SPARSE SEMI-SUPERVISED HYPERSPECTRAL UNMIXING USING A NOVEL ITERATIVE BAYESIAN INFERENCE ALGORITHM

Konstantinos E. Themelis[1,2]*, Athanasios A. Rontogiannis*[1] *and Konstantinos Koutroumbas*[1]

[1]Inst. for Space Applications and Remote Sensing, NOA, 152 36, Athens, Greece
[2] Dept. of Informatics and Telecommunications, University of Athens, 157 84, Athens, Greece
{ themelis, tronto, koutroumbas } @space.noa.gr

ABSTRACT

In this paper a novel hierarchical Bayesian model for sparse semi-supervised hyperspectral unmixing is presented. Adopting the sparsity hypothesis and taking into account the convex constraints of the estimation problem, suitable priors are selected for the model parameters. Then, a new low-complexity, iterative conditional expectations algorithm is developed to perform Bayesian inference. The proposed method converges fast to a sparse solution, which offers improved estimation accuracy. The theoretical results presented in the paper are fully verified by simulations both on synthetic and real hyperspectral data.

Index Terms— Constrained sparse regression, hyperspectral images, linear spectral unmixing

1. INTRODUCTION

Hyperspectral remote sensing has gained considerable attention in recent years, due to its wide range of applications, e.g. environmental monitoring and terrain classification [1], and the maturation of the required technology. Hyperspectral sensors are able to produce image cubes by sampling the electromagnetic spectrum in tens or hundreds of contiguous spectral bands from the visible to the near-infrared region. However, due to their low spatial resolution, more than one different materials can be mixed in a single pixel, which calls for spectral unmixing, [1]. In spectral unmixing, the measured spectrum of a mixed pixel is decomposed into a collection of constituent spectra, called *endmembers*. In addition, a set of corresponding to endmembers fractions, called *abundances*, are obtained, that indicate the proportion of each endmember contributing to the pixel.

Semi-supervised unmixing, [2, 3], which is the case considered in this paper, assumes that the spectral signatures of all the endmembers present in the image are available. The objective of semi-supervised unmixing is (a) to determine how many and which endmembers are present in the mixed pixel under study and (b) to use the selected endmembers to estimate the corresponding abundance fractions. In the latter case, the abundance fractions should satisfy two constraints in order to remain physically meaningful; they should be non-negative and sum to one. It should be emphasized that since only a small number of endmembers are mixed in a single pixel, the solution to this problem is *inherently sparse*. In this framework, a number of semi-supervised unmixing techniques have been recently proposed in [3, 4, 5], based on ℓ_1-norm penalization to enhance sparsity.

In this paper, a novel hierarchical Bayesian model is presented, which is based on the sparsity hypothesis and non-negativity property of the abundance vector. Appropriate prior distributions are assigned to the unknown parameters of the model, which reflect prior knowledge about their natural characteristics. Since the resulting joint posterior distribution of the model does not possess a tractable analytical form, a novel iterative method is developed, which can be considered as a deterministic approximation of the Gibbs sampler. More specifically, instead of sampling the conditional posterior distributions, as in conventional Gibbs sampling schemes [2], the proposed method uses their corresponding expectations, which can be expressed in tractable closed forms. The proposed Bayesian inference algorithm iterates through the derived conditional expectations, and produces a point estimate of the abundance vector. In addition, this estimate turns out to be sparse, thus, verifying the sparsity promoting nature of the proposed Bayesian approach. Besides its fast convergence, the new algorithm is computationally efficient and offers improved estimation accuracy, as corroborated by comparing its performance with that of other related algorithms via computer simulations.

Notation: We use lowercase boldface and uppercase boldface letters to represent vectors and matrices respectively. With $(\cdot)^T$ we denote transposition, and with $\|\cdot\|_2$ the ℓ_2 norm, ($\|\mathbf{x}\|_2^2 = \mathbf{x}^T\mathbf{x}$). The determinant of a matrix or the absolute value of a scalar is denoted by $|\cdot|$, while $\mathrm{diag}(\mathbf{x})$ stands for a diagonal matrix, that contains the elements of vector $\mathbf{x}$ on its diagonal. Finally, $\mathbf{0}$ denotes the zero vector, $\mathbf{1}$ the all-ones vector, and $\mathbf{I}_K$ is the $K \times K$ identity matrix.

2. PROBLEM FORMULATION

Let $\mathbf{y}$ denote a $M \times 1$ hyperspectral image pixel vector containing the measured reflectance values of the pixel in M spectral bands. Let $\mathbf{\Phi} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_N]$ stand for the $M \times N$ endmember matrix, where N is the number of all available endmembers and also let $\mathbf{w} = [w_1, w_2, \ldots, w_N]^T$ be the pixel's corresponding vector of abundances. Adopting the linear mixture model (LMM), every pixel $\mathbf{y}$ of the image is modeled as

$$\mathbf{y} = \mathbf{\Phi}\mathbf{w} + \mathbf{n}, \tag{1}$$

where the additive noise $\mathbf{n}$ is assumed to be a zero-mean, Gaussianly distributed random vector, with independent and identically distributed (i.i.d.) elements, i.e., $\mathbf{n} \sim \mathcal{N}(\mathbf{n}|0, \beta^{-1}\mathbf{I}_M)$, where β denotes the inverse of the noise variance (precision). Due to physical considerations, the abundance vector $\mathbf{w}$ satisfies the non-negativity and additivity constraints, i.e., $w_i \geq 0,\ i = 1, 2, \ldots, N$, and $\sum_{i=1}^{N} w_i = 1$,

In this work, a semi-supervised hyperspectral unmixing technique is introduced, where the full endmember matrix $\mathbf{\Phi}$ is assumed to be known a priori. This mixing matrix $\mathbf{\Phi}$ can either stem from a spectral library or it can be determined using an endmember extraction technique, e.g., [6, 7]. The aim is to estimate the abundance vector $\mathbf{w}$, subject to the previous constraints. Additionally, the reasonable assumption that only few of the available N endmembers are present in each pixel, gives rise to the sparsity property of $\mathbf{w}$, which is exploited by the hierarchical Bayesian model introduced in the following sections.

3. HIERARCHICAL BAYESIAN MODEL

3.1 Likelihood

The LMM of (1) and the Gaussian property of the noise vector $\mathbf{n}$ yield the following likelihood function for $\mathbf{y}$

$$p(\mathbf{y}|\mathbf{w}, \beta) = (2\pi)^{-\frac{M}{2}}\, \beta^{\frac{M}{2}} \exp\left[-\frac{\beta}{2}\|\mathbf{y} - \mathbf{\Phi}\mathbf{w}\|_2^2\right]. \tag{2}$$

3.2 Parameters' prior distributions

In this section, the prior distributions for the model parameters $\mathbf{w}$ and β are introduced. Accounting for the non-negativity property of $\mathbf{w}$, and assuming that all w_i's are i.i.d., a normal distribution truncated on the non-negative orthant $\mathbf{R}_+^N$ of the N-dimensional Euclidean space $\mathcal{R}^N$ is assigned to $\mathbf{w}$, i.e.,

$$p(\mathbf{w}|\boldsymbol{\gamma},\beta)=\prod_{i=1}^{N}\left[\mathcal{N}(w_i|0,\frac{\gamma_i}{\beta})\,\mathbf{I}_{\mathbf{R}_+^1}(w_i)\right]$$
$$=2^N\,(2\pi)^{-\frac{N}{2}}\,\beta^{\frac{N}{2}}\,|\boldsymbol{\Lambda}|^{\frac{1}{2}}\exp\left[-\frac{\beta}{2}\mathbf{w}^T\boldsymbol{\Lambda}\mathbf{w}\right]\mathbf{I}_{\mathbf{R}_+^N}(\mathbf{w}) \quad (3)$$

where $\mathbf{R}_+^1$ is the set of non-negative real numbers, $\mathbf{I}_{\mathbf{R}_+^N}(\cdot)$ is the indicator function[1] for $\mathbf{R}_+^N$, $\boldsymbol{\gamma}=[\gamma_1,\gamma_2,\ldots,\gamma_N]^T$ is a $N\times 1$ vector of hyperparameters, $\gamma_i\geq 0$, and $\boldsymbol{\Lambda}^{-1}=\mathrm{diag}(\boldsymbol{\gamma})$.

For β, a conjugate Gamma prior with respect to the Gaussian likelihood of (2) is selected, expressed as

$$p(\beta|\kappa,\theta)=\Gamma(\beta|\kappa,\theta)=\frac{\theta^\kappa}{\Gamma(\kappa)}\beta^{\kappa-1}\exp\left[-\theta\beta\right], \quad (4)$$

where $\beta\geq 0$, and $\kappa\geq 0$, $\theta\geq 0$ are the distribution parameters.

3.3 Hyperparameters' prior distributions

This work extends the model of [8,9], by assigning an independent Gamma distribution to every γ_i, each parameterized by a distinct hyperparameter λ_i, i.e.,

$$p(\gamma_i|\lambda_i)=\Gamma(\gamma_i|1,\frac{\lambda_i}{2})=\frac{\lambda_i}{2}\exp\left[-\frac{\lambda_i}{2}\gamma_i\right],\ i=1,2,\ldots,N, \quad (5)$$

Then, the combination of the hierarchical priors given in (3) and (5) leads to a sparsity-promoting, non-negative (truncated) Laplace distribution for $\mathbf{w}$ (this formulation gives rise to the so-called Bayesian Lasso [9]). This distribution can be obtained by marginalizing the hyperparameter vector $\boldsymbol{\gamma}$ from the model, i.e.,

$$p(\mathbf{w}|\boldsymbol{\lambda},\beta)=\int p(\mathbf{w}|\boldsymbol{\gamma},\beta)p(\boldsymbol{\gamma}|\boldsymbol{\lambda})d\boldsymbol{\gamma}$$
$$=\beta^{\frac{N}{2}}\,|\boldsymbol{\Psi}|^{\frac{1}{2}}\exp\left[-\sqrt{\beta}\sum_{i=1}^{N}\sqrt{\lambda_i}\,|w_i|\right]\mathbf{I}_{\mathbf{R}_+^N}(\mathbf{w}), \quad (6)$$

where $\boldsymbol{\lambda}=[\lambda_1,\lambda_2,\ldots,\lambda_N]^T$ and $\boldsymbol{\Psi}=\mathrm{diag}(\boldsymbol{\lambda})$. The motivation to use a hyperparameter vector $\boldsymbol{\lambda}$ instead of a single λ for all γ_i's as in [9,8], is to form a hierarchical Bayesian analogue to the adaptive Lasso, proposed in [10]. Indeed, it can be shown, that the maximum a posteriori (MAP) estimator of $\mathbf{w}$, which is distributed according to (6), is the solution to the following optimization problem,

$$\tilde{\mathbf{w}}=\arg\min_{\mathbf{w}}\left\{\frac{\beta}{2}\left\|\mathbf{y}-\boldsymbol{\Phi}\mathbf{w}\right\|_2^2+\sum_{i=1}^{N}\alpha_i|w_i|\right\},\ \text{s.t. } \mathbf{w}\in\mathbf{R}_+^N, \quad (7)$$

which, excluding the non-negativity constraint, coincides with the definition of the adaptive Lasso, [10].

It is obvious from (6) that the quality of the endmember selection procedure depends on the tuning parameter vector $\boldsymbol{\lambda}$. We choose to infer the hyperparameter vector $\boldsymbol{\lambda}$ from the data, by assuming a Gamma hyperprior for each element of $\boldsymbol{\lambda}$,

$$p(\lambda_i|r,\delta)=\Gamma(\lambda_i|r,\delta)=\frac{\delta^r}{\Gamma(r)}{\lambda_i}^{r-1}\exp\left[-\delta\lambda_i\right],\ i=1,2,\ldots,N \quad (8)$$

where r and δ are hyperparameters, with $r\geq 0$ and $\delta\geq 0$. Both Gamma priors of β, in (4), and λ_i, in (8), are flexible enough to express prior information, by properly tuning their hyperparameters. In this paper, the hyperparameters κ,θ,r,δ are set to zero as in [2,8], forming non-informative (Jeffreys') priors, although other values can, in principle, be selected.

4. BAYESIAN INFERENCE

As it is common in Bayesian inference, the estimation procedure is based on the computation of the joint posterior distribution of the parameters. For the model presented in Section 3, this posterior is

$$p\left(\mathbf{w},\beta,\boldsymbol{\gamma},\boldsymbol{\lambda}|\mathbf{y}\right)=\frac{p\left(\mathbf{y}|\mathbf{w},\beta\right)p\left(\mathbf{w}|\beta,\boldsymbol{\gamma}\right)p\left(\boldsymbol{\gamma}|\boldsymbol{\lambda}\right)p\left(\boldsymbol{\lambda}\right)p\left(\beta\right)}{p(\mathbf{y})}, \quad (9)$$

which is intractable, because $p(\mathbf{y})$ cannot be computed analytically. To overcome this obstacle, a Markovian Gibbs sampling strategy can be followed, in which the conditional posterior distributions of the associated parameters are utilized.

4.1 Posterior conditional distributions

In the following, analytical expressions are derived for the posterior conditional distributions of the model parameters $\mathbf{w}$, $\boldsymbol{\gamma}$, $\boldsymbol{\lambda}$ and β. Starting with $\mathbf{w}$, it can be easily shown that its posterior conditional density is the multivariate Gaussian, truncated in $\mathbf{R}_+^N$,

$$p(\mathbf{w}|\mathbf{y},\boldsymbol{\gamma},\boldsymbol{\lambda},\beta)=\mathcal{N}(\mathbf{w}|\boldsymbol{\mu},\boldsymbol{\Sigma})\mathbf{I}_{\mathbf{R}_+^N}=\mathcal{N}_{\mathbf{R}_+^N}(\mathbf{w}|\boldsymbol{\mu},\boldsymbol{\Sigma}) \quad (10)$$

where $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$ are respectively expressed as follows,

$$\boldsymbol{\Sigma}=\beta^{-1}\left[\boldsymbol{\Phi}^T\boldsymbol{\Phi}+\boldsymbol{\Lambda}\right]^{-1},\qquad \boldsymbol{\mu}=\beta\boldsymbol{\Sigma}\boldsymbol{\Phi}^T\mathbf{y}. \quad (11)$$

The posterior conditional for the precision parameter β, is easily shown to be a Gamma distribution, i.e.,

$$p(\beta|\mathbf{y},\mathbf{w},\boldsymbol{\gamma},\boldsymbol{\lambda})=$$
$$\Gamma\left(\beta|\frac{M+N}{2}+\kappa,\frac{1}{2}\|\mathbf{y}-\boldsymbol{\Phi}\mathbf{w}\|_2^2+\theta+\frac{1}{2}\mathbf{w}^T\boldsymbol{\Lambda}\mathbf{w}\right) \quad (12)$$

Straightforward computations yield that the conditional distribution of γ_i given $\mathbf{y},w_i,\lambda_i,\beta$ is expressed as

$$p(\gamma_i|\mathbf{y},w_i,\lambda_i,\beta)=\left(\frac{\lambda_i}{2\pi}\right)^{\frac{1}{2}}\gamma_i^{-\frac{1}{2}}\exp\left[-\frac{\beta w_i^2}{2\gamma_i}-\frac{\lambda_i}{2}\gamma_i+\right.$$
$$\left.+\sqrt{\beta\lambda_i}\,|w_i|\right],\ i=1,2,\ldots,N \quad (13)$$

Finally, the conditional posterior of λ_i given $\mathbf{y},w_i,\gamma_i,\beta$ also turns out to be a Gamma distribution,

$$p(\lambda_i|\mathbf{y},w_i,\gamma_i,\beta)=\Gamma\left(\lambda_i|1+r,\frac{\gamma_i}{2}+\delta\right),\ i=1,2,\ldots,N. \quad (14)$$

By simple inspection, it can be verified that the conditional distribution in (13) is not easy to sample. To circumvent this, we propose a deterministic approximation of the Gibbs sampler, where the randomly generated samples of the Gibbs sampler are replaced by the *means* of the corresponding conditional distributions, (10), (12), (13) and (14), as explained in Section 5. Thus, a novel scheme iterating among the conditional means of $\mathbf{w}$, β, γ_i and λ_i arises, which will be termed *Bayesian inference iterative conditional expectations* (BI-ICE) algorithm. It should be emphasized that by following this approach, we depart from the statistical framework implied by the Gibbs sampler and we end up with a new deterministic algorithm for estimating the parameters of the proposed hierarchical model.

[1] $\mathbf{I}_{\mathbf{R}_+^N}(\mathbf{x})=1(0),\text{if } \mathbf{x}\in\mathbf{R}_+^N(\mathbf{x}\notin\mathbf{R}_+^N)$.

5. THE PROPOSED BI-ICE ALGORITHM

As mentioned previously, BI-ICE needs the conditional expectations of the model parameters. These are computed analytically as described below:

Expectation of $p(\mathbf{w}|\mathbf{y},\boldsymbol{\gamma},\boldsymbol{\lambda},\beta)$ $\mathbf{w}$: As shown in (10), $p(\mathbf{w}|\mathbf{y},\boldsymbol{\gamma},\boldsymbol{\lambda},\beta)$ is a multivariate Gaussian distribution, truncated in $\mathbf{R}_+^N$. In the one-dimensional case, the expectation of the truncated Gaussian distribution in $\mathbf{R}_+^1$ can be computed as

$$x \sim \mathcal{N}_{\mathbf{R}_+^1}(x|\mu^*,\sigma^*) \Rightarrow \mathrm{E}[x] = \mu^* + \frac{\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\frac{\mu^{*2}}{\sigma^{*2}}\right)}{1-\frac{1}{2}\mathrm{erfc}\left(\frac{\mu^*}{\sqrt{2}\sigma^*}\right)}\sigma^*, \tag{15}$$

where $\mathrm{erfc}(\cdot)$ is the complementary error function. Unfortunately, to the best of our knowledge, there is no analogous closed-form expression for the N-dimensional case. However, as shown in [11], the distribution of the ith element of $\mathbf{w}$ conditioned on the remaining elements $\mathbf{w}_{\neg i} = [w_1,\ldots,w_{i-1},w_{i+1},\ldots,w_N]^T$, can be expressed as

$$w_i|\mathbf{w}_{\neg i} \sim \mathcal{N}_{\mathbf{R}_+^1}(w_i|\mu_i^*,\sigma_{ii}^*) \tag{16}$$

$$\mu_i^* = \mu_i + \boldsymbol{\sigma}_{\neg i}^T\boldsymbol{\Sigma}_{\neg i\neg i}^{-1}(\mathbf{w}_{\neg i}-\boldsymbol{\mu}_{\neg i}) \tag{17}$$

$$\sigma_{ii}^* = \sigma_{ii} - \boldsymbol{\sigma}_{\neg i}^T\boldsymbol{\Sigma}_{\neg i\neg i}^{-1}\boldsymbol{\sigma}_{\neg i}, \tag{18}$$

where matrix $\boldsymbol{\Sigma}_{\neg i\neg i}$ is formed by removing the ith row and the ith column from $\boldsymbol{\Sigma}$, the $(N-1)\times 1$ vector $\boldsymbol{\sigma}_{\neg i}$ is the ith column of $\boldsymbol{\Sigma}$ after removing its ith element σ_{ii} and μ_i is the ith element of $\boldsymbol{\mu}$. Based on this result, an iterative procedure is proposed in order to compute the mean of the posterior $p(\mathbf{w}|\mathbf{y},\boldsymbol{\gamma},\boldsymbol{\lambda},\beta)$. Specifically, the j-th iteration, $j=1,2,\ldots$, of this procedure is described as follows[2]

$$\begin{aligned}
&1.\ w_1^{(j)} = \mathrm{E}[p(w_1|w_2^{(j-1)},w_3^{(j-1)},\ldots,w_N^{(j-1)})]\\
&2.\ w_2^{(j)} = \mathrm{E}[p(w_2|w_1^{(j)},w_3^{(j-1)},\ldots,w_N^{(j-1)})]\\
&\quad\vdots\\
&N.\ w_N^{(j)} = \mathrm{E}[p(w_N|w_1^{(j)},w_2^{(j)},\ldots,w_{N-1}^{(j)})]
\end{aligned} \tag{19}$$

This procedure is repeated iteratively until convergence. Experimental results have shown that the iterative scheme in (19) converges to the mean of $\mathbf{w} \sim \mathcal{N}_{\mathbf{R}_+^N}(\mathbf{w}|\boldsymbol{\mu},\boldsymbol{\Sigma})$ after a few iterations.

Expectation of $p(\beta|\mathbf{y},\mathbf{w},\boldsymbol{\gamma},\boldsymbol{\lambda})$: The mean value of the Gamma distribution of (12) is given by

$$\mathrm{E}[p(\beta|\mathbf{y},\mathbf{w},\boldsymbol{\gamma},\boldsymbol{\lambda})] = \frac{\frac{M+N}{2}+\kappa}{\frac{1}{2}\|\mathbf{y}-\boldsymbol{\Phi}\mathbf{w}\|_2^2+\theta+\frac{1}{2}\mathbf{w}^T\boldsymbol{\Lambda}\mathbf{w}} \tag{20}$$

Expectation of $p(\gamma_i|\mathbf{y},w_i,\lambda_i,\beta)$: It can be shown that the expectation of (13) is expressed as

$$\mathrm{E}[p(\gamma_i|\mathbf{y},w_i,\lambda_i,\beta)] = \left(\frac{2\lambda_i}{\pi}\right)^{\frac{1}{2}}\left(\frac{\beta w_i^2}{\lambda_i}\right)^{\frac{3}{4}}\exp\left[\sqrt{\beta\lambda_i}\,|w_i|\right] K_{3/2}\left(\sqrt{\beta\lambda_i}\,|w_i|\right), \tag{21}$$

where $K_\nu(\cdot)$ stands for the modified Bessel function of second kind of order ν.

Expectation of $p(\lambda_i|\mathbf{y},w_i,\gamma_i,\beta)$: Finally, the mean value of the Gamma distribution (14) is

$$\mathrm{E}[\lambda_i|\mathbf{y},w_i,\gamma_i,\beta)] = \frac{1+r}{\frac{1}{2}\gamma_i+\delta}. \tag{22}$$

[2]In the following, for notational simplicity, the expectation $E_{p(x|y)}[x]$ of a random variable x with conditional distribution $p(x|y)$ is denoted as $E[p(x|y)]$.

Table 1. The BI-ICE algorithm

Input $\boldsymbol{\Phi}$, $\mathbf{y}$, κ,θ,r,δ
Initialize $\boldsymbol{\gamma}^{(0)} = \boldsymbol{\lambda}^{(0)} = \mathbf{1}$, $\beta^{(0)} = 0.01\,\|\mathbf{y}\|_2$
for $t = 1,2,\ldots$ **do**
- Compute $\mathbf{w}^{(t)}$ as follows
  - Compute $\boldsymbol{\Sigma}^{(t)}, \boldsymbol{\mu}^{(t)}$ using (11)
  - Set $\mathbf{v}^{(0)} = \boldsymbol{\mu}^{(t)}$
  - Compute $v_1^{(1)} = \mathrm{E}\left[p(v_1|v_2^{(0)},\ldots,v_N^{(0)})\right]$, using (17), (18), and (15)
  - Compute $v_2^{(1)} = \mathrm{E}\left[p(v_2|v_1^{(1)},v_3^{(0)},\ldots,v_N^{(0)})\right]$, using (17), (18), and (15)
  - $\vdots$
  - Compute $v_N^{(1)} = \mathrm{E}\left[p(v_N|v_1^{(1)},v_2^{(1)},\ldots,v_{N-1}^{(1)})\right]$, using (17), (18), and (15)
  - Set $\mathbf{w}^{(t)} = \mathbf{v}^{(1)}$
- Compute $\beta^{(t)} = \mathrm{E}\left[p(\beta|\mathbf{y},\mathbf{w}^{(t)},\boldsymbol{\gamma}^{(t-1)},\boldsymbol{\lambda}^{(t-1)}\right]$, using (20)
- Compute $\gamma_i^{(t)} = \mathrm{E}\left[p(\gamma_i|\mathbf{y},w_i^{(t)},\lambda_i^{(t-1)},\beta^{(t)})\right]$, $i=1,2,\ldots,N$, using (21)
- Compute $\lambda_i^{(t)} = \mathrm{E}\left[p(\lambda_i|\mathbf{y},w_i^{(t)},\gamma_i^{(t)},\beta^{(t)})\right]$, $i=1,2,\ldots,N$, using (22)

endfor

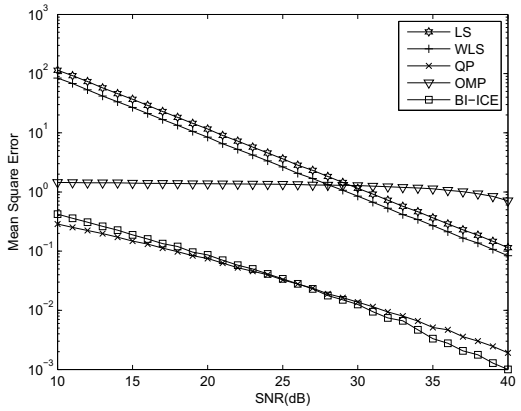
The basic steps of the proposed BI-ICE algorithm are summarized in Table 1. Regarding the updating of parameter $\mathbf{w}^{(t)}$, an auxiliary variable $\mathbf{v}$ has been utilized in Table 1. This is initialized with $\boldsymbol{\mu}^{(t)}$ (the value of $\boldsymbol{\mu}$ at iteration t) and is updated by performing a *single* iteration of the scheme described in (19). The resulting value of $\mathbf{v}$ is assigned to $\mathbf{w}^{(t)}$. The rationale behind this choice is that for $\boldsymbol{\Sigma}$ diagonal (which happens when the columns of $\boldsymbol{\Phi}$ are orthogonal), it easily follows from (17), (18) that the w_i's in (19) are uncorrelated. Thus, a single iteration is sufficient to obtain the mean of $\mathcal{N}_{\mathbf{R}_+^N}(\mathbf{w}|\boldsymbol{\mu},\boldsymbol{\Sigma})$. Although, this is not valid when $\boldsymbol{\Sigma}$ is not diagonal, experimental results have evidenced that the estimation of the mean of $\mathcal{N}_{\mathbf{R}_+^N}(\mathbf{w}|\boldsymbol{\mu},\boldsymbol{\Sigma})$ resulting after the execution of a single iteration of the scheme in (19) is also sufficient in the framework of the BI-ICE algorithm.

A basic advantage of the proposed Bayesian approach, which is the Bayesian analogue to the adaptive Lasso, is that all parameters are naturally estimated from the data. In contrast, deterministic algorithms for solving the Lasso, e.g. [10], face the problem of fine-tuning specific parameters (corresponding to $\boldsymbol{\lambda}$ of our model), that control the sparsity of the solution. As shown in the simulations presented in the next section, the BI-ICE algorithm converges very fast, and retains the sparsity of the solution. It has been further observed that by initializing $\mathbf{w}$ with $\boldsymbol{\mu}$, a single cycle is sufficient for the inner sampler to converge. The computational complexity of the proposed method can be further reduced by avoiding the explicit computation of the matrcies $\boldsymbol{\Sigma}_{\neg i\neg i}^{-1}$ in (17), (18). Due to space limitations, this issue will not be elaborated any further.

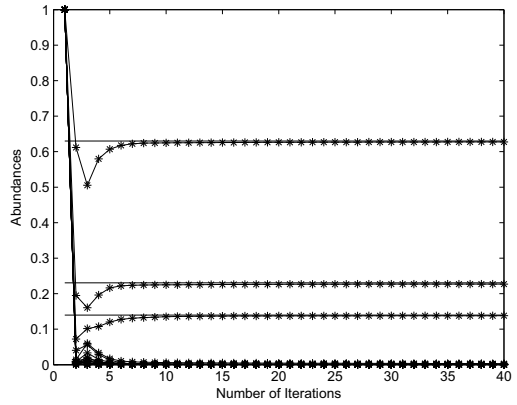
It is worth noting that the additivity constraint can be incorporated in the model deterministically, by augmenting the initial LMM of (1) with an extra equation, as in [3], i.e.,

$$\begin{bmatrix}\mathbf{y}\\ \alpha\end{bmatrix} = \begin{bmatrix}\boldsymbol{\Phi}\\ \alpha\mathbf{1}^T\end{bmatrix}\mathbf{w} + \begin{bmatrix}\mathbf{n}\\ 0\end{bmatrix} \tag{23}$$

where α is a scalar parameter, which controls the effect of the sum-to-one constraint on the estimation of $\mathbf{w}$. The larger the value of α is, the closer the sum of the estimated w_i's will be to one. Note that modifying the LMM as in (23) does not affect the proposed hierarchical Bayesian model and the subsequent analysis.



(a)



(b)

Fig. 1. (a) Mean square error of $\mathbf{w}$ for the simulated hyperspectral image. (b) Convergence of the abundance fractions for one pixel.

6. EXPERIMENTAL RESULTS AND DISCUSSION

6.1 Simulation results on synthetic data

The performance of the BI-ICE algorithm is illustrated by unmixing a synthetic hyperspectral image, using data from the USGS spectral library, [12]. Specifically, 30 endmembers were selected from the library, to construct a 470×30 endmember matrix, having condition number $1.73 \cdot 10^3$. The number of disparate endmembers composing a single pixel varied between one (pure pixel) and five, whereas the abundances were generated according to a Dirichlet distribution, so as to satisfy the positivity and sum-to-one constraints. The observations were corrupted by Gaussian noise, whose variance was determined by the SNR level. The BI-ICE algorithm was compared to (a) the least squares (LS) algorithm, (b) the weighted least squares (WLS) algorithm, where only the additivity constraint is imposed using (23), (c) the orthogonal matching pursuit (OMP) algorithm, and (d) a quadratic programming (QP) technique, which enforces the constraints, but does not exploit the problem's sparsity, [13]. For different levels of SNR, the corresponding MSE curves ($\mathrm{MSE} = \mathrm{E}[\frac{\|w-\tilde{w}\|_2^2}{\|w\|_2^2}]$) are shown in Fig. 1(a). It can be seen that the proposed algorithm outperforms the LS, WLS and OMP algorithms and has similar performance to the QP scheme. Due to the high conditioning of $\mathbf{\Phi}$, OMP fails, while for high SNRs the BI-ICE algorithm has the lowest MSE. It should be also noted that, compared to QP, the proposed BI-ICE algorithm produces estimates for all model parameters (e.g., the noise variance), and provides confidence intervals for these estimates. The fast convergence and sparsity promoting nature of the new algorithm are depicted in Fig 1(b). A pixel with three non-zero abundances (0.1397, 0.2305, 0.6298) is considered with $\mathrm{SNR} = 25\mathrm{dB}$. The curves in Fig. 1(b) are the average of 50 realizations, where not only the noise but also the positions of the non-zero abundances in $\mathbf{w}$ varies. We observe that less than 10 iterations are sufficient for the BI-ICE algorithm to converge to the correct values. Moreover, all remaining abundance fractions become zero after a few iterations.

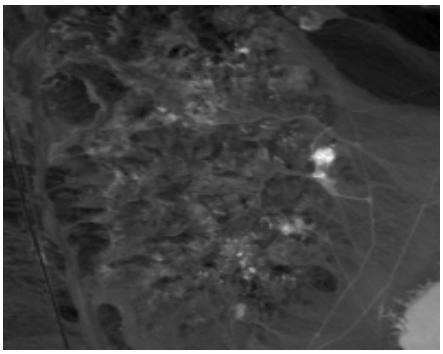
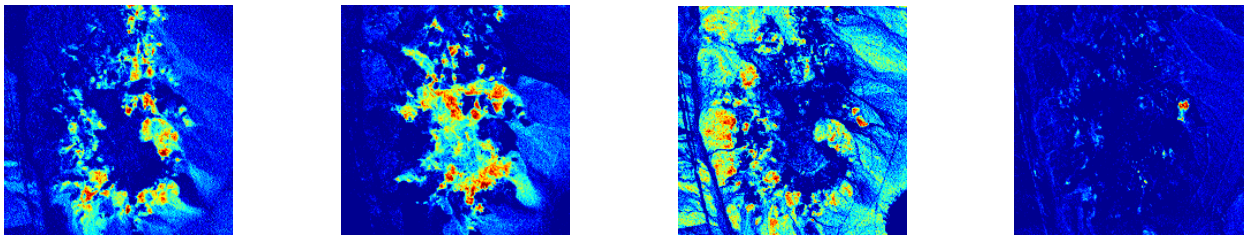


Fig. 2. Band 150 of a subimage of the Cuprite Aviris hyperspectral data set.

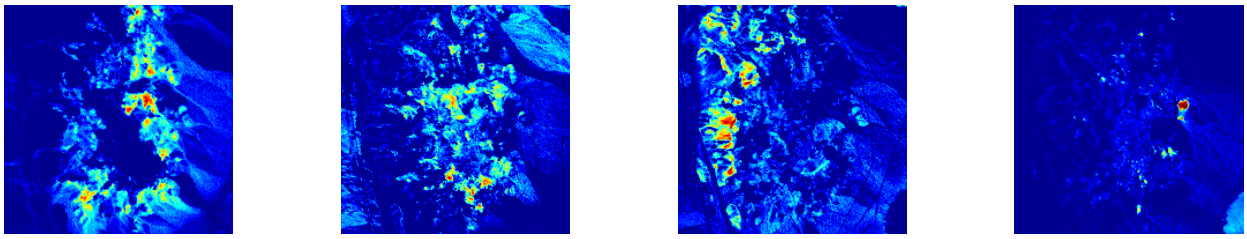
6.2 Simulation Results on Real Data

This section describes the application of the proposed BI-ICE algorithm to real hyperspectral image data. The real data were collected by the airborne visible/infrared imaging spectrometer (AVIRIS) flight over the Cuprite mining site, Nevada, in 1997, [14]. The AVIRIS sensor is a 224-channel imaging spectrometer with approximately 10-nm spectral resolution covering wavelengths ranging from 0.4 to 2.5 μm. The spatial resolution is 20 m. This data set has been widely used for remote sensing experiments [6,7]. The spectral bands 1-2, 104-113, 148-167, and 221-224 were removed due to low SNR and water-vapor absorption. Hence, a total of 188 bands were considered in this experiment. The subimage of the 150th band, including 200 vertical lines with 200 samples per line (200×200) is shown in Fig. 2.

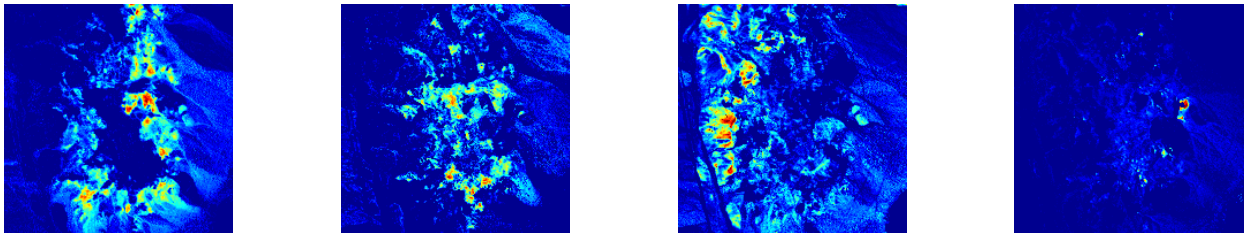
The VCA algorithm described in [6], was used to extract 14 endmembers present in the image. Using these spectral signatures, three algorithms are tested to estimate the abundances, namely the LS algorithm, the QP method, and the proposed BI-ICE algorithm. The unmixing process generates an output image for each endmember, depicting the endmember's estimated abundance fraction for each pixel. The darker the pixel, the smaller the contribution of this endmember in the pixel is. On the other hand, a light pixel indicates that the proportion of the endmember in the specific pixel is high. The abundance fractions of four endmembers, estimated using the LS, QP and BI-ICE algorithms, are shown in Fig. 3a, Fig. 3b, and Fig. 3c, respectively. Note that, for the sake of comparison, a necessary linear scaling in the range [0 1] has been performed for the LS abundance images. By simple inspection, it can be observed that the images taken using the LS algorithm clearly deviate from the images of the other two methods. The LS algorithm imposes no constraints on the estimated abundances, and hence the scaling has a major impact on the abundance fractions, resulting in performance degradation. On the contrary, the images obtained by QP and BI-ICE share a high degree of similarity and are in full agreement with previous results concerning the selected abundances and reported in [6,7], as well as with the conclusions derived in Section 6.1.



(a) LS algorithm



(b) QP algorithm



(c) BI-ICE algorithm

Fig. 3. Estimated abundance values of four endmembers using: (a) the LS algorithm, (b) the QP algorithm, (c) the proposed BI-ICE algorithm

REFERENCES

[1] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Trans. Signal Process.*, vol. 19, pp. 44–57, Jan. 2002.

[2] 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, no. 7, pp. 2684–2695, July 2008.

[3] K. Themelis, A.A. Rontogiannis, and K. Koutroumbas, "Semi-supervised hyperspectral unmixing via the weighted Lasso," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'10)*, Dallas, Texas, Mar. 2010.

[4] M-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Unmixing sparse hyperspectral mixtures," in *Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009*, Cape Town, July 2009, vol. 4, pp. 85 –88.

[5] J.M. Bioucas-Dias and M.A.T. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2010 2nd Workshop on*, Reykjavik, June 2010, pp. 1–4.

[6] 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, pp. 898–910, Apr. 2005.

[7] Tsung-Han Chan, Chong-Yung Chi, Yu-Min Huang, and Wing-Kin Ma, "A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4418 –4432, Nov. 2009.

[8] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Bayesian compressive sensing using Laplace priors," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 53 –63, Jan. 2010.

[9] T. Park and C. George, "The Bayesian Lasso," *Journal of the American Statistical Association*, vol. 103, no. 482, pp. 681–686, June 2008.

[10] H. Zou, "The adaptive Lasso and its oracle properties," *Journal of the American Statistical Association*, vol. 101, pp. 1418–1429, Dec. 2006.

[11] G. A. Rodriguez-Yam, R. A. Davis, and L. L. Scharf, "A bayesian model and Gibbs sampler for hyperspectral imaging," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop*, Aug. 2002, pp. 105–109.

[12] R. N. Clark, G. A. Swayze, R. Wise, K. E. Livo, T. M. Hoefen, R. F. Kokaly, and S. J. Sutley, "USGS digital spectral library," 2007, http://speclab.cr.usgs.gov/spectral.lib06/ds231/datatable.html.

[13] T. F. Coleman and Y. Li, "A reflective Newton method for minimizing a quadratic function subject to bounds on some of the variables," *SIAM Journal on Optimization*, vol. 6, pp. 1040–1058, 1996.

[14] "AVIRIS free standard data products," http://aviris.jpl.nasa.gov/html/aviris.freedata.html.