Robust Linear Spectral Unmixing using Anomaly Detection

Yoann Altmann(1), Steve McLaughlin(1), Alfred Hero(2)

Abstract

This paper presents a Bayesian algorithm for linear spectral unmixing of hyperspectral images that accounts for anomalies present in the data. The model proposed assumes that the pixel reflectances are linear mixtures of unknown endmembers, corrupted by an additional nonlinear term modelling anomalies and additive Gaussian noise. A Markov random field is used for anomaly 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 anomaly detection algorithm. Simulations conducted with synthetic and real hyperspectral images 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, anomaly detection.

I. 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], [4]. The first class of NLMMs consists of physical models based on the nature of the environment (e.g., intimate mixtures [5] and multiple scattering effects [6]–[8]). The second class of NLMMs contains more flexible models allowing a wider range of nonlinearities to be approximated [9], [10].

In this work, we consider a general mixing 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, variability, 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 [6]–[8] 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 [11], [12]. 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 [13]. In [12], spatial and spectral sparsity structures were 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 (relief) occurs and when the reflectance of materials present has significant variations in particular spectral ranges (e.g. due to natural variability of vegetation). This paper extends the study presented in [12] which assumed the support sets of outliers to have fixed structure, by introducing a probabilistic 3D Ising model for spatial and spectral influence of outliers thus allowing for more flexible group-sparsity structures for the support sets of anomalies.

Adopting a Bayesian framework, we assign prior distributions to the unknown model parameters to include available information (such as parameter constraints) within the es-
timation procedure. In particular, an Ising Markov random field is introduced to model spatial and spectral correlations for the anomalies. 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. More precisely, we construct an efficient stochastic gradient MCMC (SGMCMC) algorithm \[14\] that simultaneously estimates the endmember and abundance matrices along with the Ising hyperparameters.

The main contributions of this work are threefold:

1) We develop a new hierarchical outlier model taking into account spatial and spectral correlations through Markovian dependencies. This flexible model is embedded within the mixing model for robust unsupervised linear spectral unmixing via anomaly detection.

2) An adaptive Markov chain Monte Carlo algorithm is then proposed to compute the Bayesian estimates of interest and perform Bayesian inference. This algorithm is equipped with a stochastic optimisation adaptation mechanism that automatically adjusts the parameters of the Markov random field by maximum marginal likelihood estimation, thus removing the need to set the regularisation parameters by cross-validation.

3) We show the benefits of the proposed flexible model for linear spectral unmixing of synthetic and real hyperspectral images. Specifically, we demonstrate the ability of the proposed algorithm to detect structured anomalies thus enhancing endmember and abundance estimation. The simulation results obtained using real data also illustrate how the Bayesian method can be used for anomaly analysis.

The remaining sections of the paper are organized as follows. Section II introduces the mixing model for robust linear SU of HSIs, followed by Section III which summarizes the likelihood and the priors assigned to the unknown parameters of the model. The resulting joint posterior distribution and the Gibbs sampler used to sample from it are summarized in Section IV. A generalization of the proposed Bayesian model for robust Bayesian subspace identification is proposed in Section V. Some simulation results conducted on synthetic data are shown and discussed in Section VI. Conclusions and future work are reported in Section VIII.
II. PROBLEM FORMULATION

We consider a set of \( N \) observed pixels/spectra \( y_n = [y_{1,n}, \ldots, y_{L,n}]^T, n \in \{1, \ldots, N\} \) where \( L \) is the number of spectral bands. Each of these spectra is assumed to result from a linear combination of \( R \) unknown endmembers \( m_r \), corrupted by possible additive outliers and Gaussian noise. The observation model can be expressed as

\[
y_n = \sum_{r=1}^{R} m_r a_{r,n} + r_n + e_n, \quad n = 1, \ldots, N
\]

(1)

where \( m_r = [m_{r,1}, \ldots, m_{r,L}]^T \) is the spectrum of the \( r \)th material present in the scene and \( a_{r,n} \) is its corresponding proportion (abundance) in the \( n \)th pixel. In (1), \( e_n \) is an additive independently but non identically distributed zero-mean Gaussian noise sequence with diagonal covariance matrix \( \Sigma_0 = \text{diag}(\sigma^2) \), denoted as \( e_n \sim \mathcal{N}(e_n; 0, \Sigma_0) \), where \( \sigma^2 = [\sigma_1^2, \ldots, \sigma_L^2]^T \) is the vector of the \( L \) noise variances and \( \text{diag}(\sigma^2) \) is an \( L \times L \) diagonal matrix containing the elements of the vector \( \sigma^2 \). Moreover, \( r_n \) denotes the outlier vector of the \( n \)th pixel. Note that the usual matrix and vector notations \( M = [m_1, \ldots, m_R] \) and \( a_n = [a_{1,n}, \ldots, a_{R,n}]^T \) have been used in the second row of (1).

As a consequence of physical constraints, the abundance vectors \( a_n \) satisfy the following positivity and sum-to-one constraints

\[
\sum_{r=1}^{R} a_{r,n} = 1, \quad a_{r,n} > 0, \quad \forall r \in \{1, \ldots, R\}.
\]

(2)

The problem investigated in this paper is to estimate the endmember matrix \( M \), the abundance matrix \( A = [a_1, \ldots, a_N] \), the noise variances in \( \sigma^2 \) and the outlier matrix \( R = [r_1, \ldots, r_N] \) from the observation matrix \( Y = [y_1, \ldots, y_N] \). To solve this problem, we propose a hierarchical Bayesian model and a sampling method to estimate the unknown parameters.

III. ROBUST BAYESIAN LINEAR UNMIXING (RBLU)

A. Likelihood

Eq. (1) implies that \( y_n|M, a_n, r_n, \sigma^2 \sim \mathcal{N}(y_n; Ma_n + r_n, \Sigma_0) \). Assuming independence between noise sequences of the \( N \) observed pixels, the likelihood of the observation matrix \( Y \) can be expressed as

\[
f(Y|M, A, R, \sigma^2) \propto \left| \Sigma_0 \right|^{-N/2} \text{etr} \left[ -\frac{(Y-MA-R)^T \Sigma_0^{-1}(Y-MA-R)}{2} \right]
\]

(3)
where $\propto$ means “proportional to” and $\text{etr}(\cdot)$ denotes the exponential trace.

**B. Parameter priors**

1) **Prior for the abundance matrix $A$**: Each abundance vector can be written as $a_n = [c_n^T, a_{R,n}]^T$ with $c_n = [a_{1,n}, \ldots, a_{R-1,n}]^T$ and $a_{R,n} = 1 - \sum_{r=1}^{R-1} a_{r,n}$. The LMM constraints (2) impose that $c_n$ belongs to the simplex $S = \{c \mid c_r \geq 0, \forall r \in 1, \ldots, R - 1, \sum_{r=1}^{R-1} c_r \leq 1\}$. To reflect the lack of prior knowledge about the abundances, a uniform prior is assigned for each vector $c_n, n \in \{1, \ldots, N\}$, i.e., $f(c_n) \propto \mathbf{1}_S(c_n)$, where $\mathbf{1}_S(\cdot)$ is the indicator function defined on the simplex $S$. When prior knowledge about the abundances are available, the uniform prior can be replaced by more informative priors such as (mixtures of) Dirichlet distributions [15] or Gaussian mixtures using logistic coefficients [16]. Assuming prior independence between the $N$ abundance vectors $\{a_n\}_{n=1}^N$ leads to the following joint prior distribution

$$f(C) = \prod_{n=1}^N f(c_n), \quad (4)$$

where $C = [c_1, \ldots, c_N]$ is an $(R - 1) \times N$ matrix.

2) **Prior for the endmember matrix $M$**: To reflect the lack of prior knowledge about the endmembers, we use the following multivariate truncated Gaussian prior

$$f(M|\xi) \propto \prod_{r=1}^R N_{(\mathbb{R}^+)^L}(m_r; 0, \xi I_L) \quad (5)$$

where $\xi$ is fixed to a large 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 can be introduced ensuring that the endmember spectra belong to $(0, 1)$, such as a uniform, beta [17] or Gaussian distribution. Note that the prior can also include prior information from an endmember extraction algorithm, as in [18], [19].

3) **Prior for the noise variances**: A Jeffreys’ prior is chosen for the noise variance in each spectral band $\sigma^2_\ell$, i.e., $f(\sigma^2_\ell) \propto \sigma^{-2}_\ell \mathbf{1}_{\mathbb{R}^+}(\sigma^2_\ell)$ where $\mathbf{1}_{\mathbb{R}^+}(\cdot)$ denotes the indicator function defined on $\mathbb{R}^+$, which reflects the absence of knowledge about these parameters. Again, these non-informative priors can be easily replaced by conjugate inverse-Gamma priors to include prior knowledge available about the noise levels. Assuming prior independence between the noise variances, we obtain

$$f(\sigma^2) = \prod_{\ell=1}^L f(\sigma^2_\ell), \quad (6)$$
4) Priors of the outliers: As in [10], [12], the outliers are assumed to be sparse, i.e., at 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

\[
R = Z \odot X, \tag{7}
\]

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

\[
f(X|s^2) = \prod_{\ell,n} \mathcal{N}(x_{\ell,n}; 0, s^2), \tag{8}
\]

where \( s^2 \) controls the prior energy of the outliers. Note that (8) allows the outliers to be negative. Other conjugate priors, such as exponential or truncated Gaussian priors, could be used instead of (8), e.g., to enforce outlier positivity. The next paragraph presents the prior considered for the label matrix \( Z \).

5) Label matrix: For many applications, the locations of outliers are likely to be spectrally (e.g., water absorption bands) and/or spatially (e.g. weakly represented components, shadowing effects) correlated. An effective way to take correlated outliers/nonlinear effects into account is to consider Markov random fields (MRF) to build a prior for the label matrix \( Z \) [10]. 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.,

\[
P(z_{\ell,n}|Z_{\setminus z_{\ell,n}}) = P(z_{\ell,n}|Z_{V_{\ell,n}}),
\]

where \( V_{\ell,n} \) is the index set of the neighbors of \( z_{\ell,n} \), \( Z_{\setminus z_{\ell,n}} \) denotes the matrix \( Z \) whose element \( z_{\ell,n} \) has been removed and \( Z_{V_{\ell,n}} \) is the subset of \( Z \) composed of the elements whose indexes belong to \( V_{\ell,n} \). In this study, we consider that the spatial and spectral correlations can be different and thus consider two different neighborhoods. We decompose the neighborhood \( V_{\ell,n} \) as \( V_{\ell,n} = V^L_{\ell,n} \cup V^N_{\ell,n} \) where \( V^N_{\ell,n} \) (resp. \( V^L_{\ell,n} \)) denotes the spatial (resp. spectral) neighborhood of \( z_{\ell,n} \). In this paper, we consider an Ising model that can be expressed as

\[
P(Z|\beta') = \frac{1}{B(\beta')} \exp \left[ \beta'^T \phi(Z) + \phi_0(Z, \beta_0) \right] \tag{9}
\]
where \( \beta = [\beta_N, \beta_L]^T \), \( \beta' = [\beta^T, \beta_0]^T \) and
\[
\begin{align*}
\phi_L(Z) &= \sum_{n,\ell} \sum_{z_{\ell,n} \in \mathcal{V}_n^L} \delta(z_{\ell,n} - z_{\ell,n}'), \\
\phi_N(Z) &= \sum_{n,\ell} \sum_{z_{\ell,n} \in \mathcal{V}_n^N} \delta(z_{\ell,n} - z_{\ell,n}'), \\
\phi(Z) &= [\phi_L(Z), \phi_N(Z)]^T, \\
\phi_0(Z, \beta_0) &= \beta_0 \sum_{n,\ell} (1 - z_{\ell,n}) + (1 - \beta_0) \sum_{n,\ell} z_{\ell,n},
\end{align*}
\]
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. Specifically, the higher the value of \( \beta_0 \), the lower the probability of outliers in the data. The estimation of the proposed Ising model hyperparameters will be discussed in the next section. Different spectral and spatial neighbourhoods can be used in (9). In this paper, we consider a 4-neighbour structure to account for the spatial correlation and a 2-neighbour structure (i.e., the first previous and next bands) for the spectral dimension.

C. Outlier variance \( s^2 \)

The following conjugate inverse-Gamma prior is assigned to \( s^2 \)
\[
s^2 \sim IG(\gamma, \nu), \tag{10}
\]
where \( (\gamma, \nu) \) are fixed to \( (10^{-3}, 10^{-3}) \) to ensure a weakly informative prior. This choice of hyperparameters reflects the lack of prior information about the outliers variance. However, (10) could be replaced by more informative priors by suitably adapting \( (\gamma, \nu) \).

D. Joint posterior distribution

Assuming the parameters \( M, A, Z, X \) and \( \sigma^2 \) are a priori independent, the joint posterior of the parameter vector \( \theta = \{M, A, X, Z, \sigma^2\} \) and the parameter \( s^2 \) can be expressed as
\[
f(\theta, s^2 | Y, \phi, \beta') \propto f(Y | \theta)f(\theta | s^2, \xi, \beta)f(s^2 | \gamma, \nu) \tag{11}
\]
where
\[
f(\theta | s^2, \xi, \beta') = f(M | \xi)f(A)f(\sigma^2)f(X | s^2)P(Z | \beta') \tag{12}
\]
and \( \phi = [\xi, \gamma, \nu]^T \) a vector of model hyperparameters. The MRF parameter vector \( \beta' \) will be determined by maximum marginal likelihood estimation during the inference procedure. The directed acyclic graph (DAG) summarizing the structure of proposed Bayesian model is depicted in Fig. [1] The next section describes MCMC method for sampling from this posterior distribution in order to estimate the unknown model parameters.
IV. BAYESIAN INFERENCE

The Bayesian model defined in Section III specifies the joint posterior density for the unknown parameters $\theta, s^2$ given the observations $Y$ and the hyperparameters $\xi, \gamma, \nu$ and $\beta'$. This posterior distribution models the complete knowledge about the unknown parameters given the observed data and the prior information available. In this section we define suitable Bayesian estimators to summarize this knowledge and perform robust hyperspectral unmixing and outlier detection. Specifically, we propose the following Bayesian estimators for hyperspectral unmixing and nonlinearity detection: the marginal posterior mean or minimum mean square error estimator of the abundance and endmember matrices

$$
\left( \hat{A}_{\text{MMSE}}, \hat{M}_{\text{MMSE}} \right) = E \left[ A, M \mid Y, \phi, \hat{\beta}' \right],
$$

where the expectation is taken with respect to the marginal posterior density $f(A, M \mid Y, \phi, \hat{\beta}')$ (by marginalizing out $Z, X, \sigma^2$ and $s^2$, this density takes into account their uncertainty), the marginal maximum a posteriori (MMAP) estimator for the outlier support

$$
z_{n,\ell}^{\text{MMAP}} = \arg\max_{z_{n,\ell} \in \{0,1\}} f(z_{n,\ell} \mid Y, \phi, \hat{\beta}'),
$$

and, conditionally on the estimated outliers location, the minimum mean square error estimator of the outlier values

$$
r_{n,\ell}^{\text{MMSE}} = E \left[ x_{n,\ell} \mid z_{n,\ell} = z_{n,\ell}^{\text{MMAP}}, Y, \phi, \hat{\beta}' \right],
$$

where

$$
f(z_{n,\ell} \mid Y, \phi, \hat{\beta}') = \int f(\theta, s^2 \mid Y, \phi, \hat{\beta}') d\theta \mid_{z_{n,\ell}} ds^2,
$$

and

$$
f(x_{n,\ell} \mid z_{n,\ell}, Y, \phi, \hat{\beta}') = \frac{\int f(\theta, s^2 \mid Y, Y, \phi, \hat{\beta}') d\theta \mid_{z_{n,\ell}} ds^2}{f(z_{n,\ell} \mid Y, \phi, \hat{\beta}')}.
$$

Note that the outlier estimator (15) is sparse by construction (i.e., $E \left[ x_{n,\ell} \mid z_{n,\ell} = 0, Y, \phi, \hat{\beta}' \right] = 0$).

Computing (13), (14) and (15) is challenging because it requires having access to the joint marginal density of $(M, A)$, the univariate marginal densities of $z_{n,\ell}$ and the joint marginal densities of $(x_{n,\ell}, z_{n,\ell})$, which in turn require computing the posterior (11) and performing an integration over a very high-dimensional space. Fortunately, these estimators can be efficiently approximated with arbitrarily good accuracy by Monte Carlo integration. More precisely, it is
possible to compute (13), (14) and (15) by first using a Markov Chain Monte Carlo (MCMC) computational method to generate samples asymptotically distributed according to (11), and subsequently using these samples to approximate the required marginal probabilities and expectations. Note that in (13), (14) and (15), we have set $\beta' = \hat{\beta}'$, which denotes the maximum marginal likelihood estimator of the Ising regularisation hyperparameter vector $\beta'$ given the observed data $Y$, i.e.,

$$\hat{\beta}' = \arg\max_{\beta' \in B} f(Y|\phi, \beta').$$

(18)

This is an empirical Bayes approach for specifying $\beta'$ where hyperparameters with unknown values are replaced by point estimates computed from observed data (as opposed to being fixed a priori or integrated out of the model by marginalisation). As explained in [14], this strategy has several important advantages for MRF hyperparameters such as $\beta'$ having intractable conditional distributions. In particular, it allows automatic adjustment of the value of $\beta'$ to each image, producing significantly better estimation results than using a single fixed value of $\beta'$ for all images. Furthermore this empirical Bayes strategy has significantly lower computational cost as compared to that of competing approaches, such as including $\beta'$ in the model and subsequently marginalizing it during the inference procedure [20].

A. Bayesian estimation algorithm

Exact computation of the MMSE and MMAP estimators (13), (14) and (15) is very challenging because it involves calculating expectations with respect to posterior marginal densities, which in turn require evaluating the full posterior (11) and integrating it over a very high-dimensional space. Exact computation of $\hat{\beta}'$ is also difficult because it involves solving an intractable optimisation problem (it is not possible to evaluate the exact marginal likelihood $f(Y|\alpha_3)$ or its gradient $\nabla f(Y|\alpha_3)$). Here we follow the approach proposed in [14] and design a stochastic optimisation and simulation algorithm to compute (13), (14) and (15) simultaneously. We construct a stochastic gradient Markov chain Monte Carlo (SGMCMC) algorithm that simultaneously estimates $\hat{\beta}'$ and generates a chain of $N_{MC}$ samples $\{M^{(t)}, A^{(t)}\}_{t=1}^{N_{MC}}$ asymptotically distributed according to the marginal density $f(M, A|Y, \hat{\beta}')$ (this algorithm is summarised in Algorithm 1 below). Once the samples have been generated, the estimators (13), (14) and (15) are approximated by Monte Carlo integration [21] Chap.
\[ \hat{M}_{MMSE} = \frac{1}{N_{MC} - N_{bi}} \sum_{t=N_{bi}+1}^{N_{MC}} M^{(t)}, \]
\[ \hat{A}_{MMSE} = \frac{1}{N_{MC} - N_{bi}} \sum_{t=N_{bi}+1}^{N_{MC}} A^{(t)}, \]
\[ \hat{z}_{n,\ell}^{MMAP} = \begin{cases} 0 & \text{if } \text{card} (Z_{n,\ell}) \leq (N_{MC} - N_{bi}) / 2 \\ 1 & \text{else,} \end{cases} \]
\[ \hat{r}_{n,\ell}^{MMSE} = \begin{cases} \frac{1}{\text{card}(Z_{n,\ell})} \sum_{z_{n,\ell}^{(t)} \in Z_{n,\ell}} r_{n,\ell}^{(t)} & \text{if } z_{n,\ell}^{MMAP} = 1 \\ 0 & \text{else,} \end{cases} \]
with \( Z_{n,\ell} = \{ z_{n,\ell}^{(t)} | t \in \{N_{bi} + 1, \ldots, N_{MC}\}, z_{n,\ell}^{(t)} = 1 \} \) and where the samples from the first \( N_{bi} \) iterations (corresponding to the transient regime or burn-in period) are discarded. The main steps of this algorithm are detailed in below.

1) **Sampling the labels:** It can be seen from (11) that
\[ f(z_{n,\ell} = i | Y, \theta \setminus z_{n,\ell}, \beta', s^2) \propto \tilde{\pi}^{(i)}_{n,\ell}, \quad \forall (n, \ell), \] (19)
where \( i \in \{0, 1\} \) and
\[ \log \left( \frac{\tilde{\pi}^{(i)}_{n,\ell}}{\tilde{\pi}^{(0)}_{n,\ell} + \tilde{\pi}^{(1)}_{n,\ell}} \right) = -\frac{(y_{\ell,n} - m_{\ell,\beta_n - i \alpha_{\ell,n}})^2}{2\sigma^2_{\ell}} - \beta^T \phi(Z) - \phi_0 (Z, \beta_0). \] (20)
Consequently, the label \( z_{n,\ell} \) can be drawn from its conditional distribution by drawing randomly from \( \{0, 1\} \) with probabilities given by
\[ f(z_{n,\ell} = i | Y, \theta \setminus z_{n,\ell}, \beta', s^2) = \frac{\tilde{\pi}^{(i)}_{n,\ell}}{\tilde{\pi}^{(0)}_{n,\ell} + \tilde{\pi}^{(1)}_{n,\ell}}, \] (21)

2) **Sampling the endmembers:** It can be easily shown that
\[ f(M|Y, \theta \setminus M, s^2, \phi, \beta') = \prod_{\ell=1}^{L} f(m_{\ell,:}|Y, \theta \setminus m_{\ell,:}, s^2, \xi), \] (22)
i.e., the rows of \( M \), denoted as \( \{m_{\ell,:}\} \) are a posteriori independent (conditioned on the other parameters). Moreover,
\[ m_{\ell,:} | Y, \theta \setminus m_{\ell,:}, s^2, \xi \sim N_{(R^+)^n} (m_{\ell,:}, \hat{m}_{\ell,:}, S_{\ell}^{(M)}). \] (23)
where
\[
\tilde{\mathbf{m}}_\ell = \sigma_\ell^{-2} \tilde{y}_\ell, A^T S_\ell^{(M)} \\
S_\ell^{(M)} = \left( \sigma_\ell^{-2} A A^T + \xi^{-2} I_L \right)^{-1}
\] (24)
and \(\tilde{y}_\ell\) is the \(\ell\)th row of the \(L \times N\) matrix \(\tilde{Y} = Y - R\). Sampling from (23) can be achieved efficiently by using the Hamiltonian method recently proposed in [22] or by successive sampling from \(R\) truncated Gaussian distributions (via Gibbs sampling).

3) Sampling the abundances: In a similar fashion to obtaining (22), it can be easily shown that
\[
f(C|Y, \theta, C, s^2, \phi, \beta') = \prod_{n=1}^{N} f(c_n|y_n, \theta, c_n, s^2),
\] (25)
i.e., the columns of \(A\) are \textit{a posteriori} independent (conditioned on the other parameters). Moreover, the conditional distribution of \(c_n|y_n, \theta, s^2\) is a multivariate Gaussian distribution restricted to the simplex \(S\), which can be sampled efficiently using the method proposed in [22].

4) Sampling the latent variable matrix \(X\): In a similar manner to the abundance matrix in (25), the elements of \(X\) are \textit{a posteriori} independent (conditioned on the other parameters) and can be sampled independently. Since the prior (8) is conjugate to the Gaussian distribution, the full conditional distribution of \(x_{\ell,n}\) is given by
\[
x_{\ell,n} | y_{\ell,n}, \theta, x_{\ell,n}, s^2, \beta' \sim N(x_{\ell,n}; \tilde{x}_{\ell,n}, \tilde{\sigma}_{\ell,n}^2)
\] (26)
where
\[
\tilde{x}_{\ell,n} = \frac{z_{\ell,n}(y_{\ell,n} - \mathbf{m}_\ell, A_n)}{\tilde{\sigma}_{\ell,n}^2}
\]
\[
\tilde{\sigma}_{\ell,n}^2 = \frac{\sigma_\ell^2 s^2}{\sigma_\ell^2 + z_{\ell,n}s^2}.
\] (27)

5) Sampling the noise variances: Sampling the noise variances can be easily achieved by sampling from the following \(L\) independent inverse-Gamma distributions
\[
\sigma_\ell^2 | Y, \theta, s^2 \sim IG\left(\frac{N}{2}, \frac{\|y_{\ell:} - \mathbf{m}_\ell, A - r_{\ell:}\|^2}{2}\right)
\] (28)

6) Sampling the outlier variance \(s^2\): Finally, in a similar fashion to the noise variances, it can be shown that
\[
s^2 | Y, \theta, \phi \sim IG\left(\frac{NL}{2} + \gamma + \sum_{n,\ell} \frac{x_{\ell,n}^2}{2}\right)
\] (29)
7) **Updating the Ising regularisation model parameter vector** $\beta'$: If the marginal likelihood $f(Y|\phi, \beta')$ was tractable, we could update $\beta'$ from one MCMC iteration to the next by using a classic gradient descent step 

$$\beta^{(t)} = \beta^{(t-1)} + \delta_t \nabla \log f(Y|\phi, \beta^{(t-1)}),$$

with $\delta_t = t^{-3/4}$, such that $\beta^{(t)}$ converges to $\hat{\beta}'$ as $t \to \infty$. However, this gradient has two levels of intractability, one due to the marginalisation of $(\theta, s^2)$ and another one due to the intractable normalising constant of the Ising model. We address this difficulty by following the approach proposed in [14]; that is, by replacing $\nabla \log f(Y|\phi, \beta^{(t-1)}$ with an estimator computed with the samples generated by the MCMC algorithm at iteration $t$, and a set of two auxiliary variables $(Z') \sim \mathcal{K}(Z'|Z^{(t)}, \beta^{(t-1)})$ generated with an MCMC kernel $\mathcal{K}$ with target density (9) (in our experiments we used a Gibbs sampler implemented using a colouring scheme such that half of the elements of $Z'$ are generated in parallel). The updated value $\beta^{(t)}$ is then projected onto the domain $B_t = [0, B_t] \times [0, B_t] \times [0, 1]$ to guarantee the constraints of $\beta_N, \beta_L$ and $\beta_0$ and the stability of the stochastic optimisation algorithm, where $B_t$ is an arbitrarily large upper bound on $\beta_N$ and $\beta_L$ that can be increased at every iteration. In our experiments we have used $B_t = 10$.

It is worth mentioning that if it was possible to simulate the auxiliary variables $Z'$ exactly from (9) then the estimator of $\nabla \log f(Y|\phi, \beta^{(t-1)})$ used in Algorithm 1 would be unbiased and as a result $\beta^{(t)}$ would converge exactly to $\hat{\beta}'$. However, exact simulation from (9) is not computationally feasible and therefore we resort to the MCMC kernel $\mathcal{K}$ and obtain a biased estimator of $\nabla \log f(Y|\phi, \beta^{(t-1)})$ that drives $\beta^{(t)}$ to a neighbourhood of $\hat{\beta}'$ [14]. We found that computing this biased estimator is significantly less expensive than alternative approaches, e.g., using an approximate Bayesian computation algorithm as in [20], and that it leads to very accurate unmixing results.

Note that in Algorithm 1, $P_{[0,B_t]}(\cdot)$ denotes the projection onto $[0, B_t]$. In this paper, we propose an MCMC method that sequentially updates $Z, M, A, X, \sigma^2$ and $s^2$ at each sampler iteration. However, some of these variables, such as $(Z, X), (M, X)$ or $(A, X)$ could be updated simultaneously or could be re-sampled within a given sampler iteration, thus improving the mixing properties of the proposed sampling scheme, as in [12].

V. **GENERALIZATION TO ROBUST BAYESIAN PRINCIPAL COMPONENT ANALYSIS**

As mentioned previously, one of the main contributions of this paper is the introduction of a 3D Ising model to model the possible correlation between the outlier locations. This
outlier model has been applied to linear SU in Section III. However, this model can be applied to many other blind source separation (subspace identification) and inverse problems (nonlinear spectral unmixing) where the observations are corrupted by additive Gaussian noise and outliers. In this section, we discuss the generalization of the robust Bayesian principal component analysis model studied in [12], [23]. As in [23], we consider the following observation model (expressed in matrix form)

\[ Y = L + R + E \]  

where \( R \) (resp. \( E \)) are \( L \times N \) matrices representing the outliers (resp. the additive Gaussian noise) and \( L = D \Lambda W^T \) corresponds to a low rank matrix. The \( L \times R \) matrices \( D \) and \( W \) are matrices of the left- and right-singular vectors, respectively, and \( \Lambda = \text{diag} ([\lambda_1, \ldots, \lambda_R]) \) is a diagonal matrix consisting of the singular values \( \{\lambda_r\}_{r=1}^R \). Following the model considered in [12], [23], we can set \( \lambda_r = \zeta_r \eta_r \) where \( \zeta_r \in \{0,1\} \), which allows estimation of the dimension of the principal subspace. Since the columns of \( E \) are i.i.d., i.e.,

\[ e_n \sim \mathcal{N} (e_n; 0_L, \Sigma_0), \quad \forall n, \]  

we can assign appropriate prior distributions to \( \sigma^2, D, W, \{\zeta_r\} \) and \( \{\eta_r\} \), and consider the Bayesian model proposed in Section III for the outlier matrix \( R \) (i.e., Eqs. (7)-(10)). The resulting robust Bayesian principal component analysis model can then be used to estimate the data principal subspace in the presence of non-i.i.d. additive Gaussian noise and possibly correlated outliers. The sampling strategy associated with with Bayesian model is similar to those proposed in [12], [23] and that presented in Section IV and is not further developed here due to space constraints.

VI. SIMULATIONS USING SYNTHETIC DATA

A. Linear mixtures corrupted by outliers

The performance of the proposed method, referred to as “RBLU” (Robust Bayesian linear unmixing), is investigated on two synthetic \( 60 \times 60 \) pixel hyperspectral images composed of \( R = 3 \) endmembers and observed at \( L = 207 \) different spectral bands (see Fig. 2). 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 \( \sigma^2_\ell = 10^{-4}, \forall \ell \), which corresponds to an average signal-to-noise ratio (SNR) of 30dB. The first image \( I_1 \) does not contain outliers whereas the parameter \( s^2 \) 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 (9) with $\beta = [0.25; 0.25; 0.55]^T$ which leads to approximately 10% of actual outliers in $R$. The proposed method has been applied to the images with $N_{MC} = 1000$ iterations (including $N_{bi} = 300$). The endmember matrix was initialized using the VCA algorithm [24] and the abundance matrix was initialized 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

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

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

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

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

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

Finally, Table III 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 by virtue of 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). Here, we investigated the performance of RBLU in the presence of outliers in the data. The next paragraph discusses the unmixing problem when both linearly and nonlinearly mixed pixels are present in the image.
B. Unmixing of linear and nonlinear mixtures

As mentioned previously, many nonlinear models of the spectral unmixing literature can be expressed as (1). This is the case, for instance, for polynomial models [6]–[8] introduced to handle multiple scattering effects. The proposed model (1) seems particularly well adapted to detecting nonlinear effects that are often spatially localized, e.g., regions with significant relief variations, and/or spectrally concentrated such as nonlinear interactions (if present) varying smoothly over wavelengths. To evaluate the performance of the RBLU algorithm in terms of endmember estimation, abundance estimation and nonlinearity detection, we synthesized a 60×60 pixel image composed of the three endmembers considered in the previous paragraph. Most of the pixels (75%) were generated according to the classical LMM while the remaining 25% were generated according to the generalized bilinear model [8]

\[
y_n = M a_n + \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} \gamma_{i,j,n} a_{i,n} a_{j,n} m_i \odot m_j + e_n, \tag{34}
\]

where \(0 \leq \gamma_{i,j,n} \leq 1\) characterizes the level of interaction between the endmembers \(m_i\) and \(m_j\) in the \(n\)th pixel. The abundances of each pixel (linearly or nonlinearly) mixed were uniformly drawn from the simplex defined by the abundance positivity and sum-to-one constraints. The nonlinearity parameters in (34) were fixed to \(\gamma_{i,j,n} = 1\), which corresponds to the choice made in Fan’s model [25]. All pixels have been corrupted by a zero-mean additive Gaussian noise independently and identically distributed, with variance \(\sigma^2\), which corresponds to an average SNR of 28dB. The RBLU algorithm was applied to the data using \(N_{MC} = 2000\) iterations (including \(N_{bi} = 500\)).

Table IV compares the estimation performance of RBLU to the results obtained with 1) BLU [18], 2) VCA+FCLS [1], [24], 3) NfindR + FCLS [1], [26], 4) o-FCLS [1], 5) NfindR [26] followed by the gradient-based inversion step proposed in [8] based on the GBM (34) (referred to as “NfindR + GBM” in the table), 6) VCA [24] followed by the gradient-based inversion step proposed in [8] based on the GBM (34) (referred to as “VCA + GBM” in the table), and 7) the GBM method [8] applied to the data assuming the endmembers are known (referred to as “o-GBM” in the table for oracle-GBM). Due to its anomaly detection ability, RBLU generally provides better abundance estimates than LMM-based methods. Also RBLU does not rely on the GBM, it provides better abundance estimates than BLU, in particular for nonlinearly mixed pixels. For this data set containing pure pixels, N-FindR and VCA can
provide better endmember estimates than RBLU, which could be different in the absence of pure pixels (RBLU does not require the pure pixel assumption).

Fig. 3 shows the actual and estimated outlier energy in each pixel (i.e., $\|\hat{\mathbf{r}}_n\|_2^2$) and illustrates the ability of RBLU to detect bilinear mixtures in an image containing both linear and non-linear mixtures.

VII. SIMULATIONS USING REAL HYPER SPECTRAL DATA

A. Moffett data set

The first real image considered in this section is composed of $L = 189$ spectral bands and was acquired in 1997 by the AVIRIS satellite. The acquired image covers a region over Moffett Field, CA. A subimage of size $50 \times 50$ pixels has been chosen here to evaluate the performance of RBLU. The AVIRIS Moffet Field dataset has been previously used for comparing methods of linear [8], [18], [27]–[29] and nonlinear [8] unmixing. The subimage of interest is mainly composed of water, vegetation, and soil. As in in the previous subsection, the RBLU algorithm was applied with $N_{MC} = 2000$ iterations (including $N_{bi} = 500$).

Fig. 4 depicts the endmembers estimated by N-FindR and RBLU. Fig. 5 compares the abundance maps provided by RBLU to those obtained with N-FindR followed by FCLS. These figures show that the endmembers and abundance maps estimated by RBLU are in good agreement with those obtained by standard methods. In addition to the endmember and abundances estimates, RBLU provides spectral and spatial information about the possible outliers/nonlinearities. Fig. 6 (left) shows average spectral outlier energy over wavelength for the Moffett subimage. In addition to significant outlier levels near the water absorption bands, around $1400\text{nm}$ and $1800\text{nm}$, RBLU identifies important deviations from the linear mixing model for the spectral bands between $400\text{nm}$ and $800\text{nm}$. Fig. 6 (right) displays the estimated outlier energy (i.e., $\|\hat{\mathbf{r}}_n\|_2^2$) map over all pixels in the Moffett scene and shows that the deviations from the linear mixing model are mainly located in the coastal area, which is in agreement with the results obtained in [8], [27]. Fig. 7 compares the estimated spectrum of the outliers in the pixel $(40, 31)$ (located in the coastal area) to the estimated endmembers. Although RBLU promotes groups of outliers, it does not explicitly enforce spatial nor spectral dependencies for the outlier values. However, the outliers in this pixel seem to be spectrally correlated. These results show that RBLU is able to distinguish structured outliers from Gaussian noise. These results also show that the outlier spectrum and the vegetation spectral signature (in red) seem correlated, especially in the visible spectrum.
possible explanation for this correlation could be local significant changes of the vegetation, e.g., chlorophyll and water content, additional vegetation species, multiple scattering effects. Finally, the reconstruction errors obtained by N-FindR+FCLS and RBLU are compared in Fig. 8. Due to its outlier detection ability, RBLU provides lower reconstruction errors in the coastal area, which is the region where outliers/anomalies are the most dominant.

B. Villelongue data set

The second real image considered in this paper was acquired in 2010 under the Madonna project and collected by the Hyspex hyperspectral scanner over Villelongue, France (00°03’W and 42°57’N). \( L = 160 \) spectral bands were recorded from the visible to near infrared with a spatial resolution of 0.5m. This dataset has previously been studied in [10], [30], [31] and is mainly composed of forested and urban areas. More details about the data acquisition and pre-processing steps can be found in [30]. A sub-image of size 300 × 250 pixels is chosen here to evaluate the proposed unmixing procedure and is depicted in Fig. 9. The scene is composed mainly of trees and grass, resulting in \( R = 4 \) endmembers (soil, grass, trees and shade).

Fig. 10 compares the \( R = 4 \) endmembers estimated by N-FindR and RBLU for this second real image. Although it is difficult to assess the performance of the two EEAs for this image, it is interesting to note that the results obtained by the two methods are similar and lead to abundance maps (depicted in Fig. 11) that are in agreement with the true color image in Fig. 9. In particular, the two algorithms are able to identify the path (soil) in the scene. This path is barely visible in Fig. 9 but its presence can be confirmed using the Google Map image for this region (see Fig. 12).

Fig. 13 (left) depicts the anomaly energy map estimated by RBLU over the Villelongue image and highlights two main regions where significant deviations from the linear mixing model occur. The first region, on the top of Fig. 13 (left) is located where trees are identified and the deviations are likely to be due to the presence of different tree species. Note that this region can also be identified in Fig. 13 (right) (lighter green region). The second region representing a line in the centre of Fig. 13 (left) is more difficult to interpret and is barely visible in the true color image. However, it is interesting to note that the anomalies in this region form a stripe whose energy is higher in the pixels composed of grass than it is in those containing trees. Consequently, it is reasonable to postulate that the physical phenomena causing the deviations from the linear mixing model occur on the ground or below the surface,
i.e., probably not in the canopy. One possible cause of these spectral signature changes could be the presence of different kinds of surface vegetation. Finally, Fig. 14 presents examples of anomaly spectra of the first (green box in Fig. 13) and second (red box in Fig. 13) spatial regions. This figure shows that the deviations from the linear mixing model occur in specific spectral bands and are relatively similar within each spatial region. Moreover, the anomalies are spectrally different in the two regions, which confirms that they are probably due to different physical phenomena.

VIII. CONCLUSION

In this paper, we have proposed a Bayesian algorithm for robust linear spectral unmixing of hyperspectral images that performs joint endmember estimation, abundance estimation and outlier/anomaly detection. Appropriate prior distributions were used to enforce endmember and abundance positivity in addition to abundance sum-to-one constraints. Moreover, a 3D spatial-spectral Ising Markov random field was used to model correlations between outliers. Finally, an adaptive MCMC method was proposed to sample from the resulting posterior distribution in order to estimate the unknown model parameters. Simulations conducted on synthetic data showed superior performance of the proposed method for linear SU and the detection of outliers in hyperspectral images. The proposed method was also applied to real hyperspectral images and provided interesting results in terms of outlier analysis. Future work might include generalization of the proposed method for non-Gaussian observation noise.

Fig. 1. Directed acyclic graph representing the proposed hierarchical Bayesian model (fixed quantities appear in boxes).
Fig. 2. Actual endmembers (red lines) used to generate the synthetic images and endmembers estimated by VCA (black lines) and RBLU (dashed blue lines) for $I_2$.

Fig. 3. Actual (left) and estimated (right) anomaly energy in each pixel of the synthetic image composed of linearly and nonlinearly mixed pixels.
Fig. 4. The $R = 3$ endmembers extracted from the Moffett image by N-FindR (red) and RBLU (blue).

Fig. 5. The $R = 3$ abundance maps associated with the Moffett image and estimated by FCLS (top) and RBLU (bottom).
Fig. 6. Left: Average spectral energy of the outliers in the Moffett scene estimated by RBLU. Right: Estimated outlier energy $\|r_n\|_2^2$ in each pixel of the Moffett image.

Fig. 7. Endmembers and outlier signature of the pixel (40,31) estimated by RBLU for the Moffett image.
Fig. 8. Reconstruction errors of the Moffett image using N-FindR+FCLS (left) and RBLU.

Fig. 9. True color image of the Villelongue area (left) and sub-image of interest (right).
Fig. 10. The $R = 4$ endmembers extracted from the Villelongue image by N-FindR (red) and RBLU (blue).
Fig. 11. The $R = 4$ abundance maps associated with the Villelongue image and estimated by FCLS (right) and RBLU (left).
Fig. 12. Soil abundance map estimated by RBLU for the Villelongue image (left) and corresponding Google Map image (right) highlighting the presence of a path in the region of interest.

Fig. 13. Left: Anomaly energy map estimated by RBLU for the Villelongue scene. Right: sub-image of interest in true colors.
Fig. 14. Anomaly spectra in the green (top) and red (bottom) boxes depicted in Fig 13.

|       | $I_1$ | $I_2$ |
|-------|-------|-------|
|       | SAM ($\times 10^{-2}$) | RBLU | VCA-FCLS | $\alpha$-FCLS |
|       |       | $m_1$ | 0.21     | 0.68 | -          |
|       |       | $m_2$ | 0.17     | 0.92 | -          |
|       |       | $m_3$ | 0.26     | 1.96 | -          |
|       | RN MSE ($\times 10^{-2}$) | 0.68 | 1.60 | **0.67** |
|       | SAM ($\times 10^{-2}$) | $m_1$ | 0.19     | 4.03 | -          |
|       |       | $m_2$ | 0.20     | 3.08 | -          |
|       |       | $m_3$ | 0.29     | 4.26 | -          |
|       | RN MSE ($\times 10^{-2}$) | **0.74** | 8.27 | 6.95 |

TABLE I
ESTIMATION PERFORMANCE.
| $z = 0$ | $z = 1$ | Total |
|--------|--------|-------|
| $\hat{z} = 0$ | 652724 | 7190 | 659914 |
| $\hat{z} = 1$ | 789 | 84497 | 85286 |
| Total | 653513 | 91687 | 745200 |

**TABLE II**

OUTLIER DETECTION ($I_2$): CONFUSION MATRIX.

| $s^2 = 0.01$ | $s^2 = 0.1$ |
|--------------|--------------|
| RBLU | o-FCLS | RBLU | o-FCLS |
| 10% | 0.77 | 2.00 | 0.74 | 6.95 |
| 20% | 0.80 | 3.06 | 0.77 | 8.89 |
| 30% | 0.86 | 3.66 | 0.87 | 10.34 |

**TABLE III**

ABUNDANCE RNMSE ($\times 10^{-2}$) FOR DIFFERENT OUTLIER ENERGIES AND PROPORTIONS.

| RNMSE ($\times 10^{-2}$) | SAM ($\times 10^{-2}$) |
|--------------------------|------------------------|
| o-FCLS | m$_1$ | m$_2$ | m$_3$ |
| VCA+FCLS | 4.46 | 0.35 | 0.55 | 0.90 |
| NfindR + FCLS | 5.96 | 0.65 | 3.09 | 0.80 |
| o-GBM | 2.28 | -- | -- | -- |
| VCA+GBM | 2.15 | 0.35 | 0.55 | 0.90 |
| NfindR + GBM | 2.26 | 0.65 | 3.09 | 0.80 |
| BLU | 4.03 | 1.76 | 1.94 | 1.33 |
| RBLU | 3.81 | 2.79 | 3.56 | 0.59 |

**TABLE IV**

ABUNDANCE AND ENDMEMBER ESTIMATION FOR THE SYNTHETIC IMAGE COMPOSED OF LINEARLY AND NONLINEARLY MIXED PIXELS.
Algorithm 1

RBLU algorithm

1: Fixed input parameters: Number of endmembers $R$, number of burn-in iterations $N_{bi}$, total number of iterations $N_{MC}$

2: Initialization ($t = 0$)
   - Set $A^{(0)}, M^{(0)}, X^{(0)}, Z^{(0)}, \sigma^{2(0)}, s^{2(0)}, \beta^{(0)}$

3: Iterations ($1 \leq t \leq N_{MC}$)

4: Sample $Z^{(t)}$ from (21)

5: Sample $M^{(t)}$ from (23)

6: Sample $A^{(t)}$ from (25)

7: Sample $X^{(t)}$ from (26)

8: Sample $\sigma^{2(t)}$ from (28)

9: Sample $s^{2(t)}$ from (29)

10: if $t < N_{bi}$ then

11:   Sample $Z^{'} \sim K(Z|Z^{(t)}), \beta^{(t-1)}$

12:   Set $\beta^{(t)}_N = P_{[0, B_t]} \left( \beta^{(t-1)}_N + \Delta \beta_N \right)$ with

13:   $\Delta \beta_N = \delta_t \left[ \frac{d}{d\beta_N} \log f(Z|\beta^{'}) - \frac{d}{d\beta_N} \log f(Z'|\beta^{'}) \right]$

14:   Set $\beta^{(t)}_L = P_{[0, B_t]} \left( \beta^{(t-1)}_L + \Delta \beta_L \right)$ with

15:   $\Delta \beta_L = \delta_t \left[ \frac{d}{d\beta_L} \log f(Z|\beta^{'}) - \frac{d}{d\beta_L} \log f(Z'|\beta^{'}) \right]$

16:   Set $\beta^{(t)}_0 = P_{[0, 1]} \left( \beta^{(t-1)}_0 + \Delta \beta_0 \right)$ with

17:   $\Delta \beta_0 = \delta_t \left[ \frac{d}{d\beta_0} \log f(Z|\beta^{'}) - \frac{d}{d\beta_0} \log f(Z'|\beta^{'}) \right]$

18: else

19:   Set $\beta^{(t)} = \beta^{(t-1)}$

20: end if

21: Set $t = t + 1.$

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] R. Heylen, M. Parente, and P. Gader, “A review of nonlinear hyperspectral unmixing methods,” *IEEE J. Sel. Topics in Appl. Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 1844–1868, June 2014.

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

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] J. Nascimento and J. Bioucas-Dias, “Hyperspectral unmixing based on mixtures of dirichlet components,” *IEEE Trans. Geosci. and Remote Sensing*, vol. 50, no. 3, pp. 863–878, March 2012.

[16] O. Eches, N. Dobigeon, and J.-Y. Tourneret, “Enhancing hyperspectral image unmixing with spatial correlations,” *IEEE Trans. Geosci. and Remote Sensing*, vol. 49, no. 11, pp. 4239–4247, Nov. 2011.

[17] 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.

[18] 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.

[19] 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.

[20] M. Pereyra, N. Dobigeon, H. Batatia, and J.-Y. Tourneret, “Estimating the granularity coefficient of a Potts-Markov random field within an MCMC algorithm,” *IEEE Trans. Image Processing*, vol. 22, no. 6, pp. 2385–2397, June 2013.

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

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

[23] 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.
[24] 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.

[25] W. Fan, B. Hu, J. Miller, and M. Li, “Comparative study between a new nonlinear model and common linear model for analysing laboratory simulated-forest hyperspectral data,” Remote Sensing of Environment, vol. 30, no. 11, pp. 2951–2962, June 2009.

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

[27] O. Besson, N. Dobigeon, and J.-Y. Tourneret, “Minimum mean square distance estimation of a subspace,” IEEE Trans. Signal Processing, vol. 59, no. 12, pp. 5709–5720, Dec. 2011.

[28] O. Eches, N. Dobigeon, C.Mailhes, and J.-Y. Tourneret, “Bayesian estimation of linear mixtures using the normal compositional model,” IEEE Trans. Image Processing, vol. 19, no. 6, pp. 1403–1413, June 2010.

[29] O. Eches, N. Dobigeon, and J.-Y. Tourneret, “Estimating the number of endmembers in hyperspectral images using the normal compositional model and a hierarchical bayesian algorithm,” IEEE J. Sel. Topics Signal Processing, vol. 3, no. 3, pp. 582–591, June 2010.

[30] D. Sheeren, M. Fauvel, S. Ladet, A. Jacquin, G. Bertoni, and A. Gibon, “Mapping ash tree colonization in an agricultural mountain landscape: Investigating the potential of hyperspectral imagery,” in Proc. IEEE Int. Conf. Geosci. and Remote Sensing (IGARSS), July 2011, pp. 3672–3675.

[31] Y. Altmann, N. Dobigeon, S. McLaughlin, and J. Tourneret, “Nonlinear spectral unmixing of hyperspectral images using Gaussian processes,” IEEE Trans. Signal Process., vol. 61, no. 10, pp. 2442–2453, May 2013.