Variational methods for spectral unmixing
of hyperspectral images

Olivier Eches[1], Nicolas Dobigeon[1], Jean-Yves Tourneret[1]
and Hichem Snoussi[2]

Abstract

This paper studies a variational Bayesian unmixing algorithm for hyperspectral images based on the standard linear mixing model. Each pixel of the image is modeled as a linear combination of *endmembers* (assumed to be known in this paper) weighted by corresponding fractions or *abundances*. These mixture coefficients, constrained to be positive and summing to one, are estimated within a Bayesian framework. This approach requires to define prior distributions for the parameters of interest and the related hyperparameters. This paper assumes the abundances are positive and lower than one, thus ignoring the sum-to-one constraint. Therefore, the abundance coefficients are independent. The corresponding prior distribution is chosen to ensure these new constraints. The complexity of the joint posterior distribution is handled by using variational methods that allow one to approximate the joint distribution of the unknown parameters and hyperparameter. An iterative algorithm successively updates the parameter estimates. Simulations show similar performance than previous Markov chain Monte Carlo unmixing algorithms but with significantly lower computational cost.

Index Terms

Bayesian inference, variational methods, spectral unmixing, hyperspectral images.

I. INTRODUCTION

Hyperspectral imagery has been receiving growing interest in various fields of applications. The problem of spectral unmixing, crucial step in hyperspectral images analysis, has been investigated for several decades in both the signal processing and geoscience communities where many solutions have been proposed (see for instance [1] and references therein). Hyperspectral unmixing (HU) consists of decomposing the measured pixel reflectances into mixtures of pure spectra (called endmembers) whose fractions are referred to as abundances. Assuming the image pixels are linear combinations of endmembers is very common in HU. More precisely, the

linear mixing model (LMM) considers the spectrum of a mixed pixel as a linear combination of endmembers corrupted by additive noise [1] which can be written

$$y = \sum_{r=1}^{R} m_r a_r + n \tag{1}$$

where $m_r = [m_{r,1}, \ldots, m_{r,L}]^T$ denotes the spectrum of the rth material, L is the number of available spectral bands for the image, a_r is the fraction of the rth material in the pixel, R is the number of endmembers present in the observed scene and n is the additive noise considered as white Gaussian, i.e.,

$$n \sim \mathcal{N}\left(0_L, s^2 I_L\right). \tag{2}$$

Supervised algorithms assume that the R endmember spectra m_r are known. In practical applications, they can be obtained by an endmember extraction procedure such as the well-known N-FINDR algorithm developed by Winter [2] or the minimum volume simplex analysis (MVSA) algorithm presented in [3]. Once the endmembers have been determined, the abundances have to be estimated in the so-called *inversion* step. Due to physical considerations, the abundances satisfy the following positivity and sum-to-one constraints

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

A lot of inversion algorithms respecting these constraints have been proposed in the literature. The fully constrained least squares (FCLS) [4] and scaled gradient (SGA) [5] algorithms are two optimization techniques that ensure the positivity and sum-to-one constraints inherent to the unmixing problem. Another interesting approach introduced in [6] consists of assigning appropriate prior distributions to the abundances and to solve the unmixing problem within a Bayesian framework. Although this approach has shown very interesting results, the algorithm relies on the use of Markov chain Monte Carlo (MCMC) methods that are computationally expensive.

Within a Bayesian estimation framework, an alternative to MCMC methods are the so-called *variational methods* [7]. These methods aim to approximate the joint posterior distribution of the parameters $f(\theta|y)$ by a variational distribution of simpler form $q(\theta)$ that minimizes the Kullback-Leibler divergence between the two distributions. The machine learning community has widely adopted variational methods, e.g., to solve problems relying on graphical models

[8], [9]. In the signal processing community, the variational methods have been used as an alternative of MCMC in various topics such as speech processing [10], identification of non-Gaussian auto-regressive models [11] and for inverse problems applied on hyperspectral imagery [12]. In this paper, we present a new variational Bayesian framework based on the Bayesian model developed in [6] for the LMM unmixing problem. In this paper we choose to ignore the sum-to-one constraint for the abundances since it allows these abundances to be considered as independents. However, we constrain these coefficients to be lower than 1 as they represent proportions. We introduce the likelihood function and the prior distributions of the unknown parameters and the associated hyperparameter. Then, the joint posterior distribution is derived by using Bayes theorem. However, this distribution is too complex to obtain closed form expression for the standard Bayesian estimators. In order to solve this problem, we propose to approximate the posterior of interest by a product of simpler distributions based on variational techniques.

The paper is organized as follows. Section 2 recalls the hierarchical Bayesian model initially derived in [6]. Section 3 summarizes the variational method principle and provides the approximate distributions for the different parameters. Simulation results conducted on synthetic and real data are presented in Section 4. Conclusions are reported in Section 5.

II. BAYESIAN MODEL

This section briefly recalls the hierarchical Bayesian model used for the LMM that has been initially introduced in [6].

A. Likelihood

The linear mixing model defined in (1) and the statistical properties of the noise allow us to write $\boldsymbol{y}|\boldsymbol{a}, s^2 \sim \mathcal{N}\left(\boldsymbol{M}\boldsymbol{a}, s^2 \boldsymbol{I}_L\right)$ where $\boldsymbol{M} = [\boldsymbol{m}_1, \ldots, \boldsymbol{m}_R]$. Therefore the likelihood function of $\boldsymbol{y}$ is

$$f\left(\boldsymbol{y}\,|\boldsymbol{a}, s^2\right) \propto \frac{1}{s^L} \exp\left[-\frac{\|\boldsymbol{y} - \boldsymbol{M}\boldsymbol{a}\|^2}{2s^2}\right] \tag{4}$$

where $\propto$ means proportional to and $\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}^T \boldsymbol{x}}$ is the standard ℓ^2 norm. The next sub-section summarizes the prior distributions adapted in [6] for the unknown parameters, i.e., the abundances, the noise variance, and the associated hyperparameter.

B. Parameter prior

1) Abundance prior: A natural choice of prior distribution for the abundances would be to introduce Dirichlet or logistic distributions [13]. However, both prior distributions lead to complicated posteriors which cannot be handled easily using variational methods. Instead, we propose here to relax the sum-to-one constraint and to assign independent uniform priors for the abundances. These uniform priors are defined on the interval $(0, 1)$, reflecting the fact that abundances are proportions. The resulting prior for the abundances can be expressed as

$$f(\boldsymbol{a}) = \prod_{r=1}^{R} f(a_r) = \prod_{r=1}^{R} \mathbf{1}_{(0,1)}(a_r) \tag{5}$$

where $\mathbf{1}_{\mathcal{E}}(\cdot)$ is the indicator function defined on the set $\mathcal{E}$.

2) Noise variance prior: A conjugate inverse-gamma distribution has been chose as prior distribution for the noise variance

$$s^2 | \nu, \delta \sim \mathcal{IG}(\nu, \delta) \tag{6}$$

where $\nu = 1$ (as in [14]) and δ is included in the Bayesian model resulting in a hierarchical strategy. As in any hierarchical Bayesian algorithm, a prior distribution for the hyperparameter δ has to be defined. As in [6], a Jeffreys' prior is assigned to δ, i.e.,

$$f(\delta) \propto \frac{1}{\delta} \mathbf{1}_{\mathbb{R}^+}(\delta). \tag{7}$$

This prior reflects the lack of knowledge regarding δ.

C. Posterior distribution

By using the Bayes' rule, the joint posterior distribution of the unknown parameter vector $\{\boldsymbol{a}, s^2, \delta\}$ is

$$f(\boldsymbol{a}, s^2, \delta | \boldsymbol{y}) \propto \frac{\mathbf{1}_{\mathbb{R}^+}(\delta)}{\delta s^{4+L}} \exp\left(-\frac{\|\boldsymbol{y} - \boldsymbol{M}\boldsymbol{a}\|^2 + 2\delta}{2s^2}\right) \mathbf{1}_{(0,1)^R}(\boldsymbol{a}). \tag{8}$$

The posterior distribution (8) is too complex to derive simple Bayesian estimators of the unknown parameter vector $\boldsymbol{\theta} = \{\boldsymbol{a}, s^2, \delta\}$. To alleviate this problem, the estimators were approximated in [6] from samples of (8) generated by MCMC methods. The next section studies an alternative approximating the Bayesian estimator of $\boldsymbol{\theta}$ using variational methods.

III. POSTERIOR DISTRIBUTION APPROXIMATION

A. *Variational methods*

The main idea of variational methods is to approximate the posterior distribution of an unknown parameter vector $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_n\}$ by a so-called *variational* distribution of simpler form. A very popular choice consists of factorizing the joint posterior[1] into disjoint factors [15, Chap. 10]

$$f(\boldsymbol{\theta}|\boldsymbol{y}) \approx q(\boldsymbol{\theta}) = \prod_{i=1}^{N} q(\theta_i). \tag{9}$$

The optimal approximation is obtained by minimizing the Kullback-Leibler Divergence (KLD) between the distribution $f(\cdot)$ and its approximate $q(\cdot)$

$$\mathrm{KL}(f, q) = \int q(\boldsymbol{\theta}) \log \left(\frac{q(\boldsymbol{\theta})}{f(\boldsymbol{\theta}|\boldsymbol{y})} \right) d\boldsymbol{\theta}.$$

Using Bayes' relation, the KLD can be reformulated in this way

$$\mathrm{KL}(f, q) = -\mathcal{F}(q) + \log f(\boldsymbol{y}), \tag{10}$$

where $\mathcal{F}(q)$, known as the *mean-field* bound to the log-evidence $\log f(\boldsymbol{y})$, is written as [15, Chap. 10]

$$\mathcal{F}(q) = \int q(\boldsymbol{\theta}) \log \left(\frac{f(\boldsymbol{\theta}, \boldsymbol{y})}{q(\boldsymbol{\theta})} \right) d\boldsymbol{\theta}. \tag{11}$$

Since the KLD is strictly positive, minimizing $\mathrm{KL}(f : q)$ is equivalent to maximizing $\mathcal{F}(q)$. Therefore, the solution of the optimization problem $\frac{d\mathcal{F}}{dq} = 0$ provides the approximating functions. As demonstrated in [15, p. 466], the optimal solution $q(\cdot)$ of this problem is given by the following expression

$$q(\theta_i) \propto \exp\langle \log f(\boldsymbol{y}, \boldsymbol{\theta}) \rangle_{\boldsymbol{\theta}_{-i}}. \tag{12}$$

where $\langle \cdot \rangle_{\boldsymbol{\theta}_{-i}}$ denotes the expectation with respect to the $q(\cdot)$ distributions over the $j \neq i$ components of $\boldsymbol{\theta}$. Once the optimal approximation has been computed, the unknown parameters can be estimated by the means of the variational distribution according to the minimum mean square error (MMSE) principle.

---

[1]The approximate distribution obviously depends on $\boldsymbol{y}$. For the sake of conciseness, the notation $q(\boldsymbol{\theta})$ abusively holds for this distribution.

B. Posterior distribution approximation

We propose in this paper to approximate the posterior distribution (8) of the unknown parameter vector $\boldsymbol{\theta} = \{\boldsymbol{a}, s^2, \delta\}$ as follows

$$f(\boldsymbol{\theta}|\boldsymbol{y}) \approx q(\boldsymbol{a})q(s^2)q(\delta), \tag{13}$$

where the individual approximating distributions $q(\boldsymbol{a})$ $q(s^2)$ and $q(\delta)$ are derived below.

1) Abundance vector: Since the abundance coefficients are independent, this distribution can be written as the product of the marginals $q(a_r)$

$$q(\boldsymbol{a}) = \prod_{r=1}^{R} q(a_r).$$

Using (12), the variational distribution[2] for a_r is

$$q(a_r) \propto \exp \left\langle -\frac{L}{2} \log s^2 - \frac{\|\boldsymbol{y} - \boldsymbol{M}\boldsymbol{a}\|^2}{2s^2} + \log f(\boldsymbol{a}) \right\rangle,$$

where $\boldsymbol{a}_{\text{-}r} = [a_1, \ldots, a_{r-1}, a_{r+1}, \ldots, a_R]$. Let $\boldsymbol{\varepsilon}_r = \boldsymbol{y} - \sum\limits_{\substack{i=1 \\ i \neq r}}^{R} a_i \boldsymbol{m}_i$. Straightforward computations (see Appendix) show the marginal distributions $q(a_r)$ are truncated normal

$$a_r \sim \mathcal{N}_{(0,1)} \left(\mu_r, \sigma_r^2 \right), \tag{14}$$

where $\mathcal{N}_{\mathcal{A}} \left(m, \nu^2 \right)$ stands for the Gaussian distribution truncated on $\mathcal{A}$ with hidden mean m and hidden variance ν^2 and

$$\begin{cases} \sigma_r^2 = \left[\langle \frac{1}{s^2} \rangle \|\boldsymbol{m}_r\|^2 \right]^{-1}, \\ \mu_r = \sigma_r^2 \langle \boldsymbol{\varepsilon}_r \rangle^T \boldsymbol{m}_r \langle \frac{1}{s^2} \rangle. \end{cases} \tag{15}$$

Finally, setting $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$ and using $\langle a_r \rangle = \int a_r q(a_r) da_r$, the updating rules for a_r is

$$\langle a_r \rangle = \mu_r + \sigma_r \frac{\phi\left(\frac{-\mu_r}{\sigma_r}\right) - \phi\left(\frac{1-\mu_r}{\sigma_r}\right)}{\frac{1}{2}\left[\text{erf}\left(\frac{1-\mu_r}{\sigma_r}\right) - \text{erf}\left(\frac{-\mu_r}{\sigma_r}\right)\right]}. \tag{16}$$

Similarly, the variance of $q(a_r)$ is given by

$$\text{Var}(a_r) = \sigma_r^2 \left\{ 1 + \frac{\left(\frac{1-\mu_r}{\sigma_r}\right)\phi\left(\frac{-\mu_r}{\sigma_r}\right) - \left(\frac{1-\mu_r}{\sigma_r}\right)\phi\left(\frac{1-\mu_r}{\sigma_r}\right)}{\frac{1}{2}\left[\text{erf}\left(\frac{1-\mu_r}{\sigma_r}\right) - \text{erf}\left(\frac{-\mu_r}{\sigma_r}\right)\right]} - \left[\frac{\phi\left(\frac{-\mu_r}{\sigma_r}\right) - \phi\left(\frac{1-\mu_r}{\sigma_r}\right)}{\frac{1}{2}\left[\text{erf}\left(\frac{1-\mu_r}{\sigma_r}\right) - \text{erf}\left(\frac{-\mu_r}{\sigma_r}\right)\right]}\right]^2 \right\}. \tag{17}$$

---

[2]We have ignored the $\boldsymbol{\theta}_{\text{-}i}$ in the notation $\langle \cdot \rangle_{\boldsymbol{\theta}_{\text{-}i}}$ for conciseness.

2) Noise variance: By using (12), the approximated distribution for the noise variance s^2 is the following inverse-gamma distribution

$$q(s^2) \propto \frac{1}{s^{L+2}} \exp\left[-\frac{1}{s^2}\left(\frac{1}{2}\langle\|\boldsymbol{y} - \boldsymbol{Ma}\|^2\rangle + \langle\delta\rangle\right)\right]. \tag{18}$$

Therefore, the noise variance updating rule is

$$\langle s^2 \rangle = \frac{2}{L}\left(\frac{1}{2}\langle\|\boldsymbol{y} - \boldsymbol{Ma}\|^2\rangle + \langle\delta\rangle\right), \tag{19}$$

with $\langle\|\boldsymbol{y} - \boldsymbol{Ma}\|^2\rangle = \text{tr}\left[\boldsymbol{yy}^T - 2\boldsymbol{y}\langle\boldsymbol{a}^T\rangle\boldsymbol{M}^T + \boldsymbol{M}\langle\boldsymbol{aa}^T\rangle\boldsymbol{M}^T\right]$ and $\langle\boldsymbol{aa}^T\rangle = \boldsymbol{\Lambda} + \langle\boldsymbol{a}\rangle\langle\boldsymbol{a}\rangle^T$ where $\boldsymbol{\Lambda}$ covariance matrix of $\boldsymbol{a}$ whose diagonal elements are given by Eq. 17. Note that, since this distribution is an inverse-gamma, the term $\langle\frac{1}{s^2}\rangle$ in Eq. (15) is equal to $\frac{L/2}{(L/2-1)\langle s^2\rangle}$.

3) Hyperparameter: The approximated distribution for δ is the following gamma distribution

$$q(\delta) \propto \delta^{\nu-1} \exp\left\{-\delta\left\langle\frac{1}{s^2}\right\rangle\right\}. \tag{20}$$

The mean of this approximated distribution is

$$\langle\delta\rangle = \frac{\nu}{\langle\frac{1}{s^2}\rangle}. \tag{21}$$

To summarize, the variational Bayesian inference is conducted via an iterative algorithm consisting of updating $q(\boldsymbol{a})$, $q(s^2)$ and $q(\delta)$ using (16), (19) and (21). When convergence has been achieved, the abundance means are normalized to ensure they sum to one. Then, the means of these variational distributions are used to approximate the MMSE estimates of the unknown parameters. The complete algorithm is detailed in Algo. III-B3.

Variational methods applied to hyperspectral unmixing

1) Initialization:
   - Sample $\langle\boldsymbol{a}\rangle^{(0)}$ using Eq. (5),
   - Sample $\langle s^2\rangle^{(0)}$ using Eq. (**??**),
   - Sample $\langle\delta\rangle^{(0)}$ using Eq. (5),
   - Set $i = 1$,

2) Iterations: while $\|\langle\boldsymbol{a}\rangle^{(i+1)} - \langle\boldsymbol{a}\rangle^{(i)}\|^2 < \varepsilon$ do
   - Update $\langle\boldsymbol{a}\rangle^{(i)}$ using Eq. (16),
   - Update $\langle s^2\rangle^{(i)}$ using Eq. (19),
   - Update $\langle\delta\rangle^{(i)}$ using Eq. (21),
   - Update $i = i + 1$,

IV. SIMULATION RESULTS

A. Synthetic data

This section studies the accuracy of the proposed variational algorithm for the unmixing of a synthetic pixel, resulting from the combination of $R = 3$ endmembers whose abundance vector is $a = [0.12, \ 0.37, \ 0.51]^T$ and with noise variance $s^2 = 0.001$ (corresponding to a signal-to-noise ratio SNR= 20dB). The endmember spectra used in the mixture have been extracted from the ENVI spectral library (construction concrete, green grass and micaceous loam, as in [6]) observed in $L = 276$ spectral bands. Fig. 1 shows the average of estimates (with the corresponding standard deviation) of the 50 Monte Carlo simulation results obtained for the abundances (left) and noise variance (right) as functions of the algorithm iterations. These figures show us the algorithm clearly converges to the actual parameter values with a weak standard deviation.
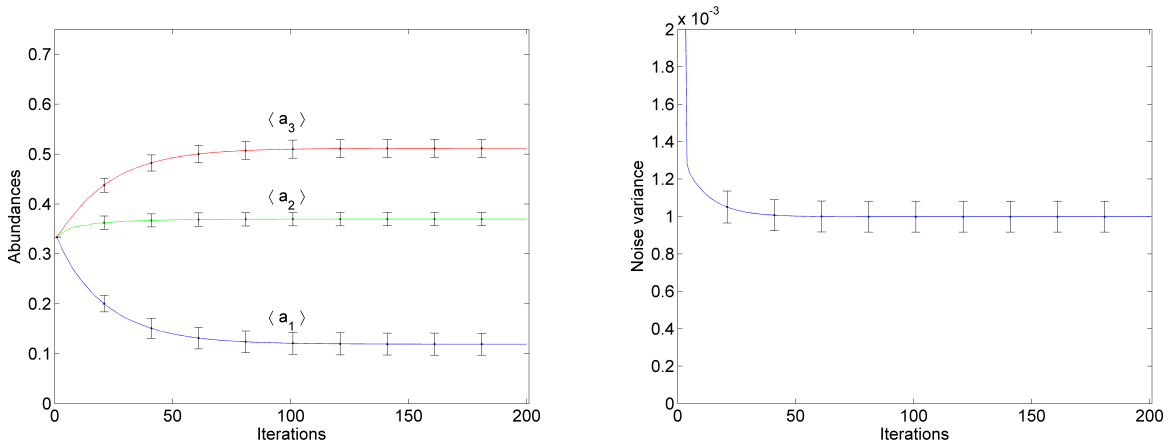

Fig. 1.   Average of estimates (solid line) and standard deviation (vertical bars) of $\langle a \rangle$ (left) and $\langle s^2 \rangle$ (right).

The proposed variational Bayesian algorithm has to be compared with other existing unmixing strategies. For this comparison, we propose to consider the MCMC-based strategy derived in [6] which is based on the same hierarchical Bayesian model where the abundance sum-to-one constraint has not been ignored. An image of $P = 625$ synthetic pixels has been generated from the mixture of $R = 6$ endmembers, with $s^2 = 1 \times 10^{-4}$ (SNR $\approx$ 30dB). The MVSA algorithm [3] has been used to estimate the endmember spectra. Finally, the material abundances have been estimated using the proposed variational algorithm and the MCMC strategy of [6]. To evaluate the algorithm performance, the mean square errors (MSEs) of the abundances have then been

computed as

$$\text{MSE}^2 = \frac{1}{P} \sum_{p=1}^{P} (\hat{\boldsymbol{a}}_p - a_p)^2, \tag{22}$$

where $\hat{\boldsymbol{a}}_p$ denotes the estimate of the abundance vector a_p of the pth pixel. Table I reports the corresponding results as well as the time of simulation required by each algorithm. These results show that the variational approach achieves similar performance to the Gibbs sampler of [6] with a significantly reduced computational cost.

TABLE I

MEAN SQUARE ERRORS OF ABUNDANCE VECTORS.

	LMM Gibbs	Variational
MSE^2	1.5×10^{-3}	1.6×10^{-3}
Time (sec.)	7.85×10^3	7.96×10^2

B. Real hyperspectral images

The proposed algorithm has also been applied to a real hyperspectral dataset, acquired over Moffett Field (CA, USA) in 1997 by the JPL spectro-imager AVIRIS. Previous works have used this image to illustrate and compare unmixing strategies for hyperspectral images, e.g., [16]. The region of interest is a 50×50 image represented in Fig. 2. The data set has been reduced from the original 224 bands to 189 bands by removing water absorption bands. A principal component analysis has been conducted as a preprocessing step to determine the number of endmembers present in the scene yielding $R = 3$. The $R = 3$ endmember spectra have been extracted with the N-FINDR algorithm proposed by Winter in [2] and are given Fig. 3. The abundance maps estimated by the proposed algorithm are represented in Fig. 4 (top). These results can be compared with those obtained with the MCMC-based algorithm that are depicted in Fig. 4 (bottom). The abundance maps obtained with the two algorithms are clearly similar.

V. CONCLUSION

A new variational Bayesian algorithm was proposed for hyperspectral image unmixing. Prior distributions were assigned to the parameters and hyperparameter associated with the proposed
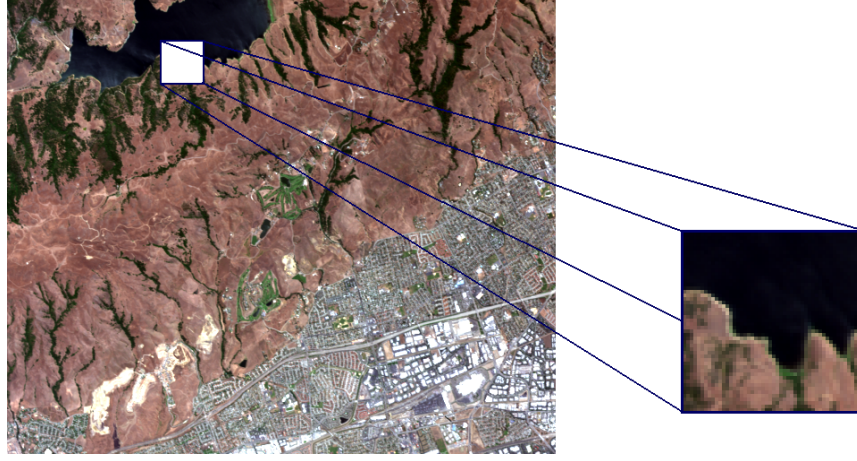
Fig. 2. Real hyperspectral data: Moffett field acquired by AVIRIS in 1997 (left) and the region of interest shown in true colors (right).
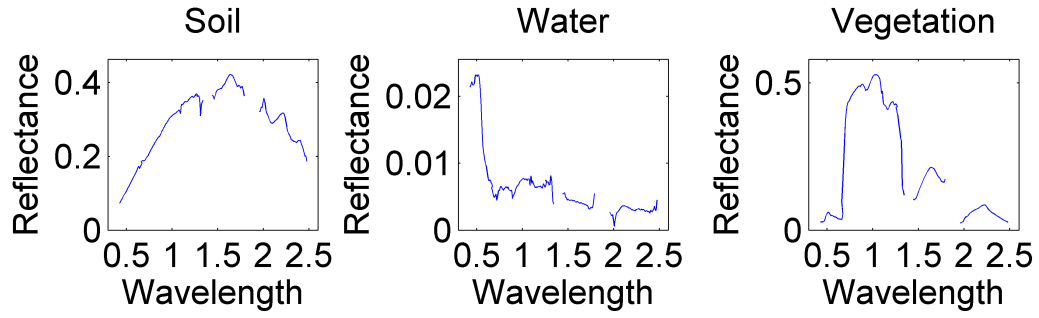

Fig. 3. The $R = 3$ endmember spectra obtained by the N-FINDR algorithm.

linear mixing model To handle the complexity of the resulting posterior distribution, a variational approximation is proposed as an alternative of the use of MCMC method used in [6]. The approximated distribution for the abundances was identified as a truncated Gaussian distribution. An iterative algorithm was employed to update the different variational distributions. At convergence, the mean of these distributions were used as estimates of the unknown parameters. Simulations showed satisfactory results with significantly lower computational cost than MCMC methods.
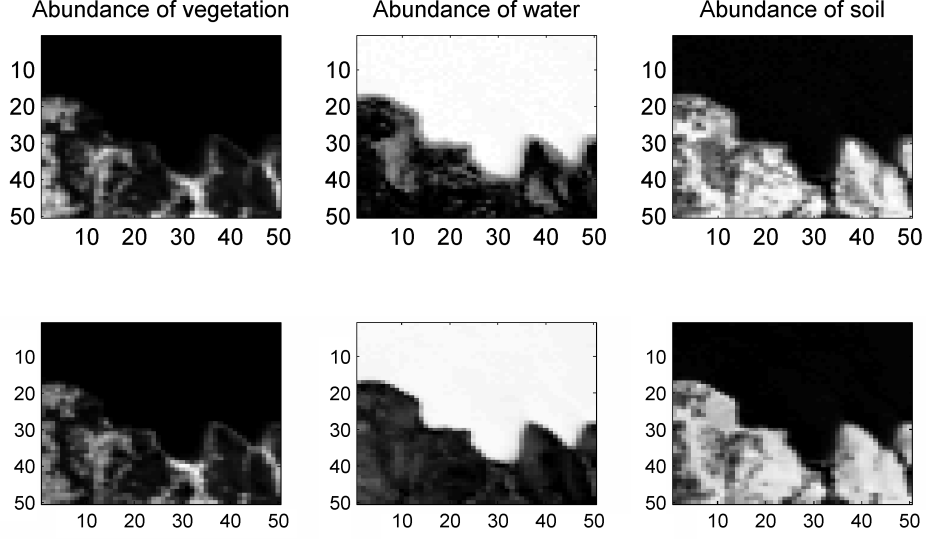
Fig. 4. Top: fraction maps estimated by the variational Bayesian algorithm. Bottom: fraction maps estimated by the MCMC Bayesian algorithm (black (resp. white) means absence (resp. presence) of the material).

APPENDIX A

APPROXIMATE MARGINALS $q(a_r)$ COMPUTATIONS

Since the abundance coefficients are independent, this distribution can be written as the product of the marginals $q(a_r)$

$$q(\boldsymbol{a}) = \prod_{r=1}^{R} q(a_r).$$

Using (12), the variational distribution for a_r is

$$q(a_r) \propto \exp \left\langle \log f(\boldsymbol{y}|\boldsymbol{a}, s^2) + \log f(\boldsymbol{a}) \right\rangle_{s^2, \delta, \boldsymbol{a}_{\text{-}r}},$$

where $\boldsymbol{a}_{\text{-}r} = [a_1, \ldots, a_{r-1}, a_{r+1}, \ldots, a_R]$. By ignoring the different terms that do not depend on $\boldsymbol{a}$ (considered as constant), this distribution can be rewritten as

$$q(\boldsymbol{a}) \propto \exp \left(- \left\langle \frac{1}{s^2} \right\rangle_{s^2} \frac{\langle \|\boldsymbol{y} - \boldsymbol{M}\boldsymbol{a}\|^2 \rangle_{\boldsymbol{a}_{\text{-}r}}}{2} \right) \mathbf{1}_{(0,1)}(\boldsymbol{a}).$$

By letting $\boldsymbol{\varepsilon}_r = \boldsymbol{y} - \sum_{\substack{i=1 \\ i \neq r}}^{R} a_i \boldsymbol{m}_i$, we have

$$\left\langle \frac{1}{s^2} \right\rangle_{s^2} \langle \|\boldsymbol{y} - \boldsymbol{M}\boldsymbol{a}\|^2 \rangle_{\boldsymbol{a}_{\text{-}r}} = \left\langle \frac{1}{s^2} \right\rangle_{s^2} \langle \|\boldsymbol{\varepsilon}_r - a_r \boldsymbol{m}_r\|^2 \rangle_{\boldsymbol{a}_{\text{-}r}},$$

$$= \left\langle \frac{1}{s^2} \right\rangle \left(\|\boldsymbol{m}_r\|^2 a_r^2 - 2\langle \boldsymbol{\varepsilon}_r \rangle_{\boldsymbol{a}_{\text{-}r}}^{T} \boldsymbol{m}_r a_r + \langle \boldsymbol{\varepsilon}_r \rangle_{\boldsymbol{a}_{\text{-}r}}^{T} \langle \boldsymbol{\varepsilon}_r \rangle_{\boldsymbol{a}_{\text{-}r}} \right),$$

with $\langle \boldsymbol{\varepsilon}_r \rangle_{\boldsymbol{a}_{\text{-}r}} = \boldsymbol{\varepsilon}_r = \boldsymbol{y} - \sum\limits_{\substack{i=1 \\ i \neq r}}^{R} \langle a_i \rangle \boldsymbol{m}_i.$

By identification, the marginal distribution is then equal to

$$q(a_r) \propto \exp\left[-\frac{(a_r - \mu_r)^2}{2\sigma_r^2} \right] \mathbf{1}_{(0,1)}(\boldsymbol{a}), \tag{23}$$

with

$$\begin{cases} \sigma_r^2 = \left[\langle \frac{1}{s^2} \rangle_{s^2} \|\boldsymbol{m}_r\|^2 \right]^{-1}, \\ \mu_r = \frac{\boldsymbol{\varepsilon}_r^T \boldsymbol{m}_r}{\|\boldsymbol{m}_r\|^2}, \end{cases} \tag{24}$$

which is equivalent to $\mu_r = \sigma_r^2 \boldsymbol{\varepsilon}_r^T \boldsymbol{m}_r \langle \frac{1}{s^2} \rangle_{s^2}.$

This distribution corresponds to a truncated normal distribution.

REFERENCES

[1] N. Keshava and J. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, pp. 44–56, Jan. 2002.

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

[3] 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)*, vol. 3, Boston, MA, Jul. 2008, pp. 250–253.

[4] 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. and Remote Sensing*, vol. 39, no. 3, pp. 529–545, March 2001.

[5] C. Theys, N. Dobigeon, J.-Y. Tourneret, and H. Lantéri, "Linear unmixing of hyperspectral images using a scaled gradient method," in *Proc. IEEE-SP Workshop Stat. and Signal Processing (SSP)*, Cardiff, UK, Aug. 2009, pp. 729–732.

[6] N. Dobigeon, J.-Y. Tourneret, and C.-I Chang, "Semi-supervised linear spectral using a hierarchical Bayesian model for hyperspectral imagery," *IEEE Trans. Signal Processing*, vol. 56, no. 7, pp. 2684–2696, July 2008.

[7] T. S. Jaakkola and M. Jordan, "Bayesian parameter estimation via variational methods," *Statistics and Computing*, vol. 10, no. 1, pp. 25–37, Jan. 2000.

[8] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul, "An introduction to variational methods for graphical models," *Machine Learning*, vol. 37, no. 2, pp. 183–233, Jan. 1999.

[9] H. Attias, "A variational Bayesian framework for graphical models," in *Advances in Neural Information Processing Systems*, S. Solla, T. K. Leen, and E. K. L. Mueller, Eds. MIT Press, 2000, vol. 12, pp. 209–215.

[10] S. Watanabe, Y. Minami, A. Nakamura, and N. Ueda, "Variational Bayesian estimation and clustering for speech recognition," *IEEE Trans. Speech, Audio Proc.*, vol. 12, no. 4, pp. 365–381, July 2004.

[11] S. J. Roberts and W. D. Penny, "Variational Bayes for generalized autoregressive models," *IEEE Trans. Signal Processing*, vol. 50, no. 2, pp. 2245–2257, Sept. 2002.

[12] N. Bali and A. Mohammad-Djafari, "Bayesian approach with hidden Markov modeling and mean field approximation for hyperspectral data analysis," *IEEE Trans. Image Processing*, vol. 17, no. 2, pp. 217–225, Feb. 2008.

[13] A. Gelman, F. Bois, and J. Jiang, "Physiological pharmacokinetic analysis using population modeling and informative prior distributions," *J. Amer. Math. Soc.*, vol. 91, no. 436, pp. 1400–1412, Dec. 1996.

[14] 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 Processing*, vol. 55, no. 4, pp. 1251–1263, April 2007.

[15] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer Verlag, 2006.

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