Robust hyperspectral unmixing accounting for residual components

Abderrahim Halimi, Paul Honeine, José Bioucas-Dias

ROBUST HYPERSPECTRAL UNMIXING ACCOUNTING FOR RESIDUAL COMPONENTS

Abderrahim Halimi [1], *Paul Honeine* [2], *Jose Bioucas-Dias* [3]

[1] Heriot-Watt University, School of Engineering and Physical Sciences, Edinburgh, U.K
[2] LITIS lab, Université de Rouen, France
[3] Instituto de Telecomunicacões and Instituto Superior Técnico, Universidade de Lisboa, Portugal.

ABSTRACT

This paper presents a new hyperspectral mixture model jointly with a Bayesian algorithm for supervised hyperspectral unmixing. Based on the residual component analysis model, the proposed formulation assumes the linear model to be corrupted by an additive term that accounts for mismodelling effects (ME). The ME formulation takes into account the effect of outliers, the propagated errors in the signal processing chain and copes with some types of endmember variability (EV) or nonlinearity (NL). The known constraints on the model parameters are modeled via suitable priors. The resulting posterior distribution is optimized using a coordinate descent algorithm which allows us to compute the maximum a posteriori estimator of the unknown model parameters. The proposed model and estimation algorithm are validated on both synthetic and real images showing competitive results regarding the quality of the inferences and the computational complexity when compared to the state-of-the-art algorithms.

Index Terms— Hyperspectral imagery, robust unmixing, Bayesian estimation, coordinate descent algorithm, Gaussian process, gamma Markov random field

1. INTRODUCTION

Spectral unmixing (SU) is a source separation problem consisting of recovering the spectral signatures (endmembers) of the materials present in the scene, and quantifying their proportions (abundances) within each hyperspectral image pixel. The linear mixture model (LMM) is the widely used model for SU mainly because of its simplicity. However, this model can be inappropriate for some hyperspectral scenarios, namely in presence of nonlinearity (such as multiple scattering or intimate mixtures), endmember variability, or outliers. In addition, the estimated LMM abundances can be badly affected by the propagated errors in the signal processing chain such as a bad estimation of the endmembers or their number (especially under a supervised unmixing, i.e., estimating only the abundances). These effects emphasize the need for a robust hyperspectral unmixing strategy to deal with these mismodelling effects [1, 2].

This paper introduces a residual component (RC) mixture model for robust hyperspectral unmixing, the model generalizes the well-known LMM by accounting for the illumination variation and the presence of a residual term. This term aims at reducing the effect of smooth spatial/spectral outliers that can be due to the presence of NL [3, 4], EV [5] or to signal processing chain errors. Estimating the abundances associated with this mixture model is a challenging problem. We propose here a hierarchical Bayesian model to estimate the parameters and hyperparameters associated with the RC model. This hierarchical model introduces prior distributions that enforce known physical constraints on the estimated parameters such as the sum-to-one and positivity of the abundances, and the smooth spectral behavior of the ME. Moreover, the spatial correlation of the residual term has been introduced by considering Markov random fields (MRF) [6]. Using the likelihood and the considered prior distributions, the joint posterior distribution of the unknown parameter vector is then derived. The minimum mean square error (MMSE) and maximum a posteriori (MAP) estimators of these parameters cannot be easily computed from the obtained joint posteriors. In this paper, the MAP estimator is evaluated by considering a coordinate descent algorithm (CDA) [7–9] that sequentially updates the abundances, the noise variances and the residual term. This provides a reduced computational cost when compared to the Markov chain Monte-Carlo based algorithms [10–12]. The proposed Bayesian model and estimation algorithm are validated using synthetic and real hyperspectral images. The obtained results are very promising and show the potential of the proposed mixture and Bayesian models and their associated inference algorithm.

The paper is structured as follows. Section 2 introduces the proposed mixture model to deal with the ME. The proposed hierarchical Bayesian model and its estimation algorithm are introduced in Sections 3 and 4. Section 5 is devoted to testing and validating the proposed techniques using synthetic images with known ground truth. Section 6 shows results obtained using a real hyperspectral image. Conclusions and future work are finally reported in Section 7.

2. MIXTURE MODEL

The LMM is widely used because of its simplicity. However, there are a lot of situations where the linear model is not valid because of the presence of EV, NL or other ME (due to physical outliers or signal processing errors). This paper proposes a new formulation based on a residual component model [13] that is expressed as the sum of a linear model and a residual term. The general observation model for the $(L \times 1)$ pixel spectrum $\boldsymbol{y}_{i,j}$ is given by

$$\boldsymbol{y}_{i,j} \quad = \quad \boldsymbol{S}_{i,j}\boldsymbol{a}_{i,j} + \boldsymbol{\phi}_{i,j}\left(\boldsymbol{S}_{i,j}, \boldsymbol{a}_{i,j}\right) + \boldsymbol{e}_{i,j}, \qquad (1)$$

where $\boldsymbol{a}_{i,j} = (a_{1,i,j}, \cdots, a_{R,i,j})^T$ is an $(R \times 1)$ vector of abundances associated with the pixel (i, j) and satisfying the positivity and sum-to-one constraints ($a_{r,i,j} \geq 0, \forall r \in \{1, \ldots, R\}$ and $\sum_{r=1}^{R} a_{r,i,j} = 1$), R is the number of endmembers, $\boldsymbol{e}_{i,j} \sim \mathcal{N}(\boldsymbol{0}_L, \boldsymbol{\Sigma})$ is an additive centered Gaussian noise with a diagonal covariance matrix $\boldsymbol{\Sigma} = \text{diag}\left(\boldsymbol{\sigma}^2\right)$, $\boldsymbol{\sigma}^2 = \left(\sigma_1^2, \cdots, \sigma_L^2\right)^T$ is an $(L \times 1)$ vector containing the noise variances, L is the number of spectral bands, $\boldsymbol{S}_{i,j}(\boldsymbol{M}) = \boldsymbol{S}_{i,j}$ is the endmember matrix that depends on each pixel to introduce EV effect, $\boldsymbol{M}$ is a fixed endmember

This work was supported in part by the HYPANEMA ANR Project under Grant ANR-12-BS03-003, and in part by the Portuguese Fundacão para a Ciência e Tecnologia (FCT), grant UID/EEA/5008/2013

matrix that is assumed known (extracted using an endmember extraction algorithm) and $\boldsymbol{\phi}_{i,j}\left(\boldsymbol{S}_{i,j}, \boldsymbol{a}_{i,j}\right)$ is a residual term that might depend on the endmembers or the abundances to model NL or EV effects.

This paper accounts for mismodelling effects or the presence of outliers by considering a residual term $\boldsymbol{\phi}_{i,j}^{ME} = \boldsymbol{d}_{i,j}$ that shows spatial and spectral correlations. Moreover, it accounts for the illumination variation effect by considering $\boldsymbol{S}_{i,j} = c_{i,j}\boldsymbol{M}$. Under these considerations, the observation model (1) reduces to

$$\boldsymbol{y}_{i,j} = c_{i,j}\boldsymbol{M}\boldsymbol{a}_{i,j} + \boldsymbol{d}_{i,j} + \boldsymbol{e}_{i,j}, \tag{2}$$

where $\boldsymbol{d}_{i,j}$ is an $L \times 1$ smooth spectral vector. This term gathers the shape EV effects that can be approximated by a smooth function as already highlighted in [14]. It also gathers the multiple scattering NL effects when the endmembers are themselves smooth. We refer the reader to [14] for more details about the relation between (2) and other NL and EV models, and the possible interpretations of $\boldsymbol{d}_{i,j}$. Note that model (2) reduces to the LMM when $\boldsymbol{d}_{i,j} = \boldsymbol{0}_L$, and $c_{i,j} = 1, \forall i, j$. Note also that other models have been introduced in the literature to account for the effect of outliers such as [1] which proposed spatial/spectral correlated outliers by considering discrete MRF and [2] which proposed positive-sparse outliers that have no spatial or spectral correlation. In contrast to [10, 11] that deal only with NL effect, and [12] that deals with EV, model (2) considers both effects at the same time.

3. HIERARCHICAL BAYESIAN MODEL

This section introduces a hierarchical Bayesian model to estimate the unknown parameter vector $\boldsymbol{\Theta} = \left(\boldsymbol{A}, \boldsymbol{c}, \boldsymbol{D}, \boldsymbol{\sigma}^2, \boldsymbol{\epsilon}\right)$, where $\boldsymbol{A}$ (resp. $\boldsymbol{D}$) gathers all the abundance vectors (resp. residual terms), and $\boldsymbol{\epsilon}$ is a hyperparameter described later in the text.

3.1. Likelihood

Using the observation model (1), the Gaussian properties of the noise sequence $\boldsymbol{e}_{i,j}$, and exploiting independence between the observations in different spectral bands, yield the following Gaussian distribution for the likelihood

$$\boldsymbol{y}_{i,j}|\boldsymbol{a}_{i,j}, c_{i,j}, \boldsymbol{d}_{i,j}, \boldsymbol{\Sigma} \sim \mathcal{N}\left(\boldsymbol{\mu}_{i,j}, \boldsymbol{\Sigma}\right). \tag{3}$$

where $\boldsymbol{\mu}_{i,j} = c_{i,j}\boldsymbol{M}\boldsymbol{a}_{i,j} + \boldsymbol{d}_{i,j}$ and $\sim$ means "is distributed according to". The independence between the observed pixels leads to $f(\boldsymbol{Y}|\boldsymbol{\Theta}) = \prod_i \prod_j f(\boldsymbol{y}_{i,j}|\boldsymbol{a}_{i,j}, c_{i,j}, \boldsymbol{d}_{i,j}, \boldsymbol{\Sigma})$, where $\boldsymbol{Y}$ gather all the pixels.

3.2. Parameter priors

This section introduces the prior distributions that we have chosen for the parameters of interest $\boldsymbol{A}$, $\boldsymbol{c}$, $\boldsymbol{D}$ and $\boldsymbol{\Sigma}$.

3.2.1. Abundance matrix $\boldsymbol{A}$

Due to the positivity and sum-to-one (PSTO) constraints, the abundance vector should live in the following simplex

$$\mathcal{S} = \left\{ \boldsymbol{a}_{i,j} \middle| a_{r,i,j} \geq 0, \forall r \text{ and } \sum_{r=1}^{R} a_{r,i,j} = 1 \right\}. \tag{4}$$

Since there is no additional information about this parameter vector, we propose to assign a uniform prior in the simplex $\mathcal{S}$ to the abundances [15, 16].

3.2.2. Prior for $\boldsymbol{c}$

The variation in illumination is introduced by the parameter $c_{i,j}$ that is pixel-dependent. In many works, this parameter has been fixed to the value $\#1$, which represents the absence of illumination variability [16–18]. In others the sum-to-one (STO) was relaxed leading to an unconstrained parameter $c_{i,j}$ (free in $\mathbb{R}$). In this paper, we relax the STO constraint but include some prior knowledge on $c_{i,j}$. Indeed, we allow this parameter to fluctuate around $\#1$ by considering a conjugate Gaussian prior as follows

$$c_{i,j} \sim \mathcal{N}\left(1, \eta^2\right), \tag{5}$$

where η^2 is a small fixed variance ($\eta^2 = 0.01$ in the rest of the paper). For simplicity, we denote "$x|\theta \sim ...$", by "$x \sim ...$" when the parameter θ is a user fixed parameter. Note finally that the joint prior of $\boldsymbol{c}$ is obtained by assuming a priori independence between the coefficients $c_{i,j}$, as follows $f\left(\boldsymbol{c}\right) = \prod_{i,j} f\left(c_{i,j}\right)$.

3.2.3. Mismodelling coefficients $\boldsymbol{d}_{i,j}$

The smooth property of the vector $\boldsymbol{d}_{i,j}$ is introduced by considering a conjugate Gaussian prior as follows

$$\boldsymbol{d}_{i,j}|\epsilon_{i,j}^2 \sim \mathcal{N}\left(\boldsymbol{0}_L, \epsilon_{i,j}^2\boldsymbol{H}\right), \tag{6}$$

where $\boldsymbol{H}$ is the squared-exponential covariance function given by $H(\ell, \ell') = \exp\left[-\frac{\left(\ell-\ell'\right)^2}{(L/2)^2}\right]$, which introduces the spectral smoothness on $\boldsymbol{d}_{i,j}$. A spatial correlation is also introduced by enforcing a smooth variation of the residual energies $\left(\boldsymbol{d}_{i,j}^T\boldsymbol{H}^{-1}\boldsymbol{d}_{i,j}\right)$. This is achieved by considering a specific prior for $\epsilon_{i,j}^2$, as explained in Section 3.3. Finally the joint prior of $\boldsymbol{D}$ is obtained by assuming a priori independence between the mismodelling coefficients, i.e., $f\left(\boldsymbol{D}|\boldsymbol{\epsilon}\right) = \prod_{i,j} f\left(\boldsymbol{d}_{i,j}|\epsilon_{i,j}^2\right)$.

3.2.4. Noise variances

The noise variances are assigned a conjugate inverse gamma distribution given by

$$f\left(\boldsymbol{\sigma}^2\right) = \prod_{\ell=1}^{L} f\left(\sigma_\ell^2\right), \quad \text{with } \sigma_\ell^2 \sim \mathcal{IG}\left(\varphi_\ell, \psi_\ell\right), \tag{7}$$

where σ_ℓ^2 are assumed a priori independent. The hyperparameters φ_ℓ and ψ_ℓ are fixed to approximate the HySime estimated variances [19]. Note finally that the parameters can also be set to $\varphi_\ell = \psi_\ell = 0$ in absence of prior knowledge about σ_ℓ^2, leading to a noninformative Jeffreys' prior.

3.3. Hyperparameter priors

Due to the spatial organization of hyperspectral images, we expect the energies of the mismodelling coefficients $\boldsymbol{d}_{i,j}$ to vary smoothly from one pixel to another. This behavior is obtained by introducing an auxiliary variable $\boldsymbol{w}$ (of size $N_{\text{row}} \times N_{\text{col}}$) and assigning a gamma Markov random field prior for $\left(\boldsymbol{\epsilon}, \boldsymbol{w}\right)$ (see [6, 10] for more details regarding this prior). This prior ensures that each $\epsilon_{i,j}^2$ is connected to four neighbor elements of $\boldsymbol{w}$ and vice-versa (see Fig. 1). The energies $\epsilon_{i,j}^2$ are conditionally independent and the 1st order neighbors (i.e., the spatial correlation) is only introduced via the auxiliary variables $\boldsymbol{w}$. An interesting property of this joint prior is that the

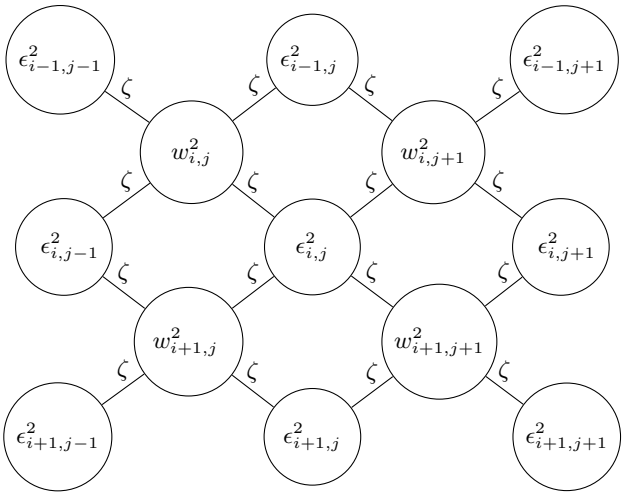

Fig. 1. Proposed 1st order gamma-MRF neighborhood structure.

conditional prior distributions of $\boldsymbol{\epsilon}$ and $\boldsymbol{w}$ reduce to conjugate inverse gamma ($\mathcal{IG}$) and gamma ($\mathcal{G}$) distributions as follows

$$\begin{aligned} \epsilon^2_{i,j}|\boldsymbol{w},\zeta &\sim \mathcal{IG}\left(4\zeta, 4\zeta\rho_{1,i,j}(\boldsymbol{w})\right), \\ w^2_{i,j}|\boldsymbol{\epsilon},\zeta &\sim \mathcal{G}\left(4\zeta, 1/(4\zeta\rho_{2,i,j}(\boldsymbol{\epsilon}))\right), \end{aligned} \tag{8}$$

where

$$\begin{aligned} \rho_{1,i,j}(\boldsymbol{w}) &= (w^2_{i,j} + w^2_{i+1,j} + w^2_{i,j+1} + w^2_{i+1,j+1})/4, \\ \rho_{2,i,j}(\boldsymbol{\epsilon}) &= (\epsilon^{-2}_{i,j} + \epsilon^{-2}_{i-1,j} + \epsilon^{-2}_{i,j-1} + \epsilon^{-2}_{i-1,j-1})/4, \end{aligned} \tag{9}$$

and ζ is a fixed coupling parameter that controls the amount of spatial smoothness enforced by the gamma-MRF.

3.4. Posterior distributions

The joint posterior distribution of the hierarchical Bayesian model can be computed from the following Bayes' rule

$$f\left(\boldsymbol{\Theta}|\boldsymbol{Y}\right) \propto f(\boldsymbol{Y}|\boldsymbol{\Theta}) f\left(\boldsymbol{A}\right) f\left(\boldsymbol{\sigma}^2\right) f\left(\boldsymbol{c}\right) f\left(\boldsymbol{D}|\boldsymbol{\epsilon}\right) f\left(\boldsymbol{\epsilon},\boldsymbol{w}\right), \tag{10}$$

where $\propto$ means "proportional to" and we have assumed a priori independence between the parameters of each model. The MMSE and MAP estimators associated with the posterior (10) are not easy to determine. In this paper, and akin to [9], we propose to evaluate the MAP estimator by using an optimization technique maximizing the posterior (10) w.r.t. the parameters of interest.

4. COORDINATE DESCENT ALGORITHM

Because of the large number of parameters to estimate, we propose a coordinate descent algorithm [7–9] that sequentially updates the different parameters. In each step of the proposed CDA-ME algorithm, the posterior distribution is maximized w.r.t. one parameter, the other being fixed. Thus, the algorithm iteratively updates each parameter by maximizing its conditional distribution as follows (see [14] for more details regarding the conditional distributions):

- Conditional of $\boldsymbol{A}$: truncated Gaussian distribution (whose one maximum is obtained with SUNSAL-FCLS[1] [18])
- Conditional of $\boldsymbol{D}$: Gaussian distribution (analytical expression of the mean)
- Conditional of $\boldsymbol{\epsilon}$: inverse gamma distribution (analytical expression of the maximum)
- Conditional of $\boldsymbol{w}$: gamma distribution (analytical expression of the maximum)
- Conditional of $\boldsymbol{\sigma}^2$: inverse gamma distribution (analytical expression of the maximum)
- Conditional of $\boldsymbol{c}$: Gaussian distribution (analytical expression of the mean)

Regarding the sequence generated by the coordinate descent algorithm, the proposition 2.7.1 in [7] asserts that its limit points are stationary points of (10) provided that the maximum of that function w.r.t. $\boldsymbol{\Theta}$ along each coordinate is unique. This is easily checked for all the parameters since they have unimode conditional distributions (Gaussian, gamma and inverse gamma distributions). Note however that the cost function is not convex, thus, the solution obtained might depend on the initial values that need to be chosen carefully. Therefore, the abundances $\boldsymbol{A}$ are initialized with SUNSAL-FCLS [18], the residual terms are initialized by 0, the noise variance is initialized by HySime [19], the illumination coefficient $\boldsymbol{c}$ is initialized by considering the sum of the abundances that are estimated using only the positivity constraint with SUNSAL-CLS [18]. With these initializations, the proposed algorithm reached minima of "good quality" in the considered simulations (see Sections 5 and 6). Thus, the CDA algorithm constitutes a good balance between computational efficiency (obtained by solving the simple problems associated with each descent step) and the quality of the solution (experimentally observed when considering a good initialization).

5. SIMULATION RESULTS ON SYNTHETIC DATA

Four synthetic images of size 100×100 pixels and $L = 207$ spectral bands were generated using $R = 3$ endmembers extracted from the ENVI software library [20]. All images were corrupted by i.i.d. Gaussian noise for a fair comparison with SU algorithms using this assumption. The images were generated using different mixture models. The first image I_1 is composed of LMM pixels. The second image I_2 is composed of a mixture of the LMM and 3 nonlinear models ($K = 4$ spatial classes were generated using a Potts-MRF with granularity parameter $\beta = 0.8$). The third image I_3 is partitioned into 4 spatial classes (the same partition map as for I_2). In each class, the pixels were generated according to the LMM while using a different set of spectra associated with the same 3 materials, to study the EV effect. The fourth image I_4 is partitioned into $K = 2$ classes using a Potts-MRF. The pixels of the first class were generated according to the RCA-ME model (2) with $\epsilon^2 = 0.002$. The pixels of the second class were generated with 4 endmembers to simulate the effect of a bad estimation of the number of endmembers. The abundances are uniformly distributed in the simplex $\mathcal{S}$ for I_1. However, to simulate a highly mixed scenario, the abundances of I_2, I_3 and I_4 are clustered inside the simplex using a Dirichlet distribution whose parameters are selected randomly in the interval $[1, 20]$. The illumination coefficient c varies in $[0.9, 1.15]$ for all images. Table 1 compares the proposed CDA-ME algorithm with the FCLS [17], SUNSAL-CLS [18], SKHype [21], and RCA-MCMC [11] algorithms. Note that SUNSAL-CLS approximates the variation of illumination by relaxing the sum-to-one constraint. The criteria used are the distance between the actual and estimated abundances $\text{RMSE}\left(\boldsymbol{A}\right) = \sqrt{\frac{1}{NR}\sum_{n=1}^{N}\|\boldsymbol{a}_n - \hat{\boldsymbol{a}}_n\|^2}$, the reconstruc-

[1] SUNSAL-FCLS satisfies the PSTO constraints while SUNSAL-CLS only ensure the positivity constraint.

tion error: RE $= \sqrt{\frac{1}{NL}\sum_{n=1}^{N}\|\hat{\boldsymbol{y}}_n - \boldsymbol{y}_n\|^2}$ and the spectral angle mapper SAM $= \frac{1}{N}\sum_{n=1}^{N}\arccos\left(\frac{\hat{\boldsymbol{y}}_n^T\boldsymbol{y}_n}{\|\boldsymbol{y}_n\|\,\|\hat{\boldsymbol{y}}_n\|}\right)$ criteria, where $\arccos(\cdot)$ is the inverse cosine operator and $\boldsymbol{y}_n$, $\hat{\boldsymbol{y}}_n$ denotes the $\#n$th measured and estimated pixel spectra. CDA-ME provides the best performance for the LMM, EV and ME based images while it shows good performance in presence of NL effects. These results highlight the robustness of CDA-ME to the different effects that can affect hyperspectral images. Finally, apart from the LMM based algorithms (FCLS and SUNSAL-CLS), the proposed CDA-ME always provides the fastest results.

Table 1. Results on synthetic images.

		($\times 10^{-2}$)			Time
		RMSE	RE	SAM	(s)
I_1 (LMM, $K=1$)	FCLS	7.78	3.58	6.24	1.2
	SUNSAL-CLS	3.42	2.27	5.62	0.1
	SKHype	1.41	–	–	541
	RCA-MCMC	4.31	–	–	6737
	CDA-ME	1.35	2.27	5.62	88
I_2 (LMM, NL $K=4$)	FCLS	24.76	15.74	10.64	1.7
	SUNSAL-CLS	16.55	4.17	7.57	0.07
	SKHype	5.87	–	–	547
	RCA-MCMC	5.66	–	–	9009
	CDA-ME	6.61	2.89	6.17	37
I_3 (EV, $K=4$)	FCLS	10.22	2.91	6.0	1.3
	SUNSAL-CLS	10.16	2.53	5.93	0.14
	SKHype	7.69	–	–	623
	RCA-MCMC	13.33	–	–	11025
	CDA-ME	4.29	2.35	5.51	316
I_4 (RCA-ME, $K=2$)	FCLS	12.59	5.01	7.19	1.21
	SUNSAL-CLS	12.65	2.91	6.94	0.12
	SKHype	11.68	–	–	374
	RCA-MCMC	18.20	–	–	6029
	CDA-ME	6.07	2.29	5.56	315

6. RESULTS ON REAL DATA

This section illustrates the performance of CDA-ME when applied to a real hyperspectral image. The considered image, denoted as Madonna, was acquired in 2010 by the Hyspex hyperspectral scanner over Villelongue, France (00 03'W and 4257'N). The dataset contains $L = 160$ spectral bands recorded from the visible to near infrared ($400 - 1000$nm) with a spatial resolution of 0.5m [22]. It has already been studied in [12, 23] and is mainly composed of forested areas (see Fig. 2 (top)). This image contains 100×100 pixels and is composed of $R = 4$ components: tree, grass, soil and shadow. The Bayesian UsLMM algorithm [24] was used to estimate $R = 4$ endmembers. The estimated CDA-ME abundance maps are in good agreement with the state-of-the-art algorithms and are not shown here for brevity. Fig. 2 (middle-right) shows the energies of the difference between the reconstructed signal and the linear model (i.e., $||\hat{\boldsymbol{y}}_{i,j} - \boldsymbol{M}\hat{\boldsymbol{a}}_{i,j}||$) . It is clear from Fig. 2 (middle-right) that the MEs are mainly located in regions with multiple physical components, and in presence of relief (near trees) where multiple scattering effect might occurs. These energies are in good agreement with the nonlinear coefficients estimated using the RCA-MCMC algorithm as shown in Fig. 2 (middle-left). However, in contrast to RCA-MCMC that interprets these energies as nonlinearities, the proposed CDA-ME highlights the presence of an illumination variation (see Fig. 2 (bottom-left)) and an additional residual term (see Fig. 2 (bottom-right)). Note that the residual term gathers NL, EV and outlier effects and is mainly located near the trees. To summarize, this section shows that the CDA-ME algorithm captures similar spatial residual effects than RCA-MCMC but gives a different interpretation. Most of these residuals appear in region of intersection between the physical elements and in presence of vegetation such as trees. CDA-ME is very flexible and can capture different physical effects such as NL, EV and outliers. Note finally that more details regarding the CDA-ME results, including the processing of another real image (the Moffett scene), are provided in [14] and not shown here for brevity.

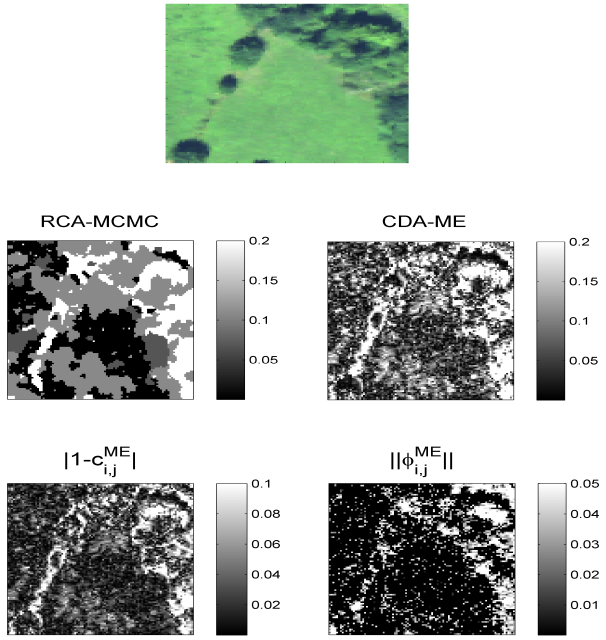

Fig. 2. Top: Real Madonna image. Middle-left: nonlinearities coefficients estimated using the RCA-MCMC algorithm [11]. Middle-right: Square root of the energies of the difference between the reconstructed signal and the linear model for the proposed CDA-ME algorithm ($||\hat{\boldsymbol{y}}_{i,j} - \boldsymbol{M}\hat{\boldsymbol{a}}_{i,j}||$). Bottom-left: estimated illumination variation $|1 - c_{i,j}|$ for the CDA-ME algorithm. Bottom-right: square root of the energies of the residual terms $\|\boldsymbol{d}_{i,j}\|$ for the CDA-ME algorithm.

7. CONCLUSIONS

This paper introduced a hyperspectral mixture model and its associated Bayesian algorithm for robust hyperspectral unmixing. The proposed model was introduced under a general formulation that can be adapted to account for other physical effects. A hierarchical Bayesian model was proposed to introduce the known constraints on their parameters. Those parameters were estimated using a coordinate descent algorithm that showed a reduced computational cost when compared to state-of-the-art algorithms. The proposed algorithm showed good performance when processing synthetic data generated with the linear model or other more sophisticated models. Results on real data confirmed the good performance of the proposed algorithm and showed its ability to extract different features in the observed scenes. Future work includes the detection of the presence of outliers using hypothesis tests.

8. REFERENCES

[1] Y. Altmann, S. McLaughlin, and A. Hero, "Robust linear spectral unmixing using anomaly detection," *IEEE Trans. Comput. Imaging*, vol. 1, no. 2, pp. 74–85, June 2015.

[2] C. Févotte and N. Dobigeon, "Nonlinear hyperspectral unmixing with robust nonnegative matrix factorization," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4810–4819, Dec 2015.

[3] R. Heylen, M. Parente, and P. Gader, "A review of nonlinear hyperspectral unmixing methods," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 7, no. 6, pp. 1844–1868, June 2014.

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

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

[6] O. Dikmen and A. Cemgil, "Gamma markov random fields for audio source modeling," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 3, pp. 589–601, March 2010.

[7] D. P. Bertsekas, *Nonlinear programming*. Belmont, Massachusetts: Athena Scientific, 1995.

[8] J. Sigurdsson, M. Ulfarsson, and J. Sveinsson, "Hyperspectral unmixing with l_q regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 6793–6806, Nov. 2014.

[9] A. Halimi, C. Mailhes, J.-Y. Tourneret, and H. Snoussi, "Bayesian estimation of smooth altimetric parameters: Application to conventional and delay/Doppler altimetry," *IEEE Trans. Geosci. Remote Sens.*, 2015, to appear.

[10] Y. Altmann, M. Pereyra, and S. McLaughlin, "Bayesian nonlinear hyperspectral unmixing with spatial residual component analysis," *IEEE Trans. Comput. Imaging*, vol. 1, no. 3, pp. 174–185, Sept 2015.

[11] 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 Process.*, vol. 23, no. 5, pp. 2148–2158, May 2014.

[12] A. Halimi, N. Dobigeon, and J.-Y. Tourneret, "Unsupervised unmixing of hyperspectral images accounting for endmember variability," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4904–4917, 2015.

[13] A. A. Kalaitzis and N. D. Lawrence, "Residual components analysis," in *Proc. ICML*, 2012, pp. 1–3.

[14] A. Halimi, P. Honeine, and J. M. Bioucas-Dias, "Hyperspectral unmixing in presence of endmember variability, nonlinearity or mismodelling effects," in *ArXiv e-prints*, Nov. 2015.

[15] N. Dobigeon, J.-Y. Tourneret, and C.-I Chang, "Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2684–2695, July 2008.

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

[17] 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. Remote Sens.*, vol. 29, no. 3, pp. 529–545, March 2001.

[18] J. Bioucas-Dias and M. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *Proc. IEEE GRSS Workshop on Hyperspectral Image and SIgnal Processing: Evolution in Remote Sensing (WHISPERS)*, June 2010, pp. 1–4.

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

[20] RSI (Research Systems Inc.), *ENVI User's guide Version 4.0*, Boulder, CO 80301 USA, Sept. 2003.

[21] 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, Jan 2013.

[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. Remote Sens. (IGARSS)*, July 2011, pp. 3672–3675.

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

[24] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A. O. Hero, "Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4355–4368, Nov. 2009.