RESIDUAL COMPONENT ANALYSIS OF HYPERSPECTRAL IMAGES FOR JOINT NONLINEAR UNMIXING AND NONLINEARITY DETECTION

Yoann Altmann[(1)]*, *Nicolas Dobigeon*[(1)], *Steve McLaughlin*[(2)], *Jean-Yves Tourneret*[(1)]

[(1)]University of Toulouse, IRIT/INP-ENSEEIHT, Toulouse, France.
[(2)]Heriot-Watt University, School of Engineering and Physical Sciences, Edinburgh, United Kingdom.

ABSTRACT

This paper presents a nonlinear mixing model for joint hyperspectral image unmixing and nonlinearity detection. The proposed model assumes that the pixel reflectances are linear mixtures of endmembers, corrupted by an additional nonlinear term and an additive Gaussian noise. A Markov random field is considered for nonlinearity detection based on the spatial structure of the nonlinear terms. The observed image is segmented into regions where nonlinear terms, if present, share similar statistical properties. A Bayesian algorithm is proposed to estimate the parameters involved in the model yielding a joint nonlinear unmixing and nonlinearity detection algorithm. Simulations conducted with synthetic and real data show the accuracy of the proposed unmixing and nonlinearity detection strategy for the analysis of hyperspectral images.

***Index Terms*—** Hyperspectral imagery, nonlinear spectral unmixing, residual component analysis, nonlinearity detection.

1. INTRODUCTION

Spectral unmixing (SU) of hyperspectral images has attracted growing interest over the last few decades. It consists of distinguishing the materials and quantifying their proportions in each pixel of the observed image. The SU problem has been widely studied for the applications where pixel reflectances are linear combinations of pure component spectra [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 relief. 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 different kinds of nonlinearities to be approximated. These flexible models can be constructed from neural networks, kernels [8], or post-nonlinear transformations [9].

While the consideration of nonlinear effects can be relevant in specific areas, the LMM is often sufficient for approximating the actual mixing models in some image pixels or homogeneous regions. Consequently, it makes sense to distinguish in any image, linearly mixed pixels which can be easily analyzed, from those nonlinearly mixed requiring deeper analysis. Conversely to the nonlinearity detector studied in [10], this paper proposes to simultaneously achieve SU and nonlinearity detection. Moreover, the consideration of spatial structures in the image, used in [11] for linear SU, is investigated to infer the locations where nonlinear effects occur. The algorithm proposed in this paper is supervised in the sense that the endmembers contained in the image are assumed to be known. This algorithm is based on a nonlinear mixing model inspired from residual component analysis (RCA) [12]. In the context of SU of hyperspectral images, the nonlinear effects are modeled by additive perturbation terms characterized by Gaussian processes (GPs). This allows the nonlinear terms to be marginalized, yielding a flexible model depending only on the nonlinearity energies. The hyperspectral image to be analyzed is partitioned into homogeneous regions in which the nonlinearities share the same GP. The proposed algorithm relies on an implicit image classification, modeled by labels whose spatial dependencies follow a Potts-Markov random field. Consideration of two classes (linear vs. nonlinear mixtures) would lead to binary detection maps. However, this paper allows for several nonlinearly mixed regions to be also identified, based on the energy of the nonlinear effects.

The remaining paper is organized as follows. Section 2 introduces the RCA model for hyperspectral image analysis, followed by Section 3 which summarizes the likelihood and the priors assigned to the parameters of the linear part of the RCA model. Section 4 is devoted to the priors associated with the nonlinear terms of the RCA model. The posterior distribution of the RCA model and the Metropolis-Within-Gibbs sampler used to sample from it are summarized in Section 5. Some simulation results conducted on real data are shown and discussed in Section 6. Conclusions and future work are finally reported in Section 7.

2. PROBLEM FORMULATION

We consider a set of N observed pixel 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 defined as a linear combination of R known spectra $\mathbf{m}_r$, referred to as endmembers, contaminated by an additional spectrum $\boldsymbol{\phi}_n$ and additive noise

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

where $\mathbf{m}_r = [m_{r,1}, \ldots, m_{r,L}]^T$ is the spectrum of the rth material present in the scene, $a_{r,n}$ is its corresponding proportion (abundance) in the nth pixel. In (1), $\mathbf{e}_n$ is an additive independently and non identically distributed zero-mean Gaussian noise sequence with diagonal covariance matrix $\boldsymbol{\Sigma}_0 = \text{diag}\left(\boldsymbol{\sigma}^2\right)$, denoted as $\mathbf{e}_n \sim$

*Part of this work was supported by the Direction Générale de l'Armement, French Ministry of Defence and by the Hypanema ANR Project, ANR Project n°12 − $BS03$ − 003. Part of this work was initiated during the thematic trimester on image processing which was held in Toulouse in 2013 and was supported by the Labex CIMI (Centre International de Mathmatiques et d'Informatique), ANR-11-LABX-0040-CIMI, within the program ANR-11-IDEX-0002-02

$\mathcal{N}(\mathbf{0}_L, \mathbf{\Sigma}_0)$, where $\boldsymbol{\sigma}^2 = [\sigma_1^2, \ldots, \sigma_L^2]^T$ is the vector of the L noise variances and diag $\left(\boldsymbol{\sigma}^2\right)$ is an $L \times L$ diagonal matrix containing the elements of the vector $\boldsymbol{\sigma}^2$. 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). Moreover, the term $\boldsymbol{\phi}_n \in \mathbb{R}^L$ in (1) is an unknown additive perturbation vector modeling nonlinear effects occurring in the nth pixel. There are several motivations for considering the mixing model described by (1): i) the model reduces to the classical linear mixing model (LMM), for $\boldsymbol{\phi}_n = \mathbf{0}_L$, ii) the model is general enough to handle different of kinds of nonlinearities such as the bilinear model studied in [13] (referred to as Fan model (FM)), the generalized bilinear model (GBM) [7], and the polynomial post-nonlinear mixing model (PPNMM) studied for nonlinear spectral unmixing in [9] and nonlinearity detection in [10]. These models assume that the mixing model consists of a linear contribution of the endmembers, corrupted by at least one additive term characterizing the nonlinear effects. In the proposed model, all additive terms are gathered in the vector $\boldsymbol{\phi}_n$. Note that a similar model, called robust LMM, has been recently introduced in [14].

Due to physical considerations, 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 addressed in this paper consists of the joint estimation of the abundance vectors and detection of nonlinearly mixed pixels (characterized by $\boldsymbol{\phi}_n \neq \mathbf{0}_L$). The next sections present the proposed Bayesian model for joint unmixing and nonlinearity detection.

3. BAYESIAN LINEAR MODEL

The unknown parameter vector associated with the proposed model (1) contains the abundances $\mathbf{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N]$, the nonlinear terms of each pixel $\{\boldsymbol{\phi}_n\}_{n=1,\ldots,N}$, and the noise variance vector $\boldsymbol{\sigma}^2$. This section presents the likelihood and the parameter priors associated with the parameters of the linear part of the model, i.e., $\mathbf{A}$ and $\boldsymbol{\sigma}^2$. The characterization of the nonlinearities will be addressed later in Section 4.

3.1. Likelihood

Eq. (1) shows that $\mathbf{y}_n|\mathbf{M}, \boldsymbol{a}_n, \boldsymbol{\phi}_n, \boldsymbol{\sigma}^2 \sim \mathcal{N}(\mathbf{M}\boldsymbol{a}_n + \boldsymbol{\phi}_n, \mathbf{\Sigma}_0)$. Assuming independence between the observed pixels, the likelihood of the observation matrix $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N]$ can be expressed as $f(\mathbf{Y}|\mathbf{M}, \mathbf{A}, \mathbf{\Phi}, \boldsymbol{\sigma}^2)$

$$\propto \quad |\mathbf{\Sigma}_0|^{-N/2} \text{etr}\left[-\frac{(\mathbf{Y}-\mathbf{X})^T\mathbf{\Sigma}_0^{-1}(\mathbf{Y}-\mathbf{X})}{2}\right] \tag{3}$$

where $\propto$ means "proportional to", $\text{etr}(\cdot)$ denotes the exponential trace, $\mathbf{X} = \mathbf{M}\mathbf{A} + \mathbf{\Phi}$ is an $L \times N$ matrix and $\mathbf{\Phi} = [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_N]^T$ is an $L \times N$ nonlinearity matrix.

3.2. Prior for the abundance matrix A

Each abundance vector can be written as[1] $\boldsymbol{a}_n = [\mathbf{c}_n^T, a_{R,n}]^T$ 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 $\mathcal{S} = \left\{\mathbf{c} \,\middle|\, c_r > 0, \forall r \in 1, \ldots, R-1, \sum_{r=1}^{R-1} c_r < 1\right\}$. 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 $f(\mathbf{C}) = \prod_{n=1}^{N} f(\mathbf{c}_n)$, where $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_N]$ is an $(R-1) \times N$ matrix.

[1] In this paper, the sum-to-one constraint is considered for the abundances. However, this constraint can be relaxed. An extended algorithm (without abundance sum-to-one constraint) has been proposed and studied in [15].

3.3. Prior for the noise variances

A Jeffreys' prior is chosen for the noise variance of each spectral band σ_ℓ^2, i.e., $f(\sigma_\ell^2) \propto \sigma_\ell^{-2}\mathbf{1}_{\mathbb{R}^+}\left(\sigma_\ell^2\right)$, which reflects the absence of knowledge for this parameter (see [16] for motivations). Assuming prior independence between the noise variances, we obtain $f(\boldsymbol{\sigma}^2) = \prod_{\ell=1}^{L} f(\sigma_\ell^2)$.

4. MODELING THE NONLINEARITIES

We propose in this paper to exploit spatial correlations between the pixels of the hyperspectral image to be analyzed. It seems reasonable to assume that nonlinear effects occurring in a given pixel are related to the nonlinear effects present in neighboring pixels. Formally, the hyperspectral image is assumed to be partitioned into K classes denoted as $\mathcal{C}_0, \ldots, \mathcal{C}_{K-1}$. Let $\mathcal{I}_k \subset \{1, \ldots, N\}$ denote the subset of pixel indexes belonging to the kth class ($k = 0, \ldots, K-1$). An $N \times 1$ label vector $\mathbf{z} = [z_1, \ldots, z_N]^T$ with $z_n \in \{0, \ldots, K-1\}$ is introduced to identify the class of each image pixel, i.e., $\mathbf{y}_n \in C_k \Leftrightarrow n \in \mathcal{I}_k \Leftrightarrow z_n = k$. In each class, the unknown nonlinearity vectors are assumed to share the same statistical properties, as will be shown in the sequel.

4.1. Prior distribution for the nonlinearity matrix Φ

As mentioned above, the mixing model (1) reduces to the LMM for $\boldsymbol{\phi}_n = \mathbf{0}_L$. For nonlinearity detection, it makes sense to consider a pixel class (referred to as class $\mathcal{C}_0$) corresponding to linearly mixed pixels. The resulting prior distribution for $\boldsymbol{\phi}_n$ conditioned upon $z_n = 0$ is given by $f(\boldsymbol{\phi}_n|z_n = 0) = \prod_{\ell=1}^{L} \delta(\phi_{\ell,n})$.

Nonlinear effects can vary, depending on the relief of the scene, the underlying components involved in the mixtures and the observation conditions to name a few factors. This makes the choice of a single informative prior distribution challenging. From a classification point of view, it is interesting to identify regions or classes where similar nonlinearities occur. For these reasons, we propose to divide nonlinearly mixed pixels into $K-1$ classes and to assign different priors for the nonlinearity vectors belonging to the different classes. In this paper, the nonlinearities associated with nonlinearly mixed pixels are assumed to be random. Assuming $\mathbf{y}_n$ belongs to the kth class, the prior distribution of the corresponding nonlinear term $\boldsymbol{\phi}_n$ is given by the following GP ($k = 1, \ldots, K-1$)

$$\boldsymbol{\phi}_n|\mathbf{M}, z_n = k, s_k^2 \sim \mathcal{N}\left(\mathbf{0}_L, s_k^2\mathbf{K}_{\mathbf{M}}\right), \tag{4}$$

where $\mathbf{K}_{\mathbf{M}}$ is an $L \times L$ covariance matrix parameterized by the endmember matrix $\mathbf{M}$ and s_k^2 is a scaling hyperparameter that tunes the energy of the nonlinearities in the kth class. Note that all nonlinearity vectors within the same class share the same prior. It has been shown that polynomial models (i.e., models involving polynomial nonlinearities with respect to the endmembers) are particularly well adapted to model scattering effects, mainly observed in vegetation and urban areas. Consequently, it makes sense to assume that the nonlinearities $\boldsymbol{\phi}_n$ depend on the endmember matrix $\mathbf{M}$. In this paper, we consider the similarity matrix of symmetric second order polynomial kernel. This kernel is defined as follows

$$[\mathbf{K}_{\mathbf{M}}]_{i,j} = \left(\mathbf{m}_{i,:}^T\mathbf{m}_{j,:}\right)^2, \quad i,j \in \{1, \ldots, L\}, \tag{5}$$

where $\mathbf{m}_{i,:}$ denotes the ith row of $\mathbf{M}$. Note that the parametrization of the matrix $\mathbf{K}_{\mathbf{M}}$ in (5) only involves bilinear and quadratic terms with respect to the endmembers. Note also that a polynomial kernel similar to (5) has been recently considered in [8] and that other kernels such as the Gaussian kernel could be investigated to model other nonlinearities as in [12]. As mentioned above, the endmembers of the scene are assumed to be known in this paper. Consequently, the proposed nonlinear model does not involve endmember estimation errors (i.e., missing or poorly estimated endmembers).

4.2. Prior distribution for the label vector z

To exploit the correlation between pixels, a Markov random field is introduced as a prior distribution for z_n given its neighbors $\mathbf{z}_{\mathcal{V}(n)}$, i.e., $f(z_n|\mathbf{z}_{\backslash n}) = f(z_n|\mathbf{z}_{\mathcal{V}(n)})$, where $\mathcal{V}(n)$ is the neighborhood of the nth pixel and $\mathbf{z}_{\backslash n} = \{z_{n'}\}_{n' \neq n}$. More precisely, this paper focuses on the Potts-Markov model since it is appropriate for hyperspectral image segmentation [11]. Given a discrete random field $\mathbf{z}$ attached to an image with N pixels, using the Hammersley-Clifford theorem, we obtain $f(\mathbf{z}) = G(\beta)^{-1} \exp[\beta \sum_{n=1}^{N} \sum_{n' \in \mathcal{V}(n)} \delta(z_n - z_{n'})]$, where $\beta > 0$ is the granularity coefficient, $G(\beta)$ is a normalizing (or partition) constant and $\delta(\cdot)$ is the Dirac delta function. Several neighborhood structures can be employed to define $\mathcal{V}(n)$. In the rest of the paper, a second-order neighborhood will be considered. The hyperparameter β tunes the degree of homogeneity of each region in the image. More precisely, small values of β yield an image with a large number of regions, whereas large values of β lead to fewer and larger homogeneous regions. In this paper, the granularity coefficient is assumed to be known. Note however that it could be also included within the Bayesian model and estimated using the strategy described in [17].

4.3. Hyperparameter priors

The performance of the proposed Bayesian model for spectral unmixing depends on the values of the hyperparameters $\left\{s_k^2\right\}_{k=1,\ldots,K}$. When the hyperparameters are difficult to adjust, it is the norm to include them in the unknown parameter vector, resulting in a hierarchical Bayesian model [18, 9]. This strategy requires the definition of prior distributions for the hyperparameters. Inverse-gamma prior distributions are assigned to the nonlinearity hyperparameters (i.e., $s_k^2|\gamma,\nu \sim \mathcal{IG}(\gamma,\nu), \forall k \in \{1,\ldots,K\}$), where the additional parameters (γ,ν) are fixed to ensure a noninformative prior for s_k^2 ($(\gamma,\nu) = (1, 1/4)$ in all simulations presented in this paper). Assuming prior independence between the hyperparameters, we obtain $f(\mathbf{s}^2|\gamma,\nu) = \prod_{k=1}^{K-1} f(s_k^2|\gamma,\nu)$, where $\mathbf{s}^2 = [s_1^2,\ldots,s_K^2]^T$.

5. BAYESIAN INFERENCE USING A METROPOLIS-WITHIN-GIBBS SAMPLER

Assuming prior independence between $\mathbf{A}$, $(\mathbf{\Phi},\mathbf{z})$ and $\boldsymbol{\sigma}^2$, the posterior distribution of $(\mathbf{\Phi},\boldsymbol{\theta})$ where $\boldsymbol{\theta} = (\mathbf{C},\mathbf{z},\boldsymbol{\sigma}^2,\mathbf{s}^2)$ can be expressed as $f(\boldsymbol{\theta},\mathbf{\Phi}|\mathbf{Y},\mathbf{M}) \propto f(\mathbf{Y}|\mathbf{M},\boldsymbol{\theta},\mathbf{\Phi}) f(\mathbf{\Phi}|\mathbf{M},\mathbf{z},\mathbf{s}^2) f(\boldsymbol{\theta})$, where $f(\boldsymbol{\theta}) = f(\mathbf{C}) f(\boldsymbol{\sigma}^2) f(\mathbf{z}) f(\mathbf{s}^2)$. This distribution can be marginalized with respect to $\mathbf{\Phi}$, leading to $f(\boldsymbol{\theta}|\mathbf{Y},\mathbf{M}) \propto f(\boldsymbol{\theta}) f(\mathbf{Y}|\mathbf{M},\boldsymbol{\theta})$ where

$$f(\mathbf{Y}|\mathbf{M},\boldsymbol{\theta}) = \int f(\mathbf{Y}|\mathbf{M},\boldsymbol{\theta},\mathbf{\Phi}) f(\mathbf{\Phi}|\mathbf{M},\mathbf{z},\mathbf{s}^2) \mathrm{d}\mathbf{\Phi} \quad (6)$$
$$\propto \prod_{k=0}^{K-1} \prod_{n \in \mathcal{I}_k} \frac{1}{|\mathbf{\Sigma}_k|^{\frac{1}{2}}} \exp\left[-\frac{1}{2} \bar{\mathbf{y}}_n^T \mathbf{\Sigma}_k^{-1} \bar{\mathbf{y}}_n\right]$$

with $\mathbf{\Sigma}_k = s_k^2 \mathbf{K}_{\mathbf{M}} + \mathbf{\Sigma}_0$ $(k = 1,\ldots,K-1)$ and $\bar{\mathbf{y}}_n = \mathbf{y}_n - \mathbf{M}\boldsymbol{a}_n$. The advantage of this marginalization is to avoid sampling the matrix $\mathbf{\Phi}$. Thus, the nonlinearities are fully characterized by the known endmember matrix, the class labels and the values of $\mathbf{s}^2$.

To overcome the challenging derivation of the Bayesian estimators associated with $f(\boldsymbol{\theta}|\mathbf{Y},\mathbf{M})$, we propose to use an efficient Markov Chain Monte Carlo (MCMC) method to generate samples asymptotically distributed according to $f(\boldsymbol{\theta}|\mathbf{Y},\mathbf{M})$. 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 [19, Chap. 10]. In this paper, we propose to sample sequentially the N labels in $\mathbf{z}$, the abundance matrix $\mathbf{A}$, the noise variances $\boldsymbol{\sigma}^2$ and $\mathbf{s}^2$ using moves that are summarized below (see [15] for further details).

Labels: Sampling z_n from its conditional distribution can be achieved by drawing a discrete value in the finite set $\{0,\ldots,K-1\}$ with known probabilities.

Abundances: The conditional distribution of $\mathbf{c}_n|\mathbf{y}_n,\mathbf{M},z_n = k, s_k^2, \boldsymbol{\sigma}^2$ is a multivariate Gaussian distribution restricted to the simplex $\mathcal{S}$, which can be sampled efficiently using the method recently proposed in [20].

Noise variance σ_ℓ^2: Sampling σ_ℓ^2 from its conditional distribution is not straightforward. We propose to use an accept/reject procedure to update σ_ℓ^2. In this paper, a classical Gaussian random walk (in the log-space) is used. Note that the noise variances are a posteriori independent and can thus be updated in a parallel manner. Note also that the variances of the L parallel Gaussian random walks have been adjusted during the burn-in period of the sampler to obtain an acceptance rate close to 0.5, as recommended in [21, p. 8].

Hyperparameter s_k^2: Due to the complexity of the conditional distribution of $s_k^2|\mathbf{Y},\mathbf{A},\mathbf{z},\boldsymbol{\sigma}^2$, Gaussian random walks are again used in the log-space to update the hyperparameters $\{s_k^2\}_{k=1,\ldots,K-1}$ (similarly to the noise variance updates). Again, the variances of the random walks have been adjusted during the burn-in period of the sampler. The reader is invited to consult [15] for further details about the proposed sampler.

After generating N_{MC} samples using the proposed MCMC method and removing N_{bi} iterations associated with the burn-in period of the sampler (N_{bi} has been set from preliminary runs), the marginal maximum a posteriori (MAP) estimator of the label vector, denoted as $\hat{\mathbf{z}}_{\text{MAP}}$, can be computed. The label vector estimator is then used to compute the minimum mean square error (MMSE) of $\mathbf{A}$ conditioned upon $\mathbf{z} = \hat{\mathbf{z}}_{\text{MAP}}$. Finally, the noise variances and the hyperparameters $\{s_k^2\}_{k=1,\ldots,K-1}$ are estimated using the empirical averages of the generated samples (MMSE estimates) conditionally to $\mathbf{z} = \hat{\mathbf{z}}_{\text{MAP}}$.

6. SIMULATIONS: REAL HYPERSPECTRAL IMAGE

The performance of the proposed algorithm has first been evaluated using synthetic data. Due to length limitations, these simulations have been included in a technical report [15]. This section studies the performance of the proposed algorithm for a real hyperspectral image. The real image was acquired in 2010 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 already been studied in [22, 23] and is mainly composed of forested and urban areas. More details about the data acquisition and pre-processing steps are available in [22]. A sub-image (of size 41×29 pixels) is chosen here to evaluate the proposed unmixing procedure and is depicted in Fig. 2 (a). The scene is composed mainly of roof, road and grass pixels, resulting in $R = 3$ endmembers. The spectral signatures of these components have been extracted from the data using the N-FINDR algorithm [24].

Different mixing models and estimation algorithms have been tested on this real image: 1) The FCLS algorithm [1] which is known to have good performance for linear mixtures, 2) The GBM-based approach [25] which is particularly adapted for bilinear nonlinearities, 3) The gradient-based approach of [9] which is based on a PPNMM and has shown nice properties for various nonlinear models, 4) The proposed RCA-SU algorithm which is based on the model (1). Note that the RCA-SU algorithm has been run with $N_{\text{MC}} = 3000$, $N_{\text{bi}} = 2000$, $K = 4$ and $\beta = 1.2$. This value of β was selected from preliminary runs but could have been estimated, as in [17]. Finally, we consider 5) the K-Hype method [8] to compare our algorithm with a state-of-the art kernel based unmixing method. The kernel used in this paper is the polynomial, second order symmetric kernel whose Gram matrix is defined by (5). This kernel provides better performance on this data set than the kernels studied in [8]. All hyperparameters of the K-Hype algorithm have been optimized using preliminary runs.

The abundance maps estimated by the RCA-SU algorithm and available in [15] are graphically in good agreement with the state-of-the art algorithms. However, Table 1 shows that K-Hype and the proposed algorithm provide a lower reconstruction error (RE) defined by $\text{RE} = \sqrt{(\sum_n \|\hat{\mathbf{y}}_n - \mathbf{y}_n\|^2)/(NL)}$, where $\mathbf{y}_n$ is the nth observation vector and $\hat{\mathbf{y}}_n$ its estimate. Since no ground truth is available for this data set, it is difficult to evaluate whether the K-Hype or the proposed algorithm provides the best abundance estimates. However, the simulation results from synthetic data [15] show that the RCA-SU generally provides better results. Fig. 1 compares the noise variances estimated by the RCA-SU for the real image with the noise variances estimated by the HySime algorithm [26]. The HySime algorithm assumes additive noise and estimates the noise covariance matrix of the image using multiple regression. Fig. 1 shows that the two algorithms provide similar noise variance estimates. These results motivate the consideration of non i.i.d. noise for hyperspectral image analysis since the noise variances increase for the highest wavelengths.

Table 1. REs($\times 10^{-2}$).

FCLS	GBM	PPNMM	K-HYPE	RCA-SU
0.65	0.65	0.54	**0.48**	**0.48**

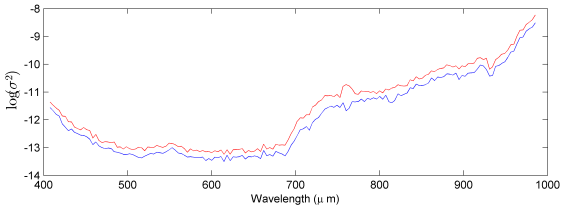
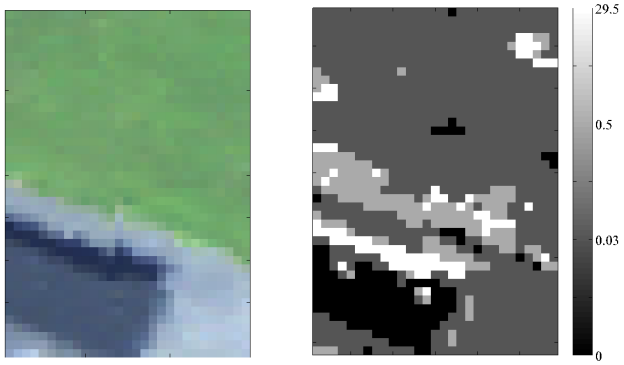


Fig. 1. Noise variances estimated by the RCA-SU (red) and the Hysime algorithm (blue) for the real Madonna image.

Fig. 2 (b) shows the detection map (map of z_n for $n = 1, \ldots, N$) provided by the proposed RCA-SU detector for the real image considered. Since no ground truth is available for this data set, in contrast to the synthetic data considered in [15], it is difficult to quantitatively assess the performance of the proposed nonlinearity detector. However, the detection map highlights structures that can also be identified in the true color image of the scene (Fig. 2 (a)). Due to the consideration of spatial structures, the proposed detector provides homogeneous regions. The estimated class $\mathcal{C}_0$ (black pixels) associated with linearly mixed pixels is mainly located in the roof region. The class $\mathcal{C}_1$ (dark grey pixels) can be related to regions where the main components in the pixels are grass or road. Mixed pixels composed of grass and road are gathered in class $\mathcal{C}_2$ (light grey pixels). Finally, shadowed pixels located between the roof and the road are associated with the last class $\mathcal{C}_3$ (white pixels). Moreover, the RCA-SU algorithm can identify three levels of nonlinearity, corresponding to $[\hat{s}_1^2, \hat{s}_2^2, \hat{s}_3^2] = [0.03, 0.50, 29.5]$. The most influent nonlinearity class is class $\mathcal{C}_3$, where shadowing effects occurs. Mixed pixels of class $\mathcal{C}_2$ contain weaker nonlinearities. Finally, the remaining pixels of class $\mathcal{C}_1$ are associated with the weakest nonlinearities. The nonlinearities of this class can probably be explained by the endmember variability and/or the endmember estimation error. It is interesting to note that the RCA-SU algorithm identifies two rather linear classes associated with homogeneous regions mainly composed of a single parameter (classes $\mathcal{C}_0$ and $\mathcal{C}_1$). The two other classes ($\mathcal{C}_2$ and $\mathcal{C}_3$) correspond to rather nonlinear regions where the pixels are mixed and shadowing effects occur.



(a) (b)

Fig. 2. (a) True color image of the scene of interest. (b) Nonlinearity detection map obtained with the RCA-SU detector for the Madonna image.

7. CONCLUSION

We have proposed a new hierarchical Bayesian algorithm for joint linear/nonlinear spectral unmixing of hyperspectral images and nonlinearity detection. The nonlinear mixtures were decomposed into a linear combination of the endmembers and an additive term representing the nonlinear effects. A Markov random field was introduced to promote spatial structures in the image. The image was decomposed into regions or classes where the nonlinearities share the same statistical properties defined by Gaussian processes, each class being associated with a level of nonlinearity. An important advantage of the proposed algorithm with respect to other strategies is the possibility of detecting several kinds of linearly and nonlinearly mixed pixels. This detection can be used to identify the image regions affected by different nonlinearities in order to characterize the nonlinear effects more deeply.

The performance of the algorithm relies on the endmember knowledge (endmembers assumed to be known in this paper). Estimating the pure component spectra present in the image, jointly with the abundance estimation and the nonlinearity detection is an important issue that should be considered in future work. Finally, estimating the number of classes and the granularity of the scene is clearly a challenging issue that we plan to investigate.

8. 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*, 2014, to appear.

[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. Van 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. 2009, vol. 7477, p. 74770I, SPIE.

[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, A. Halimi, N. Dobigeon, and J.-Y. Tourneret, "Supervised nonlinear spectral unmixing using a postnonlinear mixing model for hyperspectral imagery," *IEEE Trans. Image Processing*, vol. 21, no. 6, pp. 3017–3025, June 2012.

[10] Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Nonlinearity detection in hyperspectral images using a polynomial post-nonlinear mixing model," *IEEE Trans. Image Processing*, vol. 22, no. 4, pp. 1267–1276, April 2013.

[11] 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.

[12] A. Kalaitzis and N. D. Lawrence, "Residual components analysis," in *Proc. Int. Conf. Mach. Learning (ICML)*, 2012.

[13] 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.

[14] N. Dobigeon and C. Févotte, "Robust nonnegative matrix factorization for nonlinear unmixing of hyperspectral images," in *Proc. IEEE GRSS Workshop Hyperspectral Image SIgnal Process.: Evolution in Remote Sens. (WHISPERS)*, 2013.

[15] Y. Altmann, N. Dobigeon, S. McLaughlin, and J.-Y. Tourneret, "Residual component analysis of hyperspectral images - application to joint nonlinear unmixing and nonlinearity detection," Tech. Rep., University of Toulouse, France, Sept. 2013.

[16] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory*, John Wiley & Sons, New York, 1994.

[17] 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.

[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] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer-Verlag, New York, second edition, 2004.

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

[21] C. P. Robert and D. Cellier, "Convergence control of MCMC algorithms," in *Discretization and MCMC Convergence Assessment*, C. P. Robert, Ed., pp. 27–46. Springer Verlag, New York, 1998.

[22] 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.

[23] 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.

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

[25] A. Halimi, Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Unmixing hyperspectral images using a generalized bilinear model," in *Proc. IEEE Int. Conf. Geosci. and Remote Sensing (IGARSS)*, July 2011, pp. 1886 –1889.

[26] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. and Remote Sensing*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.