ROBUST LINEAR SPECTRAL UNMIXING USING OUTLIER DETECTION

Yoann Altmann[(1)]*, Steve McLaughlin*[(1)]†*, Alfred Hero*[(2)]‡

[(1)]Heriot-Watt University, School of Engineering and Physical Sciences, Edinburgh, United Kingdom.
{*Y.Altmann,S.McLaughlin*}*@hw.ac.uk*
[(2)]University of Michigan, Electrical Engineering and Computer Science, Ann Arbor, Michigan.
hero@eecs.umich.edu

ABSTRACT

This paper presents a Bayesian algorithm for linear spectral unmixing that accounts for outliers present in the data. The proposed model assumes that the pixel reflectances are linear mixtures of unknown endmembers, corrupted by an additional term modelling outliers and additive Gaussian noise. A Markov random field is considered for outlier detection based on the spatial and spectral structures of the anomalies. This allows outliers to be identified in particular regions and wavelengths of the data cube. A Bayesian algorithm is proposed to estimate the parameters involved in the model yielding a joint linear unmixing and outlier detection algorithm. Simulations conducted with synthetic data demonstrate the accuracy of the proposed unmixing and outlier detection strategy for the analysis of hyperspectral images.

Index Terms— Hyperspectral imagery, unsupervised spectral unmixing, Bayesian estimation, MCMC, nonlinearity detection.

1. INTRODUCTION

Spectral unmixing (SU) of hyperspectral images (HSI) has been the subject of intensive interest over the last two decades. It consists of distinguishing the materials and quantifying their proportions in each pixel of an observed image. The SU problem has been widely studied for applications where pixel reflectances are linear combinations of pure component spectra (called endmembers) [1, 2]. However, as explained in [2], the linear mixing model (LMM) can be inappropriate for some hyperspectral images, such as those containing sand-like materials or where relief is present in the scene. Moreover, LMM-based methods can also fail when the data are corrupted by (sparse) outliers, especially when extracting the endmembers from the scene. Nonlinear mixing models (NLMMs) provide an interesting alternative to overcoming the inherent limitations of the LMM. They have been proposed in the hyperspectral image literature and can be divided into two main classes [3]. The first class of NLMMs consists of physical models based on the nature of the environment (e.g., intimate mixtures [4] and multiple scattering effects [5, 6, 7]). The second class of NLMMs contains more flexible models allowing a wider range of nonlinearities to be approximated [8, 9].

In this work, we consider a general model for spectral unmixing which assumes that the observed pixels result from a convex combination of the endmembers of the scene, corrupted by an additive term modelling deviations from the classical LMM (e.g., outliers, nonlinear effects) and additive Gaussian Noise. The number of endmembers is assumed to be known whereas their spectral signatures are unknown. It is interesting to note that many nonlinear models in the literature, including polynomial models [5, 6, 7] can be expressed in a similar manner. In this paper, the additional terms are assumed to be *a priori* independent of the endmembers and/or their proportions (abundances), as in [10, 11]. This class of models for robust linear SU allows for general deviations from the LMM to be handled in blind source separation methods, i.e., nonlinear effects, outliers or possible endmember variability [12]. In [11], spatial and spectral sparsity structures haven been considered for the additional term since deviations from the LMM can occur in specific regions or spectral bands of the HSI. This is typically the case when outliers are present, but also when nonlinear effects occurs (relief) and when the reflectance of materials present has significant variations in particular spectral ranges (such those caused for instance by variations of water and chlorophyl contents in vegetation species). This paper extends the study presented in [11] which assumed a fixed partition of the data cube to perform outlier detection, by introducing spatial and spectral neighborhood relationships thus allowing for more flexible group-sparsity structures for the outliers.

In the Bayesian framework, prior distributions are assigned to the unknown model parameters to include available information (such as parameter constraints) within the estimation procedure. In particular, an Ising Markov random field is introduced to model spatial and spectral correlations for the outliers. The joint posterior distribution of the unknown parameter vector is then derived. Since classical Bayesian estimators cannot be easily computed from this joint posterior, a Markov chain Monte Carlo (MCMC) method is used to generate samples according to this posterior. Finally, the generated samples are used to compute Bayesian estimators.

The remaining sections of the paper are organized as follows. Section 2 introduces the mixing model for robust linear SU of HSIs, followed by Section 3 which summarizes the likelihood and the priors assigned to the unknown parameters/hyperparameters of the model. The resulting joint posterior distribution and the Gibbs sampler used to sample from it are summarized in Section 4. Some simulation results conducted on synthetic data are shown and discussed in Section 5. Conclusions and future work are finally reported in Section 6.

2. PROBLEM FORMULATION

We consider a set of N observed pixels/spectra $\mathbf{y}_n \in \mathbb{R}^L, n \in \{1, \ldots, N\}$ where L is the number of spectral bands. Each of these spectra is assumed to result from a linear combination of R unknown endmembers $\mathbf{m}_r$, corrupted by possible additive outliers and Gaus-

*Part of this work was supported by the Direction Générale de l'Armement, French Ministry of Defence.

†Part of this work was supported by the EPSRC via grant EP/J015180/1.

‡Part of this work was supported by US Army Research Office through Grant W911NF-11-1-0391.

sian noise. The observation model can be expressed as

$$\begin{aligned}\mathbf{y}_n &= \sum_{r=1}^{R} \mathbf{m}_r a_{r,n} + \mathbf{r}_n + \mathbf{e}_n \\ &= \mathbf{M}\boldsymbol{a}_n + \mathbf{r}_n + \mathbf{e}_n, \quad n = 1, \ldots, N \end{aligned} \tag{1}$$

where $\mathbf{m}_r = [m_{r,1}, \ldots, m_{r,L}]^T$ is the spectrum of the rth material present in the scene and $a_{r,n}$ is its corresponding proportion (abundance) in the nth pixel. In (1), $\mathbf{e}_n$ is an additive independently but non identically distributed zero-mean Gaussian noise sequence with diagonal covariance matrix $\boldsymbol{\Sigma}_0 = \text{diag}(\boldsymbol{\sigma}^2)$, denoted as $\mathbf{e}_n \sim \mathcal{N}(\mathbf{e}_n; \mathbf{0}_L, \boldsymbol{\Sigma}_0)$, where $\boldsymbol{\sigma}^2 = [\sigma_1^2, \ldots, \sigma_L^2]^T$ is the vector of the L noise variances and $\text{diag}(\boldsymbol{\sigma}^2)$ is an $L \times L$ diagonal matrix containing the elements of the vector $\boldsymbol{\sigma}^2$. Moreover, $\mathbf{r}_n$ denotes the outlier vector of the nth pixel. Note that the usual matrix and vector notations $\mathbf{M} = [\mathbf{m}_1, \ldots, \mathbf{m}_R]$ and $\boldsymbol{a}_n = [a_{1,n}, \ldots, a_{R,n}]^T$ have been used in the second row of (1).

As consequence of physical constraints, the abundance vectors $\boldsymbol{a}_n$ satisfy the following positivity and sum-to-one constraints

$$\sum_{r=1}^{R} a_{r,n} = 1, \;\; a_{r,n} > 0, \forall r \in \{1, \ldots, R\}. \tag{2}$$

The problem investigated in this paper is to estimate the endmember matrix $\mathbf{M}$, the abundance matrix $\mathbf{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N]$, the noise variances in $\boldsymbol{\sigma}^2$ and the outlier matrix $\mathbf{R} = [\mathbf{r}_1 \ldots, \mathbf{r}_N]$ from the observation matrix $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N]$. This problem is similar to the robust principal component analysis problem and a Bayesian approach similar to that proposed in [13] is investigated here to estimate the unknown parameters.

3. BAYESIAN MODEL

3.1. Likelihood

Eq. (1) shows that $\mathbf{y}_n|\mathbf{M}, \boldsymbol{a}_n, \mathbf{r}_n, \boldsymbol{\sigma}^2 \sim \mathcal{N}(\mathbf{y}_n; \mathbf{M}\boldsymbol{a}_n + \mathbf{r}_n, \boldsymbol{\Sigma}_0)$. Assuming independence between noise sequences of the N observed pixels, the likelihood of the observation matrix $\mathbf{Y}$ can be expressed as

$$f(\mathbf{Y}|\mathbf{M}, \mathbf{A}, \mathbf{R}, \boldsymbol{\sigma}^2) \propto$$

$$|\boldsymbol{\Sigma}_0|^{-N/2} \text{etr}\left[-\frac{(\mathbf{Y} - \mathbf{M}\mathbf{A} - \mathbf{R})^T \boldsymbol{\Sigma}_0^{-1} (\mathbf{Y} - \mathbf{M}\mathbf{A} - \mathbf{R})}{2}\right] \tag{3}$$

where $\propto$ means "proportional to" and $\text{etr}(\cdot)$ denotes the exponential trace.

3.2. Parameter priors

3.2.1. Prior for the abundance matrix $\mathbf{A}$

Each abundance vector can be written as $\boldsymbol{a}_n = [\mathbf{c}_n^T, a_{R,n}]^T$ with $\mathbf{c}_n = [a_{1,n}, \ldots, a_{R-1,n}]^T$ and $a_{R,n} = 1 - \sum_{r=1}^{R-1} a_{r,n}$. The LMM constraints (2) impose that $\mathbf{c}_n$ belongs to the simplex $\mathcal{S} = \left\{\mathbf{c} \left| c_r \geq 0, \forall r \in 1, \ldots, R-1, \sum_{r=1}^{R-1} c_r \leq 1 \right.\right\}$. To reflect the lack of prior knowledge about the abundances, a uniform prior is assigned for each vector $\mathbf{c}_n, n \in \{1, \ldots, N\}$, i.e., $f(\mathbf{c}_n) \propto \mathbf{1}_{\mathcal{S}}(\mathbf{c}_n)$, where $\mathbf{1}_{\mathcal{S}}(\cdot)$ is the indicator function defined on the simplex $\mathcal{S}$. Assuming prior independence between the N abundance vectors $\{\boldsymbol{a}_n\}_{n=1,\ldots,N}$ leads to the following joint prior distribution

$$f(\mathbf{C}) = \prod_{n=1}^{N} f(\mathbf{c}_n), \tag{4}$$

where $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_N]$ is an $(R-1) \times N$ matrix.

3.2.2. Prior for the endmember matrix $\mathbf{M}$

To reflect the lack of prior knowledge about the endmembers, we use the following multivariate truncated Gaussian prior

$$f(\mathbf{M}) \propto \prod_{r=1}^{R} \mathcal{N}_{\left(\mathbb{R}^+\right)^L}(\mathbf{m}_r; \mathbf{0}, \gamma^{-1}\mathbf{I}_L) \tag{5}$$

where γ is fixed to a small value, to ensure endmember positivity while using a weakly informative prior. Note that (5) is considered in order to handle the case where the data are not normalized. If the data are actual reflectance values, a prior ensuring that the endmember spectra belong to $(0; 1)$, such as a uniform, beta [14] or Gaussian distribution. Note also that the prior can also include prior information from an endmember extraction algorithm, as in [15, 16].

3.2.3. Prior for the noise variances

A Jeffreys' prior is chosen for the noise variance of each spectral band σ_ℓ^2, i.e., $f(\sigma_\ell^2) \propto \sigma_\ell^{-2}\mathbf{1}_{\mathbb{R}^+}\left(\sigma_\ell^2\right)$ where $\mathbf{1}_{\mathbb{R}^+}(\cdot)$ denotes the indicator function defined on $\mathbb{R}^+$, which reflects the absence of knowledge about these parameters. Assuming prior independence between the noise variances, we obtain $f(\boldsymbol{\sigma}^2) = \prod_{\ell=1}^{L} f(\sigma_\ell^2)$.

3.2.4. Priors of the outliers

As in [13, 11, 9], the outliers are assumed to be sparse, i.e., for most of the pixels and spectral bands, the outliers are expected to be exactly equal to zero. To model the outlier sparsity, we factorize the outlier matrix as

$$\mathbf{R} = \mathbf{Z} \odot \mathbf{X}, \tag{6}$$

where $\mathbf{Z} \in \{0, 1\}^{L \times N}$ is a label matrix, $\mathbf{X} \in \mathbb{R}^{L \times N}$ and $\odot$ denotes the Hadamard (termwise) product. This decomposition allows one to decouple the location of the sparse components from their values. More precisely, $z_{\ell,n} = [\mathbf{Z}]_{\ell,n} = 1$ if an outlier is present in the ℓth spectral band of the nth observed pixel with value equal to $r_{\ell,n} = x_{\ell,n}$. A conjugate Gaussian prior is used for $\mathbf{X}$, i.e.,

$$f(\mathbf{X}|s^2) = \prod_{\ell,n} \mathcal{N}\left(x_{\ell,n}; 0, s^2\right), \tag{7}$$

where s^2 controls the prior energy of the outliers. Note that (7) allows the outliers to be negative. Other conjugate priors, such as truncated Gaussian priors, could be used instead of (7), e.g., to enforce outlier positivity. The next section presents the prior considered for the label matrix $\mathbf{Z}$.

3.2.5. Label matrix

For many applications, the locations of outliers are likely to be spectrally (e.g., water absorption bands, material spectral variations) and/or spatially (weakly represented components, shadowing effects,...) correlated. An interesting way to take possibly correlated outliers/nonlinear effects into account is to consider Markov random fields (MRF) to build a prior for the label matrix $\mathbf{Z}$ [9]. MRFs assume that the distribution of a label $z_{\ell,n}$ conditionally to the other labels of the image equals the distribution of this label vector conditionally to its neighbors, i.e., $\text{P}(z_{\ell,n}|\mathbf{Z}_{\setminus z_{\ell,n}}) = \text{P}(z_{\ell,n}|\mathbf{Z}_{\mathcal{V}_{\ell,n}})$, where $\mathcal{V}_{\ell,n}$ is the index set of the neighbors of $z_{\ell,n}$, $\mathbf{Z}_{\setminus z_{\ell,n}}$ denotes

the matrix $\mathbf{Z}$ whose element $z_{\ell,n}$ has been removed and $\mathbf{Z}_{\mathcal{V}_{\ell,n}}$ is the subset of $\mathbf{Z}$ composed of the elements whose indexes belong to $\mathcal{V}_{\ell,n}$. In this study, we consider that the spatial and spectral correlations can be different and thus consider two different neighborhoods. We decompose the neighborhood $\mathcal{V}_{\ell,n}$ as $\mathcal{V}_{\ell,n} = \mathcal{V}^L_{\ell,n} \cup \mathcal{V}^N_{\ell,n}$ where $\mathcal{V}^N_{\ell,n}$ (resp. $\mathcal{V}^L_{\ell,n}$) denotes the spatial (resp. spectral) neighborhood of $z_{\ell,n}$. The proposed MRF can be expressed as

$$\mathrm{P}(\mathbf{Z}|\boldsymbol{\beta}') = \frac{1}{B(\boldsymbol{\beta}')} \exp\left[\boldsymbol{\beta}^T \boldsymbol{\phi}(\mathbf{Z}) + \phi_0(\mathbf{Z}, \beta_0)\right] \quad (8)$$

where $\boldsymbol{\beta} = [\beta_N, \beta_L]^T$, $\boldsymbol{\beta}' = [\boldsymbol{\beta}^T, \beta_0]^T$ and

$$\begin{cases} \phi_L(\mathbf{Z}) &= \sum_{n,\ell} \sum_{z_{\ell',n} \in \mathcal{V}^L_{\ell,n}} \delta(z_{\ell,n} - z_{\ell',n}), \\ \phi_N(\mathbf{Z}) &= \sum_{n,\ell} \sum_{z_{\ell,n'} \in \mathcal{V}^N_{\ell,n}} \delta(z_{\ell,n} - z_{\ell,n'}), \\ \boldsymbol{\phi}(\mathbf{Z}) &= [\phi_L(\mathbf{Z}), \phi_N(\mathbf{Z})]^T, \\ \phi_0(\mathbf{Z}, \beta_0) &= \beta_0 \sum_{n,\ell}(1 - z_{\ell,n}) + (1 - \beta_0) \sum_{n,\ell} z_{\ell,n}, \end{cases}$$

and $\delta(\cdot)$ denotes the Kronecker delta function. Moreover, $\beta_N > 0$ and $\beta_L > 0$ are hyperparameters that control the spatial and spectral granularity of the MRF and $0 \leq \beta_0 \leq 1$ is an additional parameter that models the probability of having outliers in the image. Precisely, the higher β_0, the lower the probability of outliers in the data. The estimation of the proposed Ising model hyperparameters will be discussed in the next section.

3.3. Hyperparameter s^2

The following weakly informative inverse-Gamma prior is assigned to s^2

$$s^2 \sim \mathcal{IG}(\gamma, \nu), \quad (9)$$

where (γ, ν) are fixed to $(\gamma, \nu) = (10^{-3}, 10^{-3})$. The next section derives the joint posterior distribution of the unknown parameters associated with (1) and studies an MCMC methods to sample from this posterior.

4. SAMPLING STRATEGY

4.1. Joint posterior distribution

Assuming the parameters $\mathbf{M}, \mathbf{A}, \mathbf{Z}, \mathbf{X}$ and $\boldsymbol{\sigma}^2$ are a priori independent, the joint posterior of the parameter vector $\boldsymbol{\theta} = \left\{\mathbf{M}, \mathbf{A}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\sigma}^2\right\}$ and hyperparameter s^2 can be expressed as

$$f(\boldsymbol{\theta}, s^2|\mathbf{Y}, \boldsymbol{\beta}) \propto f(\mathbf{Y}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|s^2, \boldsymbol{\beta}) f(s^2) \quad (10)$$

where

$$f(\boldsymbol{\theta}|s^2, \boldsymbol{\beta}) = f(\mathbf{M}) f(\mathbf{A}) f(\boldsymbol{\sigma}^2) f(\mathbf{X}|s^2) \mathrm{P}(\mathbf{Z}|\boldsymbol{\beta}). \quad (11)$$

The next paragraph presents a sampling strategy to estimate the unknown parameter vector $\boldsymbol{\theta}$ and the hyperparameter s^2.

4.2. Gibbs sampler

To overcome the challenging derivation of the Bayesian estimators associated with $f(\boldsymbol{\theta}, s^2|\mathbf{Y}, \boldsymbol{\beta})$, we propose to use an efficient Markov Chain Monte Carlo (MCMC) method to generate samples asymptotically distributed according to $f(\boldsymbol{\theta}, s^2|\mathbf{Y}, \boldsymbol{\beta})$. More precisely, we consider a hybrid Gibbs sampler described in the next part of this section. The principle of the Gibbs sampler is to sample according to the conditional distributions of the posterior of interest [17, Chap. 10]. In this paper, we propose to sample sequentially the NL labels in $\mathbf{Z}$, the endmember matrix $\mathbf{M}$, the abundance matrix $\mathbf{A}$, the latent variables in $\mathbf{X}$, the noise variances $\boldsymbol{\sigma}^2$ and s^2 using moves that are summarized below.

Labels: Sampling $z_{\ell,n}$ from its conditional distribution can be achieved by drawing in $\{0, 1\}$ with known probabilities.

Endmembers: It can be easily shown that the rows of $\mathbf{M}$ are *a posteriori* independent (conditionally to the other parameters). Moreover, the conditional distribution of $\mathbf{m}_{\ell,:}|\mathbf{Y}, \boldsymbol{\theta}_{\setminus \mathbf{m}_{\ell,:}}, s^2$ is a multivariate Gaussian distribution restricted to $\left(\mathbb{R}^+\right)^R$, which can be sampled efficiently using the method recently proposed in [18]. Note that $\mathbf{m}_{\ell,:}$ denotes the ℓth row of $\mathbf{M}$ and $\boldsymbol{\theta}_{\setminus \mathbf{m}_{\ell,:}}$ contains all the elements of $\boldsymbol{\theta}$ but $\mathbf{m}_{\ell,:}$.

Abundances: In a similar fashion, it can be easily shown that the columns of $\mathbf{A}$ are *a posteriori* independent (conditionally to the other parameters). Moreover, the conditional distribution of $\mathbf{c}_n|\mathbf{y}_n, \boldsymbol{\theta}_{\setminus \mathbf{c}_n}, s^2$ is a multivariate Gaussian distribution restricted to the simplex $\mathcal{S}$, which can be sampled efficiently using the method proposed in [18].

Latent variable matrix $\mathbf{X}$: Similarly to the abundance matrix, the columns of $\mathbf{X}$ are *a posteriori* independent (conditionally to the other parameters) and can be sampled independently. Sampling each column of $\mathbf{X}$ reduces to sampling from a multivariate Gaussian distribution (since a Gaussian conjugate prior as been assigned in (7)).

Noise variances σ^2_ℓ: Sampling the noise variances can be easily achieved by sampling from L independent inverse-Gamma distributions.

Hyperparameter s^2: Similarly to the noise variances, it can be shown that $f(s^2|\mathbf{y}_n, \boldsymbol{\theta})$ is an inverse-Gamma distribution from which it is easy to sample.

After N_{MC} iterations (including N_{bi} burn-in samples that are discarded), the label matrix is estimated using marginal maximum a posteriori (MAP) estimation. This estimator is then used to compute the minimum mean square error (MMSE) of $\mathbf{R}$ conditioned upon $\mathbf{Z} = \widehat{\mathbf{Z}}_{\mathrm{MAP}}$ using

$$\hat{\mathbf{R}} = \left(\hat{\mathbf{R}}_{\mathrm{MMSE}}|\widehat{\mathbf{Z}}_{\mathrm{MAP}}\right) \odot \widehat{\mathbf{Z}}_{\mathrm{MAP}} \quad (12)$$

Finally, the remaining parameters are estimated using the empirical averages of the generated samples (MMSE estimates). It is interesting to note that thanks to the conjugate Gaussian prior (7), the matrix $\mathbf{X}$ could have been marginalized. However, this marginalization would lead to non-standard conditional distributions for $\boldsymbol{\sigma}^2$ and s^2 and accept/reject procedures would have to be used to update these variables. In this paper, we propose not to marginalize $\mathbf{X}$ and estimate this matrix, as it would be achieved when considering more complex prior than (7) (e.g., to handle outlier positivity).

Due to space constrains, we assume here that the hyperparameter vector $\boldsymbol{\beta}$ of the Ising MRF in (8) is known. However, it is well known that fixing these parameters to arbitrary values can have a significant impact on the detection performance. Consequently, we propose to estimate $\boldsymbol{\beta}$ before applying the proposed algorithm, using an adaptive MCMC method investigated in [19] which provides a maximum marginal likelihood estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ and which can then be used in (11) instead of $\boldsymbol{\beta}$. The interested reader is invited to consult [19] for further details of the estimation of $\boldsymbol{\beta}$. In the remainder of the paper, the estimator $\hat{\boldsymbol{\beta}}$ (which is close to $\boldsymbol{\beta}$ in practice) is used in (11) instead of $\boldsymbol{\beta}$.

5. EXPERIMENTS

The performance of the proposed method, referred to as "RBLU" (Robust Bayesian linear unmixing), is investigated synthetic data.

Additional simulations, conducted on real data and not presented here due to paper length constraints, are discussed in [20]. In this section, we consider two synthetic 60×60 pixels hyperspectral image composed of $R = 3$ endmembers and observed at $L = 207$ spectral bands (see Fig. 1). The abundances of the two images are uniformly distributed in the simplex defined by the positivity and sum-to-one constraints and the noise variances have been set to $\sigma_\ell^2 = 10^{-4}, \forall \ell$, which corresponds to an average signal-to-noise ratio (SNR) of 30dB. The first image I_1 does not contain outliers whereas the hyperparameter s^2 controlling the outlier power has been set to $s^2 = 0.1$ for the second image I_2. The label matrix of I_2 has been generated using (8) with $\boldsymbol{\beta} = [0.4; 0.2; 0.501]^T$ which leads to approximately 10% of actual outliers in $\mathbf{R}$. The proposed method has been applied to the images with $N_{\text{MC}} = 600$ iterations (including $N_{\text{bi}} = 300$). The endmember matrix has been initialized using the VCA algorithm [21] and the abundance matrix using FCLS [1]. The combination VCA-FCLS is also used as a state-of-the-art method for performance comparison.

The quality of the unmixing procedures can be measured by comparing the estimated and actual abundance vector using the root normalized mean square error (RNMSE) defined by

$$\text{RNMSE} = \sqrt{\frac{1}{NR} \sum_{n=1}^{N} \|\hat{\boldsymbol{a}}_n - \boldsymbol{a}_n\|^2} \tag{13}$$

where $\boldsymbol{a}_n$ and $\hat{\boldsymbol{a}}_n$ are the actual and estimated abundance vectors for the nth pixel of the image. The quality of endmember estimation is evaluated by the spectral angle mapper (SAM) defined as

$$\text{SAM} = \arccos\left(\frac{\langle \hat{\mathbf{m}}_r, \mathbf{m}_r \rangle}{\|\hat{\mathbf{m}}_r\| \, \|\mathbf{m}_r\|}\right) \tag{14}$$

where $\mathbf{m}_r$ is the rth actual endmember and $\hat{\mathbf{m}}_r$ its estimate. The smaller $|\text{SAM}|$, the closer the estimated endmembers to their actual values.

Table 1 compares the performance of the proposed method and the VCA-FCLS unmixing strategy and shows that the proposed methods outperforms VCA-FCLS in terms of abundance and endmember estimation. Moreover, the confusion matrix of the proposed outlier detection method in Table 2illustrates the ability of the method to identify the corrupted data.

			RBLU	VCA-FCLS	o-FCLS
I_1	SAM ($\times 10^{-2}$)	$\mathbf{m}_1$	0.21	0.68	-
		$\mathbf{m}_2$	0.17	0.92	-
		$\mathbf{m}_3$	0.26	1.96	-
	RNMSE ($\times 10^{-2}$)		0.68	1.60	0.67
I_2	SAM ($\times 10^{-2}$)	$\mathbf{m}_1$	0.21	4.03	-
		$\mathbf{m}_2$	0.17	3.08	-
		$\mathbf{m}_3$	0.55	4.26	-
	RNMSE ($\times 10^{-2}$)		0.75	6.63	5.59

Table 1. Estimation performance.

	$z = 0$	$z = 1$	Total
$\hat{z} = 0$	682187	3917	686104
$\hat{z} = 1$	329	58767	59096
Total	682516	62684	745200

Table 2. Outlier detection (I_2): confusion matrix.

Finally, Table 3 compares the abundance estimation performance of RBLU to o-FCLS (which assumes perfectly known endmembers) for different outlier corruption scenarios (proportions and variances). This table shows a general performance degradation of the algorithms when the number of outliers increases. However, although RBLU also estimates the endmembers (jointly with the abundances), the performance degradation is less significant for RBLU than for o-FCLS thanks to its outlier detection ability. It is interesting to note that RBLU is also less sensitive than o-FCLS to the outlier variance (o-FCLS abundance estimation performance decreases when the outlier variance increases).

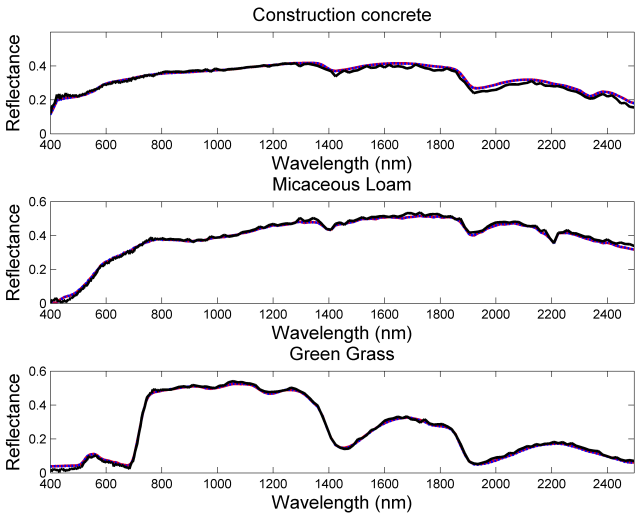

Fig. 1. Actual endmembers (red lines) used to generate the synthetic images and endmembers estimated by VCA (black lines) and RBLU (dashed blue lines) for I_2.

		$s^2 = 0.01$		$s^2 = 0.1$	
		RBLU	o-FCLS	RBLU	o-FCLS
Outlier prop.	10%	0.77	2.00	0.75	5.59
	20%	0.80	3.06	0.77	8.89
	30%	0.86	3.66	0.87	10.34

Table 3. Abundance RNMSE ($\times 10^{-2}$) for different outlier energies and proportions.

6. CONCLUSION

In this paper, we have investigated a Bayesian algorithm for robust linear spectral unmixing of hyperspectral images allowing joint endmember and abundance estimation and outlier detection. Appropriate prior distributions were used to enforce the endmember and abundance positivity and the abundance sum-to-one constraints. Moreover, an Ising Markov random field was used to model outliers spatial and spectral correlations. Finally, a Gibbs sampler was proposed to sample from the resulting posterior distribution. Simulations conducted on synthetic data showed the performance of the proposed method for linear SU and the detection of outliers in hyperspectral images. The proposed method has been applied to real hyperspectral images (the results are not presented due to paper length constraints) and has provided interesting results in terms of outlier analysis (see [20]). In this paper, the proposed outlier model was investigated for linear blind source separation. It would be interesting to extend this work to nonlinear models and to robust subspace identification problems, as in [11].

7. REFERENCES

[1] 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. 29, no. 3, pp. 529–545, March 2001.

[2] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observations Remote Sensing*, vol. 5, no. 2, pp. 354–379, April 2012.

[3] N. Dobigeon, J.-Y. Tourneret, C. Richard, J. C. M. Bermudez, S. McLaughlin, and A. O. Hero, "Nonlinear unmixing of hyperspectral images: Models and algorithms," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 82–94, Jan. 2014.

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

[5] B. Somers, K. Cools, S. Delalieux, J. Stuckens, D. V. der Zande, W. W. Verstraeten, and P. Coppin, "Nonlinear hyperspectral mixture analysis for tree cover estimates in orchards," *Remote Sensing of Environment*, vol. 113, no. 6, pp. 1183–1193, 2009.

[6] J. M. P. Nascimento and J. M. Bioucas-Dias, "Nonlinear mixture model for hyperspectral unmixing," in *Proc. SPIE Image and Signal Processing for Remote Sensing XV*, L. Bruzzone, C. Notarnicola, and F. Posa, Eds., vol. 7477. SPIE, 2009, p. 74770I.

[7] A. Halimi, Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Nonlinear unmixing of hyperspectral images using a generalized bilinear model," *IEEE Trans. Geosci. and Remote Sensing*, vol. 49, no. 11, pp. 4153–4162, Nov. 2011.

[8] J. Chen, C. Richard, and P. Honeine, "Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 480–492, 2013.

[9] Y. Altmann, N. Dobigeon, S. McLaughlin, and J.-Y. Tourneret, "Residual component analysis of hyperspectral images - application to joint nonlinear unmixing and nonlinearity detection," *IEEE Trans. Image Processing*, vol. 23, no. 5, pp. 2148–2158, May 2014.

[10] N. Dobigeon and C. Févotte, "Robust nonnegative matrix factorization for nonlinear unmixing of hyperspectral images," in *Proc. IEEE GRSS Workshop on Hyperspectral Image and SIgnal Processing: Evolution in Remote Sensing (WHISPERS)*, Gainesville, FL, June 2013.

[11] G. E. Newstadt, A. H. III, and J. Simmons, "Robust spectral unmixing for anomaly detection," in *Proc. IEEE-SP Workshop Stat. and Signal Processing*, Gold Coast, Australia, July 2014.

[12] A. Zare and K. Ho, "Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 95–104, Jan 2014.

[13] X. Ding, L. He, and L. Carin, "Bayesian robust principal component analysis," *IEEE Trans. Image Processing*, vol. 20, no. 12, pp. 3419–3430, Dec 2011.

[14] X. Du, A. Zare, P. Gader, and D. Dranishnikov, "Spatial and spectral unmixing using the beta compositional model," *IEEE J. Sel. Topics Appl. Earth Observations Remote Sensing*, vol. 7, no. 6, pp. 1994–2003, June 2014.

[15] N. Dobigeon, S. Moussaoui, M. Coulon, J.-. Tourneret, and A. O. Hero, "Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 2657–2669, Nov. 2009.

[16] Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Unsupervised post-nonlinear unmixing of hyperspectral images using a Hamiltonian Monte Carlo algorithm," *IEEE Trans. Image Processing*, vol. 23, no. 6, pp. 2663–2675, June 2014.

[17] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. New York: Springer-Verlag, 2004.

[18] A. Pakman and L. Paninski, "Exact Hamiltonian Monte Carlo for Truncated Multivariate Gaussians," *ArXiv e-prints*, Aug. 2012.

[19] M. Pereyra, N. Whiteley, C. Andrieu, and J.-Y. Tourneret, "Maximum marginal likelihood estimation of the granularity coefficient of a Potts-Markov random field within an mcmc algorithm," in *Proc. IEEE-SP Workshop Stat. and Signal Processing*, Gold Coast, Australia, July 2014.

[20] Y. Altmann, S. McLaughlin, and A. Hero, "Robust linear spectral unmixing using anomaly detection," *ArXiv e-prints*, Jan. 2015.

[21] J. M. Nascimento and J. M. Bioucas-Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. and Remote Sensing*, vol. 43, no. 4, pp. 898–910, April 2005.