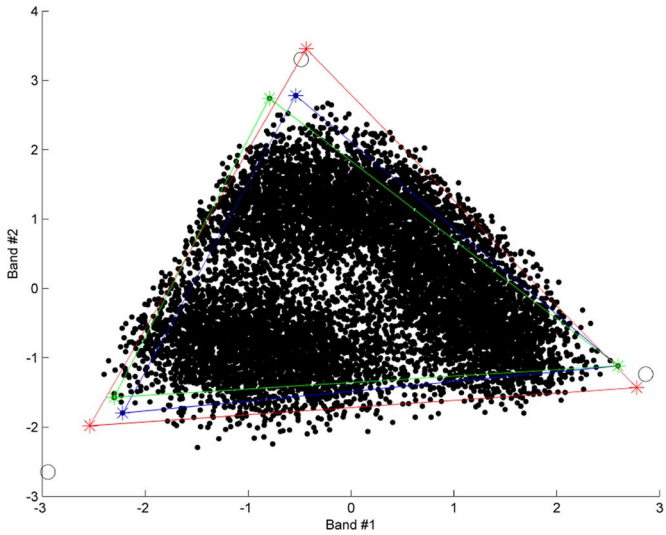


Fig. 5. Scatter plot in the lower-dimensional space $\mathcal{V}_2$: projected dataset (black points), actual endmembers (black circles), endmembers estimated by N-FINDR (blue stars), endmembers estimated by VCA (green stars) and endmembers estimated by proposed approach (red stars). All pixel spectra do not lie inside ground truth simplex due to simulated measurement noise.

These results have been compared with estimates provided by the N-FINDR or VCA algorithms, coupled with an abundance estimation procedure based on the fully constrained least-

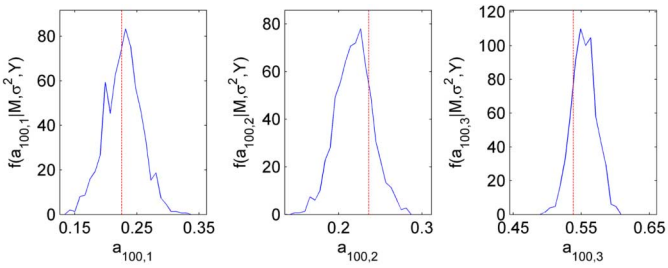


Fig. 6. Posterior distributions of $a_{p,r}$ $(r = 1, \ldots, 3)$. The actual values are depicted in red dotted lines.

TABLE IV
PERFORMANCE COMPARISON BETWEEN VCA, N-FINDR AND THE PROPOSED BAYESIAN METHOD: GMSE^2 BETWEEN THE ACTUAL AND THE ESTIMATED ABUNDANCE MAPS

	Bayesian	VCA	N-FINDR
Endm. #1	25.68	57.43	30.66
Endm. #2	29.97	74.48	46.45
Endm. #3	3.19	83.02	11.22

TABLE V
PERFORMANCE COMPARISON BETWEEN VCA, N-FINDR AND THE PROPOSED BAYESIAN METHOD IN PRESENCE OF GAUSSIAN SHAPED NOISE: MSE^2 AND SAD $(\times 10^{-1})$ BETWEEN THE ACTUAL AND THE ESTIMATED ENDMEMBER SPECTRA

Endm.	MNF/Bayes		HySime/Bayes		VCA		N-FINDR	
	MSE^2	SAD	MSE^2	SAD	MSE^2	SAD	MSE^2	SAD
#1	0.26	0.25	0.42	0.31	1.11	0.46	1.11	0.46
#2	1.99	0.79	4.35	1.16	5.78	1.33	5.78	1.33
#3	0.33	0.19	0.57	0.22	1.94	0.41	2.19	0.43

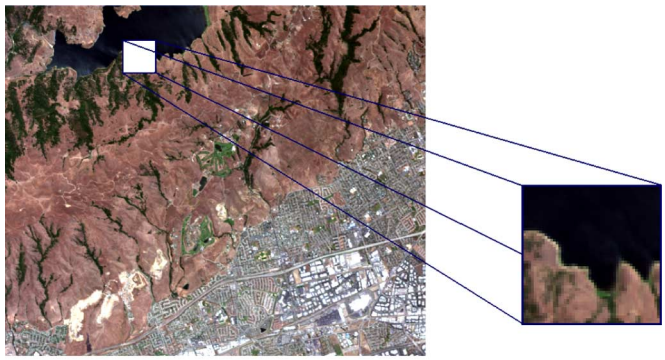
squares (FCLS) approach proposed by Heinz *et al.* [17]. The global abundance MSEs have been computed as

$$\mathrm{GMSE}_r^2 = \sum_{p=1}^{P} (\hat{a}_{p,r} - a_{p,r})^2 \tag{44}$$

where $\hat{a}_{p,r}$ is the estimated abundance coefficient of the material $\#r$ in the pixel $\#p$. These performance measures have been reported in Table IV and confirm the accuracy of the proposed Bayesian estimation method. Moreover, note that neither N-FINDR nor VCA are able to provide confidence measures such as those depicted in Fig. 6.

C. Other Simulation Scenarios

Simulations with different noise levels ($\mathrm{SNR}_{\mathrm{dB}} = 5$ dB, 15 dB, 25 dB) and with other combinations of endmembers ($R = 3$ and $R = 5$) are reported in Table II. First, VCA and N-FINDR algorithms have been applied on these different datasets, as well as the proposed approach. In addition, to complete the performance comparison, a constrained version of a nonnegative matrix factorization algorithm [27] (referred to as cNMF) has been implemented and evaluated. Finally, the minimum volume transform (MVT) based strategy described in [29] (referred to as MVC-NMF) has also been applied on the synthetic hyperspectral images. Note that this MVC-NMF algorithm has failed to unmix the pixels when using $R = 5$ endmembers for $\mathrm{SNR} = 5$ dB. Consequently, as the results were not significant, they have not been reported in Table II. Estimation performances expressed in terms of MSE and SAD show that MVC-NMF and cNMF provided the worst results for these examples. These results also corroborate the effectiveness of our Bayesian estimation procedure, when compared with others EEA approaches.



(a) (b)

Fig. 7. Real hyperspectral data: Moffett Field image acquired by AVIRIS in 1997 (left) and the region of interest (right) represented in synthetic colors.

D. Robustness to Non-i.i.d. Noise Models

In this paragraph, we illustrate the robustness of the proposed algorithm with respect to violation of the i.i.d. noise assumption. More precisely, a so-called Gaussian shaped noise inspired by [55] has been considered. The noise correlation matrix $\boldsymbol{\Sigma}_{\mathrm{n}} = \mathrm{diag}\left(\sigma_1^2, \ldots, \sigma_L^2\right)$ is designed such that its diagonal elements σ_l^2 $(l = 1, \ldots, L)$ follow a Gaussian shape centered at band $L/2$

$$\sigma_l^2 = \sigma^2 \exp\left[-\frac{\left(\frac{l-L}{2}\right)^2}{2\eta^2}\right]. \tag{45}$$

The parameter σ^2 can be tuned to choose the SNR whereas the parameter η adjusts the shape width ($\eta \to \infty$ corresponds to i.i.d. noise). For this simulation, the parameters σ^2 and η have been fixed to 1.0×10^2 and 50 respectively, leading to a noise level of $\mathrm{SNR}_{\mathrm{dB}} = 15$ dB. When the noise is not i.i.d., dimensionality reduction methods based on eigen-decomposition of observed data correlation matrix $\boldsymbol{\Upsilon}$ introduced in (12) can be inefficient. In this case, other hyperspectral subspace identification methods have to be considered. Here, the PCA-based dimension reduction step introduced in paragraph III-B-2 was replaced by two techniques: the well-known MNF transform [44] approach and the more recent HySime algorithm [55]. Both of them require the estimation of the noise covariance matrix $\boldsymbol{\Sigma}_{\mathrm{n}}$, which was implemented following [55]. The estimation performances for the proposed Bayesian estimation procedure coupled with MNF or HySime are reported in Table V and compared with VCA and N-FINDR. These results show that the proposed method i) can be easily used with other dimension reduction procedures, and ii) is quite robust to violation of i.i.d. noise assumption, at least for independent but non identically distributed noise.

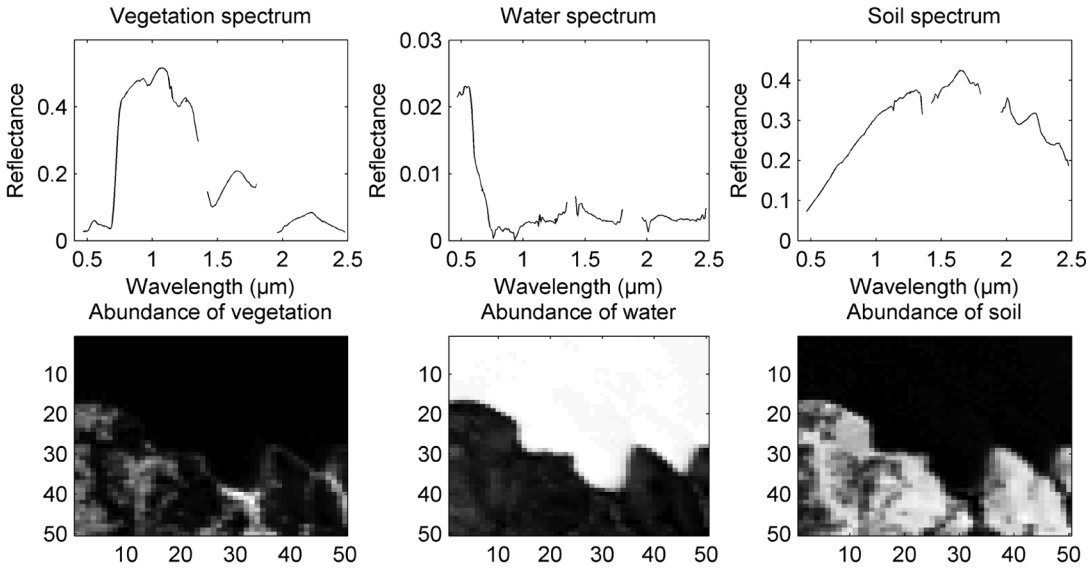


Fig. 8. Top: the $R = 3$ endmember spectra estimated by the algorithm in the Moffett Field scene. Bottom: the corresponding abundance maps (black (respectively, white) means absence (respectively, presence) of the material).

VI. Real AVIRIS Data

A. Moffett Field

This section illustrates the proposed algorithm on a real hyperspectral dataset. The considered hyperspectral image was acquired over Moffett Field (CA, USA) in 1997 by the JPL spectro-imager AVIRIS [56]. This image has been used in many works to illustrate and compare hyperspectral signal processing algorithms [57], [58].

A 50×50 subimage depicted in Fig. 7 (right) has been unmixed using the proposed Bayesian approach. The number of endmembers has been estimated as in [19]. More precisely, we retain the first $R - 1$ eigenvalues identified by PCA that capture 95% of the energy contained into the dataset. As detailed in Section III-B-1, we use also PCA to choose the subset $\mathcal{V}_{R-1}$ defined in (11). After a short burn-in period $N_{\text{bi}} = 50$, estimates of the parameters of interest are computed following the MMSE principle in (40) with $N_r = 450$. The $R = 3$ endmembers recovered by the proposed joint Bayesian LSMA algorithm are depicted in Fig. 8 (top). These endmember spectra are represented in $L = 189$ spectral bands after removing the water absorption bands.[4] These endmembers are characteristic of the coastal area that appears in the image: vegetation, water, and soil. The corresponding abundance maps, shown in Fig. 8 (bottom), are in agreement with the previous results presented in [19].

B. Cuprite

The proposed algorithm has also been evaluated on the "alunite hill" that appears in the Cuprite scene, acquired by AVIRIS over the Cuprite mining site, Nevada, in 1997. The geologic characteristics of this area of interest, represented in Fig. 9, have been investigated in [59], [60]. In this area, the endmembers are dominated by three materials: muscovite, alunite, and

[4]The water vapor absorption bands are usually discarded to avoid poor SNR in these intervals.

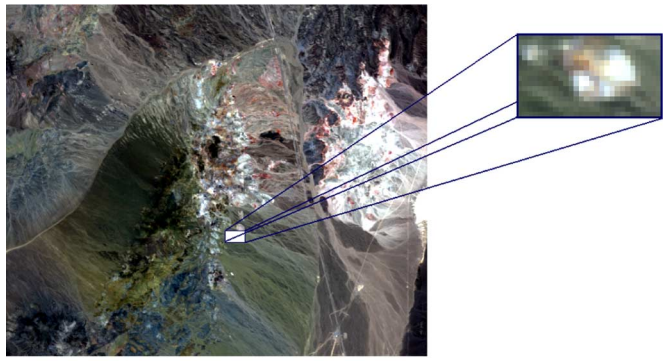


Fig. 9. Real hyperspectral data: Cuprite image acquired by AVIRIS in 1997 (left) and the region of interest (right) represented in synthetic colors.

cuprite. The rescaled endmember spectra estimated by the proposed Bayesian strategy have been depicted in Fig. 10 (top, red lines) in 50 spectral bands of interest (as in [9], [60], and [34]). The spectra estimated by the N-FINDR and VCA algorithms are also shown in this figure (blue and green lines, respectively), as well as the spectral signatures extracted from the United States Geological Survey (USGS) Spectral Library (black lines) [61]. Similarity measures between the estimated spectra and the USGS signatures, expressed in terms of SAD and MSE, are reported in Table VI. The MMSE estimates of the abundance maps computed by the Bayesian unmixing algorithm are depicted in Fig. 10 (bottom). In this figure, particularly in the second map abundance, the alunite hill is clearly recovered. These results confirm the superior performance of the proposed strategy.

VII. Conclusion

This paper presented a Bayesian model as well as an MCMC algorithm for unsupervised unmixing of hyperspectral images,

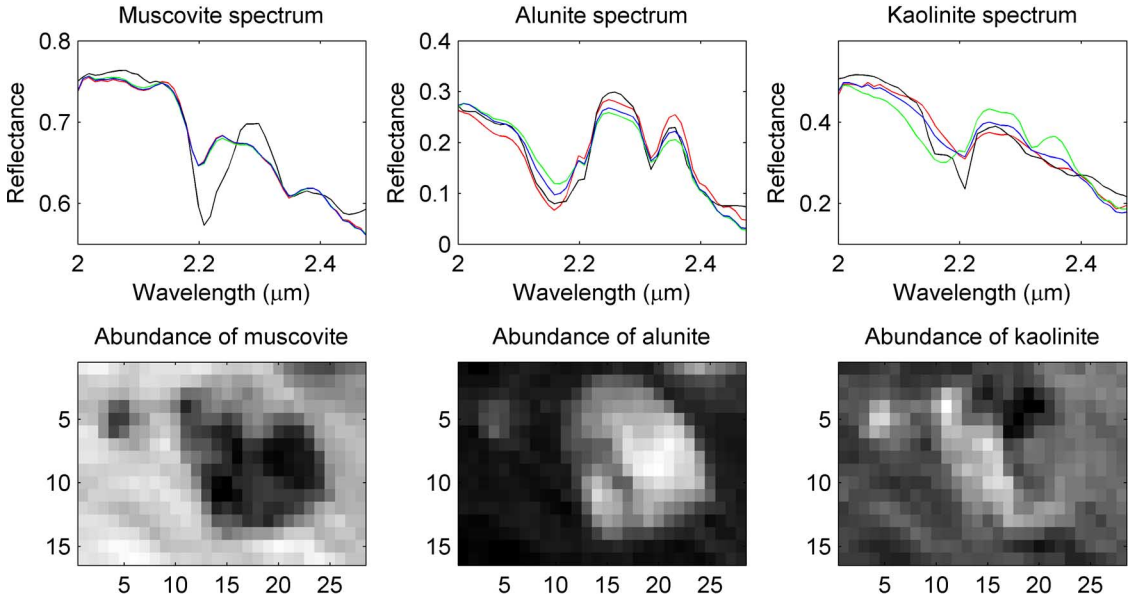


Fig. 10. Top: the $R = 3$ endmember spectra estimated by the Bayesian (red lines), N-FINDR (blue lines) and VCA (green lines) algorithms in the Cuprite scene, compared with the signatures extracted from the USGS spectral library (black lines). Bottom: the corresponding abundance maps estimated by the Bayesian algorithm (black (respectively, white) means absence (respectively, presence) of the material).

TABLE VI
MSEs AND SADs BETWEEN EXTRACTED ENDMEMBERS AND LABORATORY REFLECTANCES FOR THE BAYESIAN, VCA, AND N-FINDR ALGORITHMS

	Bayesian		VCA		N-FINDR	
	MSE^2	SAD	MSE^2	SAD	MSE^2	SAD
Muscovite	0.35	0.39	0.31	0.37	0.33	0.38
Alunite	0.19	1.00	0.36	1.37	0.21	1.03
Kaolinite	0.42	0.77	1.07	1.23	0.44	0.79

i.e., estimating the endmember spectra in the observed scene and their respective abundances for each pixel. Appropriate priors were chosen for the abundance vectors to ensure nonnegativity 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, which can be identified by a standard dimension reduction technique such as PCA, MNF, and HySime, these projections were assigned priors that satisfy positivity constraints on the reconstructed endmember spectra. Due to the complexity of the posterior distribution, 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 and real hyperspectral images illustrated the accuracy of the proposed Bayesian method when compared with other algorithms from the literature. An interesting open question is whether one can improve performance further by folding the intrinsic dimension K of the projection subspace $\mathcal{V}_K$ into the Bayesian framework, e.g., by applying Bayesian PCA or Bayesian latent variable models. This question is a topic of current research. While this paper introduced a Bayesian method in the context of hyperspectral unmixing, the method can also be used for other unmixing applications, such as blind source separation, that satisfy positivity and sum-to-one constraints.

APPENDIX
ON THE CHOICE OF UNIFORM DISTRIBUTIONS AS PRIOR DISTRIBUTIONS FOR $\mathbf{a}_r$ AND THE SIZE OF THE SIMPLEX SOLUTION OF THE BSS PROBLEM

In this appendix, we show that choosing uniform distributions as priors for the abundance vectors allows one to favor *a posteriori*, among two *a priori* equiprobable polytopes that are admissible solutions of the BSS problem, the solution corresponding to the smallest polytope.

Property: Let $\mathbf{M}^{(1)}$ and $\mathbf{M}^{(2)}$ be two R-dimensional convex polytopes of $\mathbb{R}^L$ that are admissible solutions of the BSS constrained problem, i.e.

$$\begin{aligned} \exists \mathbf{A}^{(1)} &= \left[\mathbf{a}_1^{(1)}, \ldots, \mathbf{a}_R^{(1)}\right]^T \subset \mathcal{A}^R \\ \exists \mathbf{A}^{(2)} &= \left[\mathbf{a}_1^{(2)}, \ldots, \mathbf{a}_R^{(2)}\right]^T \subset \mathcal{A}^R \end{aligned} \tag{46}$$

such as $\mathbf{Y} = \mathbf{M}^{(1)}\mathbf{A}^{(1)} = \mathbf{M}^{(2)}\mathbf{A}^{(2)}$ where $\mathcal{A}$ has been defined in (4). Then

$$\begin{aligned} & f\left(\mathbf{M}^{(1)}|\mathbf{Y}\right) \geq f\left(\mathbf{M}^{(2)}|\mathbf{Y}\right) \\ & \Leftrightarrow \\ & \mathrm{vol}\left(\mathcal{S}_{\mathbf{M}^{(1)}}\right) \leq \mathrm{vol}\left(\mathcal{S}_{\mathbf{M}^{(2)}}\right) \end{aligned} \tag{47}$$

where $\mathrm{vol}\left(\mathcal{S}_{\mathbf{M}^{(i)}}\right)$ stands for the volume of the polytope $\mathcal{S}_{\mathbf{M}^{(i)}} \subset \mathbb{R}^L$ introduced in (10) whose vertices are the columns of $\mathbf{M}^{(i)}$.

Proof: First note that, in absence of noise, as $\mathbf{y}_p = \mathbf{M}\mathbf{a}_p$,

$$\mathbf{a}_p \sim \mathcal{U}(\mathcal{A}) \Rightarrow \mathbf{y}_p|\mathbf{M} \sim \mathcal{U}(\mathcal{S}_{\mathbf{M}}) \tag{48}$$

where $\mathcal{U}(\cdot)$ stands for the uniform distribution.

Consequently

$$f(\mathbf{Y}|\mathbf{M}) = \left[\frac{1}{\mathrm{vol}(\mathcal{S}_{\mathbf{M}})}\right]^P \prod_p^P \mathbf{1}_{\mathcal{S}_{\mathbf{M}}}(\mathbf{y}_p) \tag{49}$$

which can be simplified by

$$f(\mathbf{Y}|\mathbf{M}) = \left[\frac{1}{\mathrm{vol}(\mathcal{S}_{\mathbf{M}})}\right]^P \tag{50}$$

since, by definition, the observed pixels $\mathbf{y}_p$ $(p = 1, \ldots, P)$ belong to the solution polytope $\mathcal{S}_{\mathbf{M}}$. Moreover, Bayes' paradigm allows one to state

$$f(\mathbf{M}|\mathbf{Y}) = \frac{f(\mathbf{Y}|\mathbf{M})\, f(\mathbf{M})}{f(\mathbf{Y})}. \tag{51}$$

Since the two solutions $\mathbf{M}^{(1)}$ and $\mathbf{M}^{(2)}$ are *a priori* equiprobable, from (50), it yields

$$\frac{f\left(\mathbf{M}^{(1)}|\mathbf{Y}\right)}{f\left(\mathbf{M}^{(2)}|\mathbf{Y}\right)} = \left[\frac{\mathrm{vol}\left(\mathcal{S}_{\mathbf{M}^{(2)}}\right)}{\mathrm{vol}\left(\mathcal{S}_{\mathbf{M}^{(1)}}\right)}\right]^P. \tag{52}$$

It follows (47). ■

Note that the equiprobability assumption underlying the solutions $\mathcal{S}_{\mathbf{M}^{(2)}}$ and $\mathcal{S}_{\mathbf{M}^{(2)}}$ is not a too restrictive hypothesis. Indeed if the variances s_r^2 $(r = 1, \ldots, R)$ had been chosen such that the prior distribution in (22) is sufficiently flat, then

$$f\left(\mathbf{M}^{(1)}\right) \approx f\left(\mathbf{M}^{(2)}\right). \tag{53}$$

Note also that the projection of the polytope $\mathcal{S}_{\mathbf{M}^{(i)}}$ onto the subset $\mathcal{V}_{R-1} \subset \mathbb{R}^{R-1}$ is the simplex $\mathcal{S}_{\mathbf{T}^{(i)}}$ whose vertices are the columns of $\mathbf{T}^{(i)}$.

ACKNOWLEDGMENT

The authors would like to thank the anonymous referees for their valuable and incisive comments. They are also very grateful to J. Idier and E. Le Carpentier, from IRCCyN Nantes, for interesting discussions regarding this work.

REFERENCES

[1] C.-I Chang, *Hyperspectral Imaging: Techniques for Spectral detection and classification*. New York: Kluwer, 2003.

[2] J. Plaza, R. Pérez, A. Plaza, P. Martínez, and D. Valencia, "Mapping oil spills on sea water using spectral mixture analysis of hyperspectral image data," in *Chemical and Biological Standoff Detection III*, J. O. Jensen and J. Thériault, Eds. Bellingham, WA: SPIE, 2005, vol. 5995, pp. 79–86.

[3] 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.

[4] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, pp. 44–57, Jan. 2002.

[5] R. B. Singer and T. B. McCord, "Mars: Large scale mixing of bright and dark surface materials and implications for analysis of spectral reflectance," in *Proc. 10th Lunar and Planet. Sci. Conf.*, Mar. 1979, pp. 1835–1848.

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

[7] 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.

[8] J. Boardman, "Automating spectral unmixing of AVIRIS data using convex geometry concepts," in *Proc. Summaries 4th Annu. JPL Airborne Geosci. Workshop*, Washington, DC, 1993, vol. 1, pp. 11–14.

[9] M. Winter, "Fast autonomous spectral end-member determination in hyperspectral data," in *Proc. 13th Int. Conf. on Appl. Geologic Remote Sens.*, Vancouver, BC, Apr. 1999, vol. 2, pp. 337–344.

[10] J. M. Nascimento and J. M. Bioucas-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.

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

[12] J. H. Bowles, P. J. Palmadesso, J. A. Antoniades, M. M. Baumback, and L. J. Rickard, "Use of filter vectors and fast convex set methods in hyperspectral analysis," in *Infrared Spaceborne Remote Sensing III*, M. Strojnik and B. F. Andresen, Eds. Philadelphia, PA: SPIE, 1995, vol. 2553, pp. 148–157.

[13] J. Li and J. M. Bioucas-Dias, "Minimum volume simplex analysis: a fast algorithm to unmix hyperspectral data," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, Boston, MA, Jul. 2008, vol. 3, pp. 250–253.

[14] A. Plaza, P. Martinez, R. Perez, and J. Plaza, "A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 650–663, Mar. 2004.

[15] P. J. Martinez, R. M. Pérez, A. Plaza, P. L. Aguilar, M. C. Cantero, and J. Plaza, "Endmember extraction algorithms from hyperspectral images," *Ann. Geophys.*, vol. 49, no. 1, pp. 93–101, Feb. 2006.

[16] N. Keshava, J. P. Kerekes, D. G. Manolakis, and G. A. Shaw, "Algorithm taxonomy for hyperspectral unmixing," in *Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI*, S. S. Shen and M. R. Descour, Eds. Philadelphia, PA: SPIE, 2000, vol. 4049, pp. 42–63.

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

[18] J. Settle, "On the relationship between spectral unmixing and subspace projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 4, pp. 1045–1046, Jul. 1996.

[19] 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 Process.*, vol. 56, pp. 2684–2695, Jul. 2008.

[20] P. Common, C. Jutten, and J. Herault, "Blind separation of sources. Part II: Problems statement," *Signal Process.*, vol. 24, no. 1, pp. 11–20, Jul. 1991.

[21] T. W. Lee, *Independent Component Analysis: Theory and Applications*. Hingham, MA: Kluwer Academic, 1998.

[22] J. Bayliss, J. A. Gualtieri, and R. Cromp, J. M. Seler, Ed., "Analyzing hyperspectral data with independent component analysis," in *Proc. AIPR Workshop Exploiting New Image Sources and Sens.*, Washington, DC, 1998, vol. 3240, pp. 133–143.

[23] N. Dobigeon and V. Achard, L. Bruzzone, Ed., "Performance comparison of geometric and statistical methods for endmembers extraction in hyperspectral imagery," in *Proc. Image and Signal Process. Remote Sens. XI*, Oct. 2005, vol. 5982, pp. 335–344.

[24] J. M. P. Nascimento and J. M. Bioucas-Dias, "Does independent component analysis play a role in unmixing hyperspectral data?," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 1, pp. 175–187, Jan. 2005.

[25] J. M. P. Nascimento and J. M. Bioucas-Dias, "Hyperspectral unmixing algorithm via dependent component analysis," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, Jul. 2007, pp. 4033–4036.

[26] P. Paatero and U. Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol. 5, pp. 111–126, 1994.

[27] P. Sajda, S. Du, T. R. Brown, R. Stoyanova, D. C. Shungu, X. Mao, and L. C. Parra, "Nonnegative matrix factorization for rapid recovery of constituent spectra in magnetic resonance chemical shift imaging of the brain," *IEEE Trans. Med. Imag.*, vol. 23, no. 12, pp. 1453–1465, 2004.

[28] 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.

[29] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 765–777, Mar. 2007.

[30] D. B. Rowe, "A Bayesian approach to blind source separation," *J. Interdisciplin. Math.*, vol. 5, no. 1, pp. 49–76, 2002.

[31] C. Févotte and S. J. Godsill, "A Bayesian approach for blind separation of sparse sources," *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, no. 6, pp. 2174–2188, Nov. 2006.
[32] 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, pp. 4133–4145, Nov. 2006.
[33] N. Dobigeon, S. Moussaoui, and J.-Y. Tourneret, "Blind unmixing of linear mixtures using a hierarchical Bayesian model. Application to spectroscopic signal analysis," in *Proc. IEEE-SP Workshop Stat. and Signal Processing*, Madison, WI, Aug. 2007, pp. 79–83.
[34] L. Parra, K.-R. Mueller, C. Spence, A. Ziehe, and P. Sajda, "Unmixing hyperspectral data," in *Advances in Neural Information Processing Systems (NIPS)*. Cambridge, MA: MIT Press, 2000, vol. 12, pp. 942–948.
[35] N. Dobigeon, S. Moussaoui, J.-Y. Tourneret, and C. Carteret, "Bayesian separation of spectral sources under non-negativity and full additivity constraints," *Signal Process.*, vol. 89, no. 12, pp. 2657–2669, Dec. 2009.
[36] C.-I Chang, "Further results on relationship between spectral unmixing and subspace projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 3, pp. 1030–1032, May 1998.
[37] C.-I Chang, X.-L. Zhao, M. L. G. Althouse, and J. J. Pan, "Least squares subspace projection approach to mixed pixel classification for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 3, pp. 898–912, May 1998.
[38] N. Dobigeon and J.-Y. Tourneret, "Bayesian sampling of structured noise covariance matrix for hyperspectral imagery," Univ. Toulouse, Toulouse, France, 2008 [Online]. Available: http://dobigeon.perso.enseeiht.fr/publis.html
[39] J. Li, "Wavelet-based feature extraction for improved endmember abundance estimation in linear unmixing of hyperspectral signals," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 644–649, Mar. 2004.
[40] 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.
[41] T. Veit, J. Idier, and S. Moussaoui, "Rééchantillonnage de l'échelle dans les algorithmes MCMC pour les problèmes inverses bilinéaires," *Traitement du Signal*, vol. 25, no. 4, pp. 329–343, 2008.
[42] S. Moussaoui, D. Brie, and J. Idier, "Non-negative source separation: range of admissible solutions and conditions for the uniqueness of the solution," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP)*, Philadelphia, PA, Mar. 2005, vol. 5, pp. 289–292.
[43] I. T. Jolliffe, *Principal Component Analysis*. New York: Springer-Verlag, 1986.
[44] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 1, pp. 65–74, Jan. 1988.
[45] C. P. Robert, *The Bayesian Choice: from Decision-Theoretic Motivations to Computational Implementation*, ser. Springer Texts in Statistics, 2nd ed. New York: Springer-Verlag, 2007.
[46] 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, pp. 747–758, Mar. 2002.
[47] 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, pp. 1251–1263, Apr. 2007.
[48] H. Jeffreys, *Theory of Probability*, 3rd ed. London, U.K.: Oxford Univ. Press, 1961.
[49] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. New York: Springer-Verlag, 1999.
[50] N. Dobigeon and J.-Y. Tourneret, "Efficient sampling according to a multivariate Gaussian distribution truncated on a simplex," IRIT/ENSEEIHT/TéSA, Tech. Rep., Mar. 2007 [Online]. Available: http://dobigeon.perso.enseeiht.fr/publis.html
[51] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
[52] C. P. Robert, "Simulation of truncated normal variables," *Statistics and Computing*, vol. 5, pp. 121–125, 1995.
[53] L. Devroye, *Non-Uniform Random Variate Generation*. New York: Springer-Verlag, 1986.
[54] ENVI User's Guide Version 4.0 RSI (Research systems inc.), Boulder, CO, 2003.
[55] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.
[56] AVIRIS Free Data Jet Propulsion Lab. (JPL). Calif. Inst. Technol., Pasadena, CA, 2006 [Online]. Available: http://aviris.jpl.nasa.gov/html/aviris.freedata.html
[57] 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.
[58] T. Akgun, Y. Altunbasak, and R. M. Mersereau, "Super-resolution reconstruction of hyperspectral images," *IEEE Trans. Image Process.*, vol. 14, pp. 1860–1875, Nov. 2005.
[59] R. N. Clark, G. A. Swayze, and A. Gallagher, "Mapping minerals with imaging spectroscopy, U.S. geological survey," *Office of Mineral Resources Bulletin*, vol. 2039, pp. 141–150, 1993.
[60] R. N. Clark *et al.*, "Imaging spectroscopy: Earth and planetary remote sensing with the USGS tetracorder and expert systems," *J. Geophys. Res.*, vol. 108, no. E12, pp. 5-1–5-44, Dec. 2003.
[61] R. N. Clark *et al.*, The U. S. Geological Survey, Digital Spectral Library: Version 1: 0.2 to 3.0 Microns U.S. Geological Survey, 1993 [Online]. Available: http://speclab.cr.usgs.gov

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, in 2004. He received the M.Sc. and Ph.D. degrees in 2004 and 2007, respectively, both in signal processing, from the National Polytechnic Institute of Toulouse.

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

Saïd Moussoui received the State engineering degree from Ecole Nationale Polytechnique, Algiers, Algeria, in 2001, and the Ph.D. degree in 2005 from Université Henri Poincaré, Nancy, France.

He is currently an Associate Professor with the Ecole Centrale de Nantes. Since September 2006, he is with the Institut de Recherche en Communication et Cybernétique de Nantes (IRCCYN, UMR CNRS 6597). His research interests are in statistical signal processing methods and their applications.

Martial Coulon was born in Paris, France, in 1973. He received the Ingénieur degree in computer science and applied mathematics from Ecole Nationale Supérieure d'Electronique, d'Electrotechnique, d'Informatique et d'Hydraulique et Télécommunications (ENSEEIHT), and the M.Sc. degree in signal processing from the National Polytechnic Institute of Toulouse (INPT), France, both in 1996. He received the Ph.D. degree from INPT in 1999.

He is currently an Assistant Professor with the University of Toulouse, (INP-ENSEEIHT, Department of Communications Systems), and a member of the IRIT laboratory (CNRS). His research activities are centered around statistical signal processing and communications systems, with a particular interest to change-point detection, Bayesian inference, spread spectrum systems, and multiuser detection.

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

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

Dr. Tourneret was the Program Chair of the European Conference on Signal Processing (EUSIPCO), in Toulouse, in 2002. He was also member of the organizing committee for the International Conference ICASSP'06, also 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). He is currently serving as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING.

Alfred O. Hero (S'79–M'84–SM'96–F'98) received the B.S. degree (*summa cum laude*) from Boston University, Boston, MA, in 1980, and the Ph.D. degree from Princeton University, Princeton, NJ, in 1984, both in electrical engineering.

Since 1984, he has been with the University of Michigan, Ann Arbor, where he is the R. Jamison and Betty Professor of Engineering. His primary appointment is in the Department of Electrical Engineering and Computer Science and he also has appointments, by courtesy, in the Department of Biomedical Engineering and the Department of Statistics. In 2008, he was awarded the Digiteo Chaire d'Excellence, sponsored by Digiteo Research Park in Paris, France, located at the Ecole Superieure d'Electricite, Gif-sur-Yvette, France. His recent research interests have been in signal detection, pattern analysis, information theory applied to spatio-temporal data analysis. Of particular interest are applications to network security, multimodal sensing and tracking, biomedical imaging, and genomic signal processing.

Dr. Hero is a Fellow of the IEEE and has received the IEEE Signal Processing Society Meritorious Service Award (1998), the IEEE Signal Processing Society Best Paper Award (1998), and the IEEE Third Millennium Medal (2000). He was the President of the IEEE Signal Processing Society (2006-2008) and is the Director-Elect of the IEEE for Division IX (2009).