Supervised nonlinear spectral unmixing using a post-nonlinear mixing model for hyperspectral imagery

Yoann Altmann, Abderrahim Halimi, Nicolas Dobigeon, Jean-Yves Tourneret

Supervised nonlinear spectral unmixing using a post-nonlinear mixing model for hyperspectral imagery

Yoann Altmann, Abderrahim Halimi, Nicolas Dobigeon and Jean-Yves Tourneret
E-mail : {Yoann.Altmann, Abderrahim.halimi, Nicolas.Dobigeon, Jean-Yves.Tourneret}@enseeiht.fr

TECHNICAL REPORT – 2011, November
University of Toulouse, IRIT/INP-ENSEEIHT
2 rue Camichel, BP 7122, 31071 Toulouse cedex 7, France

Abstract

This paper presents a nonlinear mixing model for hyperspectral image unmixing. The proposed model assumes that the pixel reflectances are nonlinear functions of pure spectral components contaminated by an additive white Gaussian noise. These nonlinear functions are approximated using polynomial functions leading to a polynomial post-nonlinear mixing model. A Bayesian algorithm and optimization methods are proposed to estimate the parameters involved in the model. The performance of the unmixing strategies is evaluated thanks to simulations conducted on synthetic and real data.

Index Terms

Hyperspectral imagery, spectral unmixing, post-nonlinear model.

I. Introduction

Spectral unmixing (SU) is one of the major issues when analyzing hyperspectral images. SU consists of identifying the macroscopic materials present in an hyperspectral image and quantifying the proportions of these materials in all the image pixels. Most SU strategies assume that pixel reflectances are linear combinations of pure component spectra [1]–[5]. The resulting linear mixing model (LMM) has been widely used in the literature and has provided interesting results. However, as explained in [6], the LMM can be inappropriate for some hyperspectral

images, such as those containing sand, trees or vegetation areas. Nonlinear mixing models provide an interesting alternative for overcoming the inherent limitations of the LMM. They have been proposed in the hyperspectral image literature for specific kinds of nonlinearity. More precisely, the bidirectional reflectance-based model proposed in [7] has been introduced for hyperspectral images including intimate mixtures. Conversely, the bilinear models recently studied in [8]–[11] address the problem of scattering effects, mainly observed in vegetation areas. Other more flexible unmixing techniques have been also proposed to handle wider class of nonlinearity, including radial basis function networks [12], [13] and kernel-based models [14], [15]. This paper considers a class of nonlinear mixing models referred to as *post-nonlinear mixing models* (PNMMs). PNMMs are flexible generalizations of the standard LMMs that have been introduced in [16], [17] for source separation problems. The main advantage of PNMMs is that they can accurately model many different nonlinearities (as will be shown in this paper). This paper addresses the problem of supervised SU of hyperspectral images using PNMMs. Note that "supervised" means that the endmembers contained in the image have been estimated by an endmember extraction algorithm (EEA). As a consequence, the only parameters to be estimated are the abundances and the nonlinearity coefficients for all pixels of the image. In the last decades, many EEAs have been developed to identify the pure spectral components contained in a hyperspectral image (the reader is invited to consult [18] for a recent review of these methods). Most EEAs implicitly rely on the LMM and might be inappropriate for nonlinear models such as PNMMs. However, as noticed in [6], geometric EEAs are still adapted to identify endmembers and can be reasonably employed when the mixing model involves nonlinearities. Therefore, this paper proposes to extract the endmembers contained in the hyperspectral image using a geometric EEA, known as *vertex component analysis* (VCA) [19] and the recent nonlinear EEA proposed in [20]. Once the endmembers have been extracted from the image, we propose to estimate the abundances and the nonlinearity parameters involved in the PNMM using estimation algorithms based on Bayesian and least-squares (LS) methods.

In the Bayesian framework, appropriate prior distributions are chosen for the unknown PNMM parameters. The joint posterior distribution of these parameters is then derived. However, the classical Bayesian estimators cannot be easily computed from this joint posterior. To alleviate this problem, a Markov chain Monte Carlo (MCMC) method is used to generate samples according to the posterior of interest. As in any Bayesian algorithm, the joint posterior distribution

can also be used to compute confidence intervals for the parameter estimates. However, the resulting computational complexity can be too heavy for practical applications. In order to reduce this computational complexity, we propose to study LS methods that have already received considerable attention in the hyperspectral imagery [2], [10], [14]. A first method based on Taylor series expansions is proposed to iteratively solve the LS criterion associated with the PNMM observation model. The Taylor approximations allow quadratic optimization problems to be solve at each iteration. A second approach is based on a classical gradient method dedicated to constrained problems.

The paper is organized as follows. Section II introduces the PNMM for hyperspectral image analysis. Section III presents a Bayesian unmixing algorithm associated with the proposed PNMM. Section IV studies the two alternative unmixing algorithms based on least squares methods. Some simulation results conducted on synthetic and real data are shown and discussed in Section V. Conclusions are finally reported in Section VI.

II. Polynomial Post-Nonlinear Mixing Model

This section defines the nonlinear mixing model used for hyperspectral image SU. More precisely, the L-spectrum $\mathbf{y} = [y_1, \ldots, y_L]^T$ of a mixed pixel is defined as a nonlinear transformation $\boldsymbol{g}$ of a linear mixture of R spectra $\mathbf{m}_r$ contaminated by additive noise

$$\mathbf{y} = \boldsymbol{g}\left(\sum_{r=1}^{R} a_r \mathbf{m}_r\right) + \mathbf{n} = \boldsymbol{g}\left(\mathbf{M}\mathbf{a}\right) + \mathbf{n} \tag{1}$$

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 is its corresponding proportion, R is the number of endmembers contained in the image and $\boldsymbol{g}$ is an appropriate nonlinear function. Moreover, L is the number of spectral bands and $\mathbf{n}$ is an additive independent and identically distributed (i.i.d) zero-mean Gaussian noise sequence with variance σ^2, denoted as $\mathbf{n} \sim \mathcal{N}\left(\mathbf{0}_L, \sigma^2 \mathbf{I}_L\right)$, where $\mathbf{I}_L$ is the $L \times L$ identity matrix. Note that the usual matrix and vector notations $\mathbf{M} = [\mathbf{m}_1, \ldots, \mathbf{m}_R]$ and $\mathbf{a} = [a_1, \ldots, a_R]^T$ have been used in the right hand side of (1).

The choice of the nonlinearity $\boldsymbol{g}$ deserves a specific attention. Polynomials, sigmoidal functions and combinations of polynomial and sigmoidal nonlinearities have shown interesting properties for source separation [17]. This study focuses on second order polynomial nonlinearities $\boldsymbol{g}_b$

defined by

$$\begin{aligned} \boldsymbol{g}_b: \quad [0,1]^L &\rightarrow \mathbb{R}^L \\ \mathbf{s} &\mapsto \left[s_1 + bs_1^2, \ldots, s_L + bs_L^2\right]^T \end{aligned} \tag{2}$$

with $\mathbf{s} = [s_1, \ldots, s_L]^T$. An interesting property of the resulting nonlinear model referred to as polynomial post nonlinear mixing model (PPNMM) is that it reduces to the classical LMM for $b = 0$. Thus, we can expect unmixing results at least as good as those presented in [21] and [2] where Bayesian and LS methods were investigated. Another motivation for using the PPNMM is the Weierstrass approximation theorem which states that any continuous function defined on a bounded interval can be uniformly approximated by a polynomial with any desired precision [22, p. 15]. As explained in [9], it is reasonable to consider polynomials with first and second order terms (since higher order terms can generally be neglected) which leads to (2). Higher order terms could be considered in the presence of more than two reflections. However, the resulting interaction spectra are in practice of low amplitude and are hardly distinguishable from the noise. Straightforward computations allow the PPNMM observation vector (for a given pixel of the image) to be expressed as follows

$$\mathbf{y} = \boldsymbol{g}_b(\mathbf{Ma}) + \mathbf{n} = \mathbf{Ma} + b(\mathbf{Ma}) \odot (\mathbf{Ma}) + \mathbf{n} \tag{3}$$

where $\odot$ denotes the Hadamard (term-by-term) product. Note that the resulting PPNMM includes bilinear terms such as those considered in [8]–[11]. However, the nonlinear terms are characterized by a single amplitude parameter b, leading to a less complex model when compared with the models introduced in [8], [9] and[11]. Note that the endmember $\mathbf{m}_r$ (contained in the matrix $\mathbf{M}$) can be obtained from (3) in the noise free case ($\mathbf{n} = \mathbf{0}_L$) by setting $b = 0$ and $\mathbf{a} = [\mathbf{0}_{r-1}, 1, \mathbf{0}_{R-r}]^T$ in (3).

Due to physical considerations, the abundance vector $\mathbf{a}$ satisfy the following positivity and sum-to-one constraints

$$\sum_{r=1}^{R} a_r = 1, \quad a_r \geq 0, \forall r \in \{1, \ldots, R\}. \tag{4}$$

It is straightforward to show that the function $\mathbf{s} \mapsto \boldsymbol{g}_b(\mathbf{s})$ is non-injective for a fixed b. However, the unmixing problem is identifiable since the application

$$\begin{aligned} \boldsymbol{g}: \quad \mathbb{R}^R \times \mathbb{R} &\rightarrow \mathbb{R}^L \\ (\mathbf{a}, b) &\mapsto \mathbf{Ma} + b(\mathbf{Ma}) \odot (\mathbf{Ma}) \end{aligned}$$

is injective under specific conditions related to the pure component spectra (see Appendix A for details).

III. Bayesian estimation

This section generalizes the hierarchical Bayesian model introduced in [21] to the PPNMM. The unknown parameter vector associated with the PPNMM contains the pixel abundances $\mathbf{a}$ (satisfying the constraints (4)), the nonlinearity parameter b and the additive noise variance σ^2. This section summarizes the likelihood and the parameters priors associated with the proposed hierarchical Bayesian PPNMM.

A. Likelihood

Equation (3) shows that $\mathbf{y}|\mathbf{a}, b, \sigma^2$ is distributed according to a Gaussian distribution with mean $\boldsymbol{g}_b(\mathbf{M}\mathbf{a})$ and covariance matrix $\sigma^2\mathbf{I}_L$ (denoted as $\mathbf{y}|\mathbf{a}, b, \sigma^2 \sim \mathcal{N}(\boldsymbol{g}_b(\mathbf{M}\mathbf{a})), \sigma^2\mathbf{I}_L$). As a consequence, the likelihood function of $\mathbf{y}$ can be expressed as

$$f(\mathbf{y}|\mathbf{a}, b, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{L}{2}} \exp\left(-\frac{\|\mathbf{y} - \mathbf{M}\mathbf{a} - b(\mathbf{M}\mathbf{a}) \odot (\mathbf{M}\mathbf{a})\|^2}{2\sigma^2}\right) \tag{5}$$

where $\|\mathbf{x}\| = \sqrt{\mathbf{x}^T\mathbf{x}}$ is the standard ℓ^2 norm.

B. Parameter priors

In order to satisfy the sum-to-one constraint, the abundance vector can be rewritten[1] $\mathbf{a} = [\mathbf{a}_{\backslash R}, a_R]^T$ with $\mathbf{a}_{\backslash R} = [a_1, \ldots, a_{R-1}]^T$ and $a_R = 1 - \sum_{r=1}^{R-1} a_r$. The positivity constraints in (4) impose that $\mathbf{a}_{\backslash R}$ belongs to the following simplex $\mathcal{S}$

$$\mathcal{S} = \left\{ \mathbf{a}_{\backslash R} \middle| a_r \geq 0, \forall r \neq R, \sum_{r=1}^{R-1} a_r \leq 1 \right\}. \tag{6}$$

A uniform prior distribution on $\mathcal{S}$ is chosen for $\mathbf{a}_{\backslash R}$ to reflect the absence of prior knowledge about the abundance vector.

A Jeffreys' prior is chosen for σ^2

$$f(\sigma^2) \propto \frac{1}{\sigma^2}\mathbf{I}_{\mathbb{R}^+}(\sigma^2) \tag{7}$$

[1]Note that the proposed parametrization is chosen for notation simplicity. However, the component to be discarded can be randomly chosen.

which also reflects the absence of knowledge for this parameter (see [23] for details). A conjugate Gaussian prior is finally chosen for the nonlinearity parameter b

$$b \left| \sigma_b^2 \sim \mathcal{N}\left(0, \sigma_b^2\right) \right. . \tag{8}$$

The Gaussian prior is zero-mean since the value of b can be equally likely positive or negative. Moreover, it favors small values of b and is a conjuguate prior for the parameter b which will simplify the computations.

C. Hyperparameter prior

The hyperparameter σ_b^2 is also included within the Bayesian model. A conjugate inverse-Gamma prior is assigned to σ_b^2

$$\sigma_b^2 \sim \mathcal{IG}\left(\gamma, \nu\right) \tag{9}$$

where (γ, ν) are real parameters fixed to obtain a flat prior, reflecting the absence of knowledge about the variance σ_b^2 ((γ, ν) will be set to $(1, 10^{-2})$ in the simulation section). The resulting directed acyclic graph (DAG) is depicted in Fig. 1.

D. Posterior distribution of $\boldsymbol{\theta}$

The joint posterior distribution of $\boldsymbol{\theta} = \left\{\mathbf{a}_{\setminus R}, b, \sigma^2\right\}$ and σ_b^2 can be computed using the following hierarchical structure

$$f(\boldsymbol{\theta}, \sigma_b^2|\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\sigma_b^2) f(\sigma_b^2) \tag{10}$$

where $\propto$ means "proportional to" and $f(\mathbf{y}|\boldsymbol{\theta})$ is defined in (5). By assuming the parameters σ^2, b and $\mathbf{a}_{\setminus R}$ are *a priori* independent, the joint prior distribution of the unknown parameter vector can be expressed as

$$f(\boldsymbol{\theta}|\sigma_b^2) \;\; = \;\; f(\mathbf{a}_{\setminus R}) f(b|\sigma_b^2) f(\sigma^2). \tag{11}$$

The joint posterior distribution $f(\boldsymbol{\theta}, \sigma_b^2|\mathbf{y})$ can then be computed up to a multiplicative constant

$$f(\boldsymbol{\theta}, \sigma_b^2|\mathbf{y}) \propto \frac{1}{\sigma^2} \left(\frac{1}{\sigma_b^2}\right)^{\frac{3}{2}+\gamma} f(\mathbf{y}|\mathbf{a}_{\setminus R}, \sigma^2, b) \exp\left(-\frac{b^2 + 2\nu}{2\sigma_b^2}\right) \mathbf{1}_{\mathcal{S}}(\mathbf{a}_{\setminus R}). \tag{12}$$

Unfortunately, it is difficult to obtain closed form expressions of the standard Bayesian estimators (including the maximum a posteriori (MAP) and the minimum mean square error (MMSE)

estimators) associated with (12). The last part of this section studies a Markov chain Monte Carlo (MCMC) method which can be used to generate samples asymptotically distributed according to (12). These generated samples are then used to compute the MAP or MMSE estimators of the unknown parameters $\boldsymbol{\theta}$ and σ_b^2.

E. Metropolis-within-Gibbs sampler

The principle of the Gibbs sampler is to sample according to the conditional distributions of the posterior of interest [24, Chap. 10]. The conditional distributions associated with the posterior (12) are studied below.

1) $f(a_r|\mathbf{y}, \boldsymbol{a}_{\backslash R,r} b, \sigma^2, \sigma_b^2)$: Straightforward computations lead to

$$f(a_r|\mathbf{y}, b, \mathbf{a}_{\backslash R,r}, \sigma^2, \sigma_b^2) \propto \exp\left(-\frac{\|\mathbf{y} - \mathbf{M}\mathbf{a} - b(\mathbf{M}\mathbf{a}) \odot (\mathbf{M}\mathbf{a}))\|^2}{2\sigma^2}\right) \mathbf{1}_{\mathcal{S}}(\mathbf{a}_{\backslash R}) \tag{13}$$

where $r = 1, \ldots, R-1$ and $\mathbf{a}_{\backslash R,r} = [a_1, \ldots, a_{r-1}, a_{r+1}, \ldots, a_{R-1}]^T$. Since it is not easy to sample according to (13) (mainly because of the indicator function $\mathbf{1}_{\mathcal{S}}(\mathbf{a}_{\backslash R})$), we propose to update the abundance a_r thanks to a Metropolis-Hasting move. More precisely, a new abundance coefficient is proposed following a Gaussian random walk (RW) procedure (the variance of the proposal distribution has been adjusted to obtain an acceptance rate close to 0.5, as recommended in [25, p. 8]). The generated sampler is accepted or rejected with an appropriate probability provided in Algo. 1.

2) $f(b|\mathbf{y}, \boldsymbol{a}_{\backslash R}, \sigma^2, \sigma_b^2)$: Using (5), it can be easily shown that b is distributed according to the following Gaussian distribution

$$b|\mathbf{y}, \mathbf{a}_{\backslash R}, \sigma^2 \sim \mathcal{N}\left(m_b, s_b^2\right) \tag{14}$$

where

$$m_b = \frac{\sigma_b^2 \left(\mathbf{y} - \mathbf{M}\mathbf{a}\right)^T \boldsymbol{h}(\mathbf{a})}{\sigma_b^2 \boldsymbol{h}(\mathbf{a})^T \boldsymbol{h}(\mathbf{a}) + \sigma^2}, \quad s_b^2 = \frac{\sigma_b^2 \sigma^2}{\sigma_b^2 \boldsymbol{h}(\mathbf{a})^T \boldsymbol{h}(\mathbf{a}) + \sigma^2}$$

and $\boldsymbol{h}(\mathbf{a}) = (\mathbf{M}\mathbf{a}) \odot (\mathbf{M}\mathbf{a})$. As a consequence, sampling according to (14) is straightforward.

3) $f(\sigma^2|\mathbf{y}, \boldsymbol{a}_{\backslash R}, b, \sigma_b^2)$: By considering the posterior distribution (12), it can be shown that $\sigma^2|\mathbf{y}, \mathbf{a}_{-k}, b, \sigma_b^2$ is distributed according to the following inverse-gamma distribution

$$\sigma^2|\mathbf{y}, \mathbf{a}_{\backslash R}, b, \sigma_b^2 \sim \mathcal{IG}\left(\frac{L}{2}, \frac{\|\mathbf{y} - \boldsymbol{g}_b(\mathbf{M}\mathbf{a})\|^2}{2}\right) \tag{15}$$

from which it is easy to sample.

4) $f(\sigma_b^2|\mathbf{y}, \boldsymbol{a}_{\setminus R}, b, \sigma^2)$*:* Finally, by looking at the posterior distribution (12), it can be seen that $\sigma_b^2|\mathbf{y}, \mathbf{a}_{\setminus R}, b, \sigma^2$ is distributed according to the following inverse-gamma distribution

$$\sigma_b^2|\mathbf{y}, \mathbf{a}_{\setminus R}, b, \sigma^2 \sim \mathcal{IG}\left(\frac{1}{2} + \gamma, \frac{b^2}{2} + \nu\right). \tag{16}$$

The resulting Metropolis-within-Gibbs sampler used to sample according to the posterior (12) is summarized in Algo. 1.

After generating samples using the procedures detailed above, the MMSE estimator of the unknown parameters can be approximated by computing the empirical averages of these samples, after an appropriate burn-in period[2]. Even if the sampling strategy has been observed to converge very fast, its computational complexity can be heavy for practical applications. The next section studies LS estimators which allow this computational complexity to be significantly reduced.

IV. Least squares methods

LS methods have been used successfully for linear unmixing [2]. The LS method associated with the observation equation (3) consists of minimizing the following criterion

$$J(\mathbf{a}, b) = \frac{1}{2}\left\|\mathbf{y} - \boldsymbol{g}_b(\mathbf{M}\mathbf{a})\right\|^2 = \frac{1}{2}\left\|\mathbf{y} - \mathbf{M}\mathbf{a} - b(\mathbf{M}\mathbf{a}) \odot (\mathbf{M}\mathbf{a})\right\|^2 \tag{17}$$

under the positivity and sum-to-one constraints (4). This optimization problem is not easy to handle mainly because of the constraints (4). However, the cost function $J(\mathbf{a}, b)$ is quadratic with respect to the parameter b. As a consequence, by differentiating $J(\mathbf{a}, b)$ with respect to b, the following closed-form expression for b can be obtained

$$b = \frac{(\mathbf{y} - \mathbf{M}\mathbf{a})^T \boldsymbol{h}(\mathbf{a})}{\boldsymbol{h}(\mathbf{a})^T \boldsymbol{h}(\mathbf{a})} = \beta(\mathbf{a}). \tag{18}$$

After replacing (18) in $J(\mathbf{a}, b)$, the following criterion can be obtained[3]

$$J(\mathbf{a}) = J(\mathbf{a}, \beta(\mathbf{a})) = \frac{1}{2}\left\|\mathbf{y} - \boldsymbol{\phi}(\mathbf{a})\right\|^2 \tag{19}$$

where

$$\boldsymbol{\phi}(\mathbf{a}) = \mathbf{M}\mathbf{a} + \beta(\mathbf{a})(\mathbf{M}\mathbf{a}) \odot (\mathbf{M}\mathbf{a}). \tag{20}$$

[2]The length of the burn-in period has been determined using appropriate convergence diagnoses [25].

[3]For brevity, the same notation J is chosen for the criteria depending on $\mathbf{a}$ and $(\mathbf{a}, b)$.

We introduce below two strategies to compute the optimal abundance vector

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} \; J(\mathbf{a})$$

under the constraints (4). Note that once $\hat{\mathbf{a}}$ has been computed, the nonlinearity parameter b can be estimated as follows

$$\hat{b} = \beta\left(\hat{\mathbf{a}}\right). \tag{21}$$

A. Taylor approximation

Motivated by the method introduced in [10], we propose to approximate the function $\boldsymbol{\phi}(\cdot)$ defined in (20) using a Taylor series expansion where only first-order terms are considered. Let $\mathbf{a}^{(t)}$ denotes the estimated abundance vector estimate at the tth step of the proposed iterative algorithm, and its corresponding estimated spectrum $\boldsymbol{\phi}(\mathbf{a}^{(t)})$ following (20). The Taylor approximation of $\boldsymbol{\phi}(\cdot)$ at $\mathbf{a}^{(t)}$ can be written

$$\boldsymbol{\phi}(\mathbf{a}) \approx \boldsymbol{\phi}\left(\mathbf{a}^{(t)}\right) + \nabla\boldsymbol{\phi}\left(\mathbf{a}^{(t)}\right)\left(\mathbf{a} - \mathbf{a}^{(t)}\right) \tag{22}$$

where $\nabla\boldsymbol{\phi}(\mathbf{a}^{(t)})$ is the gradient matrix of $\boldsymbol{\phi}(\mathbf{a}^{(t)})$ of size $L \times R$ and $\mathbf{a}$ is the unknown parameter vector to be estimated. The rth column of the gradient matrix $\nabla\boldsymbol{\phi}(\mathbf{a}^{(t)})$ can be derived from (3)

$$\frac{\partial\boldsymbol{\phi}(\mathbf{a})}{\partial a_r} = \mathbf{m}_r + \frac{\partial\beta(\mathbf{a})}{\partial a_r}\boldsymbol{h}(\mathbf{a}) + \beta(\mathbf{a})\frac{\partial\boldsymbol{h}(\mathbf{a})}{\partial a_r} \tag{23}$$

where $r = 1, \ldots, R$ and the partial derivatives of $\beta(\cdot)$ and $\boldsymbol{h}(\cdot)$ are available in Appendix B. Approximating $\boldsymbol{\phi}(\cdot)$ in (19) using (22), the vector $\mathbf{a}^{(t+1)}$ can then be estimated by solving the following constrained LS problem

$$\mathbf{a}^{(t+1)} = \arg\min_{\mathbf{a}} \left\| \boldsymbol{z}^{(t)} - \tilde{\mathbf{M}}^{(t)}\mathbf{a} \right\|^2, \tag{24}$$

under the constraints (4), where

$$\boldsymbol{z}^{(t)} = \mathbf{y} - \boldsymbol{\phi}\left(\mathbf{a}^{(t)}\right) + \nabla\boldsymbol{\phi}\left(\mathbf{a}^{(t)}\right)\mathbf{a}^{(t)} \tag{25}$$

and $\tilde{\mathbf{M}}^{(t)} = \nabla\boldsymbol{\phi}\left(\mathbf{a}^{(t)}\right)$ is the $L \times R$ gradient matrix. Problem (24) can finally be solved by the FCLS algorithm [2]. More precisely, the sum-to-one constraint of the abundances is considered by penalizing (24), leading to

$$\mathbf{a}^{(t+1)} = \arg\min_{\mathbf{a}} \left[\left\| \boldsymbol{z}^{(t)} - \tilde{\mathbf{M}}^{(t)}\mathbf{a} \right\|^2 + \delta(1 - \mathbf{1}_R^T\mathbf{a})^2 \right] \tag{26}$$

subject to the non-negativity constraints for the parameter vector $\mathbf{a}$, where $\delta \in \mathbb{R}^+$ controls the impact of the sum-to-one constraint. The procedure (26) is repeated until convergence and is summarized in Algo. 2. The convergence of this iterative procedure to the global minimum of the objective function (21) is difficult to prove because of the constraints (4) in (24). The next section introduces an alternative subgradient based algorithm whose convergence (to a local minimum of the associated objective function) is ensured.

B. Subgradient-based optimization

A gradient approach could be used to solve the cost function defined in (19) in absence of constraints. However, the problem is more complicated when the constraints (4) have to be considered. The estimation method studied in this section is based on a subgradient optimization (SO) algorithm [26, p. 339] that is appropriate for constrained problems. More precisely, subgradient-based optimization allows each abundance $a_1, \ldots, a_r$ to be updated independently. Thanks to the sum-to-one constraint of the abundance vector, the cost function (19) can be expressed as a function of $\mathbf{a}_{\backslash R}$ by setting $a_R = 1 - \sum_{r=1}^{R-1} a_r$. In that case, the cost function (19) can be rewritten

$$\overline{J}(\mathbf{a}_{\backslash R}) = \frac{1}{2} \left\| \mathbf{y} - \overline{\boldsymbol{\phi}}(\mathbf{a}_{\backslash R}) \right\|^2 \tag{27}$$

where

$$\overline{J}(\mathbf{a}_{\backslash R}) = J\left(a_1, \ldots, a_{R-1}, 1 - \sum_{r=1}^{R-1} a_r\right) \tag{28}$$

$$\overline{\boldsymbol{\phi}}(\mathbf{a}_{\backslash R}) = \boldsymbol{\phi}\left(a_1, \ldots, a_{R-1}, 1 - \sum_{r=1}^{R-1} a_r\right). \tag{29}$$

At a given point $\mathbf{a}_{\backslash R}$, the SO algorithm performs sequential line searches along the directions d_r defined by the partial derivatives with respect to a_r (for $r = 1, \ldots, R-1$), i.e.,

$$d_r = -\frac{\partial \overline{J}(\mathbf{a}_{\backslash R})}{\partial a_r} = \left[\mathbf{y} - \overline{\boldsymbol{\phi}}(\mathbf{a}_{\backslash R})\right]^T \frac{\partial \overline{\boldsymbol{\phi}}(\mathbf{a}_{\backslash R})}{\partial a_r}$$

where the partial derivatives of $\overline{\boldsymbol{\phi}}(\mathbf{a}_{\backslash R})$ are provided in Appendix B. Finally, the line search procedure solves the following problem

$$\hat{\lambda}_r \quad = \quad \arg\min_{\lambda_r} \quad \overline{J}(\mathbf{a}_{\backslash R} - \lambda_r \boldsymbol{u}_r). \tag{30}$$

where $\boldsymbol{u}_r = [0, \ldots, \text{sign}(d_r), 0, \ldots, 0]^T$ is a direction vector of size $(R-1) \times 1$, $0 \leq \lambda_r \leq \lambda_{r,M}$ and $\lambda_{r,M} \in \mathbb{R}^+$ (for $r = 1, \ldots, R-1$) are upper bounds for the line search parameters. More precisely, upper bounding λ_r according to the following rule

$$\lambda_{r,M} = \begin{cases} 0, & \text{if } \; d_r = 0 \\ a_r, & \text{if } \; d_r > 0 \\ a_r - \sum_{i=1, i \neq r}^{R-1} a_i, & \text{if } \; d_r < 0 \end{cases}$$

ensures the constraints (4) are satisfied. The problem (30) can be solved using the golden section method [26, p. 270]. The abundances are then updated component by component. The final algorithm is summarized in Algo. 2. Here again, the procedure is repeated until convergence. The next section presents the performance of the proposed algorithms on synthetic and real hyperspectral images.

V. Simulations

A. Synthetic data

The performance of the proposed nonlinear SU algorithms is first evaluated by unmixing 4 synthetic images of size 50×50 pixels. The $R = 3$ endmembers contained in these images have been extracted from the spectral libraries provided with the ENVI software [27] (i.e., green grass, olive green paint and galvanized steel metal). The first synthetic image I_1 has been generated using the standard linear mixing model (LMM). A second image I_2 has been generated according to the bilinear mixing model introduced in [10], referred to as "Fan model" (FM). A third image I_3 has been generated according to the generalized bilinear mixing model (GBM) presented in [11], whereas a fourth image I_4 has been generated according to the PNMM. For each image, the abundance vectors $\mathbf{a}_p, p = 1, \ldots, 2500$ have been randomly generated according to a uniform distribution over the admissible set defined by the positivity and sum-to-one constraints. All images have been corrupted by an additive white Gaussian noise of variance $\sigma^2 = 2.8 \times 10^{-3}$, corresponding to a signal-to-noise ratio $\text{SNR} = L^{-1}\sigma^{-2} \left\| \boldsymbol{g}_b(\mathbf{a}) \right\|^2 \simeq 15$dB. The nonlinearity coefficients are uniformly drawn in the set $(0, 1)$ for the GBM and the parameter b has been generated uniformly in the set $(-0.3, 0.3)$ for the PPNMM. Different estimation procedures have been considered for the four mixing models. More precisely,

- for the LMM, we have considered the standard FCLS algorithm [2] and the Bayesian algorithm of [21],

- the FM has been unmixed using the LS method introduced in [10] and a Bayesian algorithm similar to the one derived in [11] but assuming all the nonlinearity coefficients are equal to 1,
- the unmixing strategies used for the GBM are the three algorithms presented in [28], i.e., a Bayesian algorithm and two LS methods,
- the Bayesian and LS algorithms presented in Sections III and IV have been used for unmixing the proposed PPNMM. Note that all results presented in this paper have been obtained using the Bayesian MMSE estimator.

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

$$\text{RMSE} = \sqrt{\frac{1}{P}\sum_{p=1}^{P} \|\hat{\mathbf{a}}_p - \mathbf{a}_p\|^2} \tag{31}$$

where $\mathbf{a}_p$ and $\hat{\mathbf{a}}_p$ are the actual and estimated abundance vectors for the pth pixel of the image and P is the number of image pixels. Table I shows the RMSEs associated with the images $I_1, \ldots, I_4$ for the different estimation procedures. Note that the best results (in term of RMSE) for each image have been represented in underlined bold whereas the second best results have been depicted in bold. Table I shows that the abundances estimated by the Bayesian algorithm and the LS methods are similar for the PPNMM. Moreover, for these 4 images, the PPNMM seems to be more robust than the other mixing models to deviations from the actual model.

The unmixing quality can also be evaluated by the reconstruction error (RE) defined as

$$\text{RE} = \sqrt{\frac{1}{PL}\sum_{p=1}^{P} \|\hat{\mathbf{y}}_p - \mathbf{y}_p\|^2} \tag{32}$$

where $\mathbf{y}_p$ is the pth observation vector and $\hat{\mathbf{y}}_p$ its estimate. Table II compares the REs obtained for the different synthetic images. These results show that the REs are close for the different unmixing algorithms even if the estimated abundances can vary more significantly. Again, the proposed PPNMM seems to be more robust than the other mixing models to deviations from the actual model in term of RE.

Fig. 2. shows the estimated distributions of b for the images $I_1, \ldots, I_4$ using the three presented algorithms (i.e., Bayesian, linearization and subgradient). This figure shows that the two algorithms perform similarly for the estimation of the nonlinearity parameter b.

Table III shows the execution times of MATLAB implementations on a 1.66GHz Dual-Core of the proposed algorithms for unmixing the proposed images (2500 pixels for each image). The linearization-based algorithm has the lowest computational cost and also provides accurate estimations. Note that the computational cost of the Bayesian algorithm (which allows prior knowledge to be included in the unmixing procedure) can be prohibitive for larger images and a high number of endmembers. However, the computational cost of the two proposed optimization methods (linearization and gradient-based) is very reasonable which make them very useful for practical applications.

The next set of simulations analyzes the performance of the proposed nonlinear SU algorithms for different numbers of endmembers ($R \in \{3, 6, 9, 12\}$) by unmixing 4 synthetic images of 500 pixels. The endmembers contained in these images have been randomly selected from the fourteen endmembers extracted by VCA from the full Cuprite scene described in [29]. For each image, the abundance vectors $\mathbf{a}_p, (p = 1, \ldots, 500)$ have been randomly generated according to a uniform distribution over the admissible set defined by the positivity and sum-to-one constraints. All images have been corrupted by an additive white Gaussian noise corresponding to a signal-to-noise ratio $\mathrm{SNR} = 20$dB. The nonlinearity coefficients b are uniformly drawn in the set $(-0.3, 0.3)$. Tables IV and V compare the performance of the three proposed methods in term of abundance estimation and reconstruction error. These results show that the three methods perform similarly in term of reconstruction error. The Bayesian estimators tend to provide more accurate abundance estimations (i.e., smaller RMSEs) for large values of R. Indeed, the Taylor and gradient algorithms may be trapped in local minima of the LS criterion (17) for large values of R.

B. Real data

The first real image considered in this section is composed of $L = 189$ spectral bands and was acquired in 1997 by the airborne visible infrared imaging spectrometer (AVIRIS) over the Cuprite mining site in Nevada. A sub-image of size 50×50 pixels has been chosen here to evaluate the proposed unmixing procedures. The scene is mainly composed of muscovite, alunite and kaolinite, as explained in [30]. The endmembers extracted by VCA [19] and the nonlinear EEA proposed in [20] (referred to as "Heylen"), with $R = 3$ are depicted in Fig. 3. The endmembers obtained by the two methods have similar shapes. This result confirms the fact that the geometric

EEAs (such as VCA) can be used as a first approximation for endmember estimation [6]. The estimation algorithms presented in Sections III and IV have been applied independently to each pixel of the scene using the endmembers extracted by the two EEAs. Examples of abundance maps obtained by the Heylen's method are presented in Fig. 4. They are similar to the abundance maps obtained with the VCA algorithm and presented in Fig. 5. However, the advantage of the PPNMM is that it allows the nonlinearities between the observations and the abundance vectors to be analyzed. For instance, Fig. 6 shows the estimated maps of b for the Cuprite image. These results show that the observations are nonlinearly related to the endmembers (since $b \neq 0$). However, the nonlinearity is weak since the estimated values of b are close to 0.

The second real image considered in this section is composed of $L = 189$ spectral bands and was acquired in 1997 by the satellite AVIRIS over the Moffett Field, CA. A sub-image of size 50×50 pixels has also been chosen here to evaluate the proposed unmixing procedures. The scene is mainly composed of water, vegetation and soil. The endmembers extracted by VCA and Heylen's method with $R = 3$ are depicted in Fig. 7. Again, the endmembers obtained by the two methods are similar. Examples of abundance maps estimated by the Heylen's method are presented in Fig. 8. They are similar to the abundance maps obtained with estimation algorithms associated with the LMM and available in [21]. Fig. 9 shows the estimated maps of b for the Moffett image. In the water area, the observations are nonlinearly related to the endmembers (since $b \neq 0$). These nonlinearities can be due to the low amplitude of the water spectrum and nonlinear bathymetric effects.

The quality of unmixing is finally evaluated using the REs for both real images. These REs are compared in Table VI with those obtained by assuming other mixing models. The proposed PPNMM provides smaller REs when compared to other models which is a very encouraging result.

The performance of the gradient-based algorithm has been evaluated on the full Cuprite hyperspectral image composed of $L = 189$ spectral bands. The geologic characteristics of the complete data have been described in [29]. The area of interest of size 190×250 has been previously studied in [19] to test the VCA algorithm with $R = 14$ endmembers. The subgradient-based estimator has been used to estimate the parameters of the PPNMM related to the analyzed scene. Fig. 10 shows the abundance maps estimated by the subgradient-based algorithm corresponding to the $R = 14$ components identified by VCA. These maps are in good

agreement with the maps obtained using FCLS that are shown in Fig. 11. The associated average reconstruction errors are compared in Table VII. As can be seen, the proposed algorithm shows good performance for this example obtained with the full Cuprite image.

VI. Conclusions and future works

A Bayesian and two least squares algorithms were presented for nonlinear spectral unmixing of hyperspectral images. These algorithms assumed that the hyperspectral image pixels are related to the endmembers by a polynomial post-nonlinear mixing model. In the Bayesian framework, the constraints related to the unknown parameters were ensured by using appropriate prior distributions. The posterior distribution of the unknown parameter vector was then derived. The corresponding minimum mean square error estimator was approximated from samples generated using Markov chain Monte Carlo methods. Least squares methods were also investigated for unmixing the polynomial post-nonlinear model. These methods provided results similar to the Bayesian algorithm with a reduced computational cost, making them very attractive for hyperspectral image unmixing. Results obtained on synthetic and real images illustrated the accuracy of the polynomial post-nonlinear model and the performance of the corresponding estimation algorithms. Future works include the study of nonlinear EEAs appropriate for the proposed parametric PPNMM . Deriving nonlinearity detectors based on the proposed parametric PPNMM is also under investigation.

APPENDIX A
INJECTIVITY OF THE NONLINEAR FUNCTIONS

A. Non-injectivity of $\mathbf{s} \mapsto \boldsymbol{g}_b(\mathbf{s})$

The application

$$\begin{aligned} \boldsymbol{g}_b : \quad [0,1]^L &\to \mathbb{R}^L \\ \mathbf{s} &\mapsto \left[s_1 + bs_1^2, \ldots, s_L + bs_L^2\right]^T \end{aligned}$$

with $\mathbf{s} = [s_1, \ldots, s_L]^T$ is not injective. Indeed, if $\boldsymbol{g}_b(\mathbf{s}_1) = \boldsymbol{g}_b(\mathbf{s}_2)$ then

$$\forall l = 1, \ldots, L \quad s_{2,i} \in \left\{ s_{1,i}, -\frac{1}{b} - s_{1,i} \right\}$$

which leads to 2^L solutions for the problem $\boldsymbol{g}_b(\mathbf{s}_1) = \boldsymbol{g}_b(\mathbf{s}_2)$

B. Injectivity of $(\boldsymbol{a}, b) \mapsto \boldsymbol{g}(\boldsymbol{a}, b) = \mathbf{M}\boldsymbol{a} + b(\mathbf{M}\boldsymbol{a}) \odot (\mathbf{M}\boldsymbol{a})$

Let $\mathbf{a}$ and $\mathbf{a}^*$ be two abundance vectors satisfying the positivity and sum-to-one constraints, $\mathbf{M}$ the matrix containing the endmembers and $(b, b^*) \in \mathbb{R}^2$. Consider the nonlinear functional $\boldsymbol{g}(\mathbf{a}, b)$ defined as follows

$$\begin{aligned} \boldsymbol{g}(\mathbf{a}, b) &= \mathbf{M}\mathbf{a} + b(\mathbf{M}\mathbf{a}) \odot (\mathbf{M}\mathbf{a}) \\ &= \sum_{r=1}^{R} a_r \mathbf{m}_r + b \sum_{r=1}^{R} a_r^2 \mathbf{m}_r \odot \mathbf{m}_r + 2b \sum_{r=1}^{R=1} \sum_{j=r+1}^{R} a_r a_j \mathbf{m}_r \odot \mathbf{m}_j \end{aligned}$$

where $\odot$ denotes the term by term product operation. If $\boldsymbol{g}_b(\mathbf{M}\mathbf{a}) = \boldsymbol{g}_{b^*}(\mathbf{M}\mathbf{a}^*)$, then

$$\mathbf{M}(\mathbf{a} - \mathbf{a}_2^*) + b(\mathbf{M}\mathbf{a}) \odot (\mathbf{M}\mathbf{a}) - b^*(\mathbf{M}\mathbf{a}^*) \odot (\mathbf{M}\mathbf{a}^*) = 0$$

and

$$\sum_{r=1}^{R} (a_r - a_r^*)\mathbf{m}_r + \sum_{r=1}^{R} (ba_r^2 - b^* a_r^{*2})\mathbf{m}_r \odot \mathbf{m}_r + 2 \sum_{r=1}^{R=1} \sum_{j=r+1}^{R} (ba_r a_j - b^* a_r^* a_j^*)\mathbf{m}_r \odot \mathbf{m}_j = 0.$$

If the columns of the $L \times \frac{R(R+3)}{2}$ matrix

$$\check{\mathbf{M}} = \{\mathbf{m}_1, \ldots, \mathbf{m}_R, \mathbf{m}_1 \odot \mathbf{m}_1, \ldots, \mathbf{m}_R \odot \mathbf{m}_R, \mathbf{m}_1 \odot \mathbf{m}_2, \ldots, \mathbf{m}_{R-1} \odot \mathbf{m}_R\}$$

are linearly independent, then

$$\mathbf{a} = \mathbf{a}^*, \quad \text{and} \quad b = b^*. \tag{33}$$

Consequently, the identifiability of the unmixing problem assuming the proposed PPNMM is ensured when $\text{rank}(\check{\mathbf{M}}) = \frac{R(R+3)}{2}$, which is usually satisfied when using real pure spectral components. Note that the identifiability of the unmixing problem associated with the LMM requires a similar condition, i.e., $\text{rank}(\mathbf{M}) = R$, where $\mathbf{M} = [\mathbf{m}_1, \ldots, \mathbf{m}_R]^T$.

APPENDIX B
PARTIAL DERIVATIVES

A. Partial derivatives of $\beta(\cdot)$ and $\boldsymbol{h}(\cdot)$

The partial derivative of $\boldsymbol{h}(\cdot)$ with respect to $a_r, (r = 1, \ldots, R)$ is given by

$$\frac{\partial \boldsymbol{h}(\mathbf{a})}{\partial a_r} = 2\,(\mathbf{M}\mathbf{a}) \odot \mathbf{m}_r.$$

Using the following partial derivatives

$$\begin{aligned}
\frac{\partial(\mathbf{y} - \mathbf{M}\mathbf{a})}{\partial a_r} &= -\mathbf{m}_r \\
\frac{\partial(\mathbf{y} - \mathbf{M}\mathbf{a})^T \boldsymbol{h}(\mathbf{a})}{\partial a_r} &= -\mathbf{m}_r^T \boldsymbol{h}(\mathbf{a}) + 2(\mathbf{y} - \mathbf{M}\mathbf{a})^T \left((\mathbf{M}\mathbf{a}) \odot \mathbf{m}_r\right) \\
\frac{\partial \boldsymbol{h}(\mathbf{a})^T \boldsymbol{h}(\mathbf{a})}{\partial a_r} &= 2\boldsymbol{h}(\mathbf{a})^T \frac{\partial \boldsymbol{h}(\mathbf{a})}{\partial a_r}
\end{aligned}$$

and the usual differentiation rules, we obtain

$$\frac{\partial \beta(\mathbf{a})}{\partial a_r} = \frac{1}{\|\boldsymbol{h}(\mathbf{a})\|^4} \left[\frac{\partial(\mathbf{y} - \mathbf{M}\mathbf{a})^T \boldsymbol{h}(\mathbf{a})}{\partial a_r} \boldsymbol{h}(\mathbf{a})^T \boldsymbol{h}(\mathbf{a}) - (\mathbf{y} - \mathbf{M}\mathbf{a})^T \boldsymbol{h}(\mathbf{a}) \frac{\partial \boldsymbol{h}(\mathbf{a})^T \boldsymbol{h}(\mathbf{a})}{\partial a_r} \right].$$

B. Partial derivatives of $\overline{\boldsymbol{\phi}}(\cdot)$

Thanks to the sum-to-one constraint of the abundance vector, the cost function (19) can be expressed as a function of $\mathbf{a}_{\backslash R} = [a_1, \ldots, a_{R-1}]^T$ by setting $a_R = 1 - \sum_{r=1}^{R-1} a_r$. Straightforward computations lead to

$$\mathbf{M}\mathbf{a} = \overline{\mathbf{M}}\mathbf{a}_{\backslash R} + \mathbf{m}_R$$

where $\overline{\mathbf{M}} \triangleq [\mathbf{m}_1 - \mathbf{m}_R, \ldots, \mathbf{m}_{R-1} - \mathbf{m}_R] \triangleq [\overline{\mathbf{m}}_1, \ldots, \overline{\mathbf{m}}_{R-1}]$ is a matrix of size $L \times (R-1)$. The estimated spectrum $\overline{\boldsymbol{\phi}}(\mathbf{a}_{\backslash R})$ can then be expressed as

$$\overline{\boldsymbol{\phi}}(\mathbf{a}_{\backslash R}) = \overline{\mathbf{M}}\mathbf{a}_{\backslash R} + \mathbf{m}_R + \overline{\beta}(\mathbf{a}_{\backslash R})\overline{\boldsymbol{h}}(\mathbf{a}_{\backslash R})$$

where

$$\overline{\beta}(\mathbf{a}_{\backslash R}) \;=\; \beta\left(a_1,\ldots,a_{R-1},1-\sum_{r=1}^{R-1}a_r\right)$$

and

$$\overline{\boldsymbol{h}}(\mathbf{a}_{\backslash R}) \;=\; (\overline{\mathbf{M}}\mathbf{a}_{\backslash R}+\mathbf{m}_R)\odot(\overline{\mathbf{M}}\mathbf{a}_{\backslash R}+\mathbf{m}_R).$$

Using the following partial derivatives with respect to a_r

$$\begin{aligned}
\frac{\partial\overline{\boldsymbol{h}}(\mathbf{a}_{\backslash R})}{\partial a_r} &= 2\overline{\mathbf{m}}_r\odot\left(\overline{\mathbf{M}}\mathbf{a}_{\backslash R}+\mathbf{m}_R\right)\\
\frac{\partial(\mathbf{y}-\overline{\mathbf{M}}\mathbf{a}_{\backslash R}-\mathbf{m}_R)}{\partial a_r} &= -\overline{\mathbf{m}}_r\\
\frac{\partial(\mathbf{y}-\overline{\mathbf{M}}\mathbf{a}_{\backslash R}-\mathbf{m}_R)^T\overline{\boldsymbol{h}}(\mathbf{a}_{\backslash R})}{\partial a_r} &= -\overline{\mathbf{m}}_r^T\overline{\boldsymbol{h}}(\mathbf{a}_{\backslash R})+2(\mathbf{y}-\mathbf{M}\mathbf{a})^T\left(\overline{\mathbf{m}}_r\odot\left(\overline{\mathbf{M}}\mathbf{a}_{\backslash R}+\mathbf{m}_R\right)\right)\\
\frac{\partial\overline{\boldsymbol{h}}(\mathbf{a}_{\backslash R})^T\overline{\boldsymbol{h}}(\mathbf{a}_{\backslash R})}{\partial a_r} &= 2\overline{\boldsymbol{h}}(\mathbf{a}_{\backslash R})^T\frac{\partial\overline{\boldsymbol{h}}(\mathbf{a}_{\backslash R})}{\partial a_r}
\end{aligned}$$

and the usual differentiation rules, we obtain

$$\begin{aligned}
\frac{\partial\overline{\beta}(\mathbf{a}_{\backslash R})}{\partial a_r} &= \frac{1}{\left\|\overline{\boldsymbol{h}}(\mathbf{a}_{\backslash R})\right\|^4}\left\{\frac{\partial(\mathbf{y}-\overline{\mathbf{M}}\mathbf{a}_{\backslash R}-\mathbf{m}_R)^T\overline{\boldsymbol{h}}(\mathbf{a}_{\backslash R})}{\partial a_r}\left[\overline{\boldsymbol{h}}(\mathbf{a}_{\backslash R})^T\overline{\boldsymbol{h}}(\mathbf{a}_{\backslash R})\right]\right.\\
&- \left.(\mathbf{y}-\overline{\mathbf{M}}\mathbf{a}_{\backslash R}-\mathbf{m}_R)^T\overline{\boldsymbol{h}}(\mathbf{a}_{\backslash R})\frac{\partial\overline{\boldsymbol{h}}(\mathbf{a}_{\backslash R})^T\overline{\boldsymbol{h}}(\mathbf{a}_{\backslash R})}{\partial a_r}\right\}.
\end{aligned}$$

Finally, the partial derivative of the estimated spectrum $\overline{\boldsymbol{\phi}}(\mathbf{a}_{\backslash R})$ with respect to a_r is

$$\frac{\partial\overline{\boldsymbol{\phi}}(\mathbf{a}_{\backslash R})}{\partial a_r}=\overline{\mathbf{m}}_r+\frac{\partial\overline{\beta}(\mathbf{a}_{\backslash R})}{\partial a_r}\overline{\boldsymbol{h}}(\mathbf{a}_{\backslash R})+\overline{\beta}(\mathbf{a}_{\backslash R})\frac{\partial\overline{\boldsymbol{h}}(\mathbf{a}_{\backslash R})}{\partial a_r}.$$

REFERENCES

[1] M. Craig, "Minimum volume transforms for remotely sensed data," *IEEE Trans. Geosci. and Remote Sensing*, vol. 32, no. 3, pp. 542–552, may 1994.

[2] 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.

[3] 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.

[4] L. Miao, H. Qi, and H. Szu, "A maximum entropy approach to unsupervised mixed-pixel decomposition," *IEEE Trans. Image Processing*, vol. 16, no. 4, pp. 1008–1021, april 2007.

[5] Z. Yang, G. Zhou, S. Xie, S. Ding, J.-M. Yang, and J. Zhang, "Blind spectral unmixing based on sparse nonnegative matrix factorization," *IEEE Trans. Image Processing*, vol. 20, no. 4, pp. 1112–1125, april 2011.

[6] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, pp. 44–57, Jan. 2002.

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

[8] 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.

[9] 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, no. 1. SPIE, 2009, p. 74770I.

[10] 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.

[11] 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.

[12] K. J. Guilfoyle, M. L. Althouse, and C.-I. Chang, "A quantitative and comparative analysis of linear and nonlinear spectral mixture models using radial basis function neural networks," *IEEE Geosci. and Remote Sensing Lett.*, vol. 39, no. 8, pp. 2314–2318, Aug. 2001.

[13] Y. Altmann, N. Dobigeon, S. McLaughlin, and J.-Y. Tourneret, "Nonlinear unmixing of hyperspectral images using radial basis functions and orthogonal least squares," in *Proc. IEEE Int. Conf. Geosci. and Remote Sensing (IGARSS)*, july 2011, pp. 1151 –1154.

[14] J. Broadwater, R. Chellappa, A. Banerjee, and P. Burlina, "Kernel fully constrained least squares abundance estimates," in *Proc. IEEE Int. Conf. Geosci. and Remote Sensing (IGARSS)*, Barcelona, Spain, 2007, pp. 4041–4044.

[15] K.-H. Liu, E. Wong, and C.-I. Chang, "Kernel-based linear spectral mixture analysis for hyperspectral image classification," in *in Proc. IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, aug. 2009, pp. 1–4.

[16] C. Jutten and J. Karhunen, "Advances in nonlinear blind source separation," in *4th Int. Symp. on Independent Component Analysis and Blind Signal Separation (ICA2003)*, Nara, Japan, April 2003, pp. 245–256.

[17] M. Babaie-Zadeh, C. Jutten, and K. Nayebi, "Separating convolutive post non-linear mixtures," in *Proc. of the 3rd Workshop on Independent Component Analysis and Signal Separation (ICA2001)*, San Diego, 2001, pp. 138–143.

[18] M. Parente and A. Plaza, "Survey of geometric and statistical unmixing algorithms for hyperspectral images," in *Proc. IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Reykjavík, Iceland, 2010.

[19] 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.

[20] R. Heylen, D. Burazerovic, and P. Scheunders, "Non-linear spectral unmixing by geodesic simplex volume maximization," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 3, pp. 534 –542, june 2011.

[21] 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.

[22] V. J. Mathews and G. L. Sicuranza, *Polynomial Signal Processing*. New York: Wiley, 2000.

[23] E. Punskaya, C. Andrieu, A. Doucet, and W. Fitzgerald, "Bayesian curve fitting using MCMC with applications to signal segmentation," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 747–758, March 2002.

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

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

[26] M. Bazaraa, H. Sherali, and C. Shetty, *Nonlinear programming: Theory and algorithms*, 2nd ed., Wiley, Ed., 1993.

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

[28] 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.

[29] R. N. Clark *et al.*, "Imaging spectroscopy: Earth and planetary remote sensing with the USGS Tetracorder and expert systems," *J. Geophys. Res.*, vol. 108, no. E12, pp. 5–1–5–44, Dec. 2003.

[30] 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. 2657–2669, Nov. 2009.

ALGORITHM 1

Gibbs Sampling Algorithm

1: Initialization ($t = 0$)

- $\mathbf{a}^{(0)}, b^{(0)}, \sigma^{2(0)}, \sigma_b^{2(0)}$.

2: Iterations ($t \geq 1$)

3: Set $\boldsymbol{c} = [c_1, \ldots, c_{R-1}]^T = \mathbf{a}_{\setminus R}^{(t-1)}$

4: **for** $r = 1 : R - 1$ **do**

5:
- Sample a candidate ζ_r using a Gaussian proposal distribution $\mathcal{N}\left(a_r^{(t-1)}, \sigma_r^2\right)$.
- Compute $\rho = \min\left\{\dfrac{f(\zeta_r|\mathbf{y}, b, \boldsymbol{c}_{\setminus r}, \sigma^2, \sigma_b^2)}{f(c_r|\mathbf{y}, b, \boldsymbol{c}_{\setminus r}, \sigma^2, \sigma_b^2)}, 1\right\}$
- Set $c_r = \begin{cases} \zeta_r & \text{with probability} \quad \rho \\ c_r & \text{with probability} \quad 1 - \rho \end{cases}$

6: **end for**

7: Set $\mathbf{a}_{\setminus R}^{(t)} = [c_1, \ldots, c_{R-1}]^T$

8: Set $a_R^{(t)} = 1 - \sum_{r=1}^{R-1} a_r^{(t)}$

9: Sample $b^{(t)}$ from the pdf in (14)

10: Sample $\sigma^{2(t)}$ from the pdf in (15)

11: Sample $\sigma_b^{2(t)}$ from the pdf in (16)

12: Set $t = t + 1$.

ALGORITHM 2

Taylor Approximation Algorithm

1: Initialization ($t = 0$)

- Set $\mathbf{a}^{(0)}$

2: Iterations ($t \geq 0$)

3: Compute the gradient matrix of $\boldsymbol{\phi}$ at $\mathbf{a}^{(t)}$ using (23)

4: Compute $\mathbf{a}^{(t)}$ using (24)

5: Compute $b^{(t)}$ using (21)

6: Set $t = t + 1$.

ALGORITHM 3

Constrained Subgradient Algorithm

1: Initialization ($t = 0$)

- Set $\mathbf{a}^{(0)}$

2: Iterations ($t \geq 1$)

3: Set $\boldsymbol{c} = [c_1, \ldots, c_{r-1}]^T = \mathbf{a}_{\backslash R}^{(t-1)}$

4: **for** $r = 1 : R-1$ **do**

5:
- Compute $d_r = -\frac{\partial \overline{J}(\mathbf{c})}{\partial c_r}$
- Compute $\lambda_{r,M}$ from (31)
- Compute $\hat{\lambda}_r$ from (30)
- Set $c_r = c_r - \hat{\lambda}_r d_r$

6: **end for**

7: Set $\mathbf{a}_{\backslash R}^{(t)} = \boldsymbol{c}$

8: Set $a_R^{(t)} = 1 - \sum_{r=1}^{R-1} a_r^{(t)}$

9: Compute $b^{(t)}$ using (21)

10: Set $t = t + 1$.

TABLE I

ABUNDANCE RMSES ($\times 10^{-2}$): SYNTHETIC IMAGES .

		I_1 (LMM)	I_2 (FM)	I_3 (GBM)	I_4 (PPNMM)
LMM	Bayesian [21]	**1.58**	27.54	15.16	18.88
	FCLS [2]	**1.58**	24.72	9.49	16.87
FM	Bayesian	22.67	**1.51**	13.63	16.84
	Taylor [10]	22.67	**1.49**	12.61	26.33
GBM	Bayesian [28]	3.24	17.49	9.09	16.18
	Taylor [28]	6.32	14.67	7.07	15.61
	Gradient [28]	4.28	4.26	**3.01**	15.05
PPNMM	Bayesian	2.75	3.43	**3.22**	**2.93**
	Taylor	**2.70**	3.83	3.26	**3.33**
	Gradient	2.93	3.43	3.43	**2.93**

TABLE II

REs ($\times 10^{-2}$): SYNTHETIC IMAGES .

		I_1 (LMM)	I_2 (FM)	I_3 (GBM)	I_4 (PPNMM)
LMM	Bayesian [21]	**5.28**	6.54	5.65	5.89
	FCLS [2]	**5.28**	5.74	5.42	5.48
FM	Bayesian	5.61	**5.29**	5.38	5.76
	Taylor [10]	5.61	**5.28**	5.38	5.75
GBM	Bayesian [28]	**5.29**	5.49	5.33	5.44
	Taylor [28]	5.31	5.40	5.30	5.42
	Gradient [28]	**5.29**	5.30	**5.28**	5.41
PPNMM	Bayesian	**5.28**	**5.29**	**5.28**	**5.28**
	Taylor	**5.29**	**5.29**	**5.28**	**5.28**
	Gradient	**5.29**	**5.29**	**5.28**	**5.28**

TABLE III

COMPUTATIONAL TIMES OF THE UNMIXING ALGORITHMS FOR 2500 PIXELS (IN SECOND).

	I_1	I_2	I_3	I_4
Bayesian	5960	6200	6600	5970
Taylor	5	10	8	7
Subgradient	84	102	96	101

TABLE IV

AVERAGE RMSEs($\times 10^{-2}$): SYNTHETIC IMAGES .

	Bayesian		Taylor	Gradient
	MMSE	MAP		
R=3	12.99	18.06	16.34	16.31
R=6	18.46	27.86	30.99	29.79
R=9	17.07	28.68	35.69	34.24
R=12	16.38	27.98	38.66	36.66

TABLE V

AVERAGE RES($\times 10^{-2}$): SYNTHETIC IMAGES .

	Bayesian		Taylor	Gradient
	MMSE	MAP		
R=3	4.18	4.22	4.17	4.17
R=6	4.22	4.24	4.20	4.20
R=9	4.27	4.29	4.24	4.24
R=12	4.18	4.19	4.13	4.13

TABLE VI

RES ($\times 10^{-2}$): CUPRITE AND MOFFETT IMAGES

		VCA		Heylen	
		Cuprite	Moffett	Cuprite	Moffett
LMM	Bayesian [21]	2.14	2.70	2.35	2.02
	FCLS [2]	2.11	2.62	2.10	2.00
FM	Bayesian	7.36	2.31	2.30	1.92
	Taylor [10]	3.05	2.29	2.29	1.92
GBM	Bayesian [28]	2.24	2.57	2.11	1.99
	Taylor [28]	2.34	2.41	2.03	2.01
	Gradient [28]	**2.02**	2.30	2.04	1.93
PPNMM	Bayesian	**1.19**	1.59	**1.91**	**1.85**
	Taylor	**1.19**	**1.54**	**1.90**	**1.84**
	Gradient	**1.19**	**1.55**	**1.90**	1.87

TABLE VII

AVERAGE RES($\times 10^{-2}$) FOR THE FULL CUPRITE SCENE.

FCLS	Taylor FM [10]	Gradient GBM [28]	Gradient
0.49	0.53	0.48	0.44

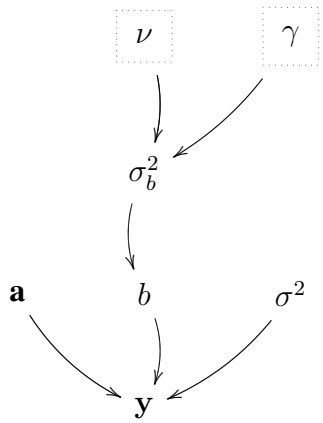

Fig. 1. DAG for the parameter priors and hyperpriors (the fixed parameters appear in dashed boxes).

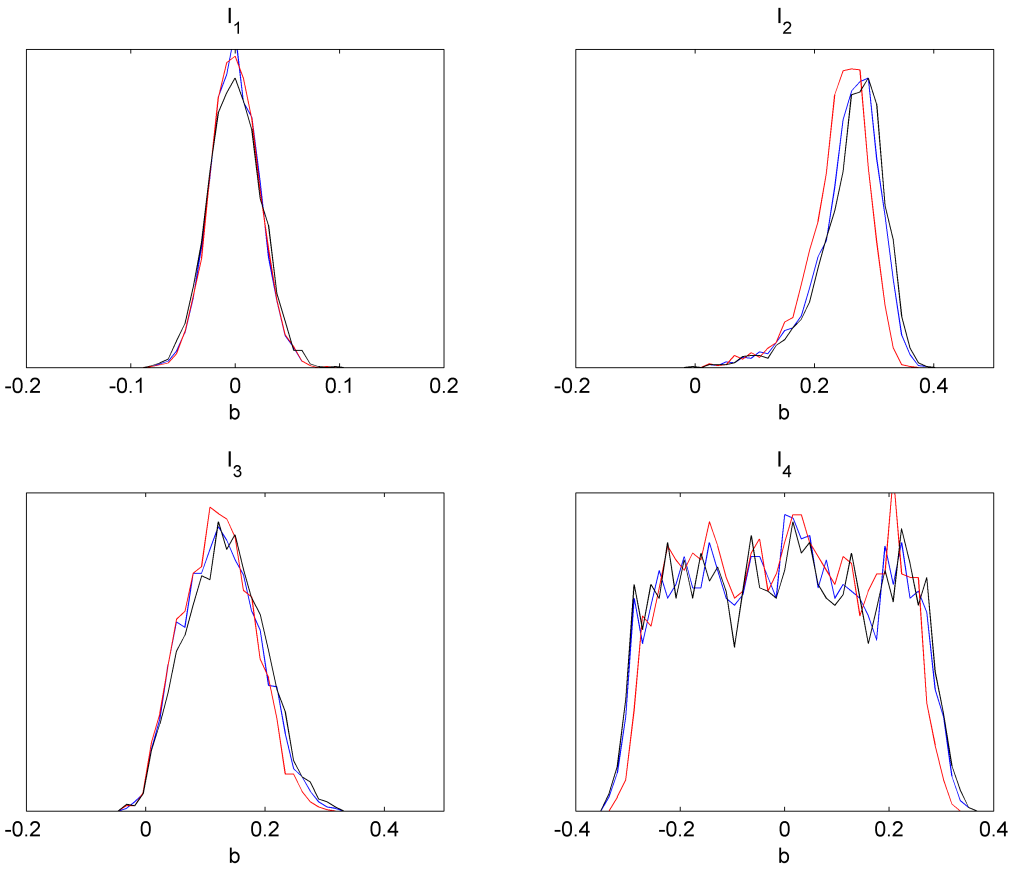

Fig. 2. Histograms of the estimated nonlinearity parameter $\hat{b}$ for the four synthetic images estimated by the Bayesian (black), linearization-based (red) and subgradient-based (blue) algorithms.

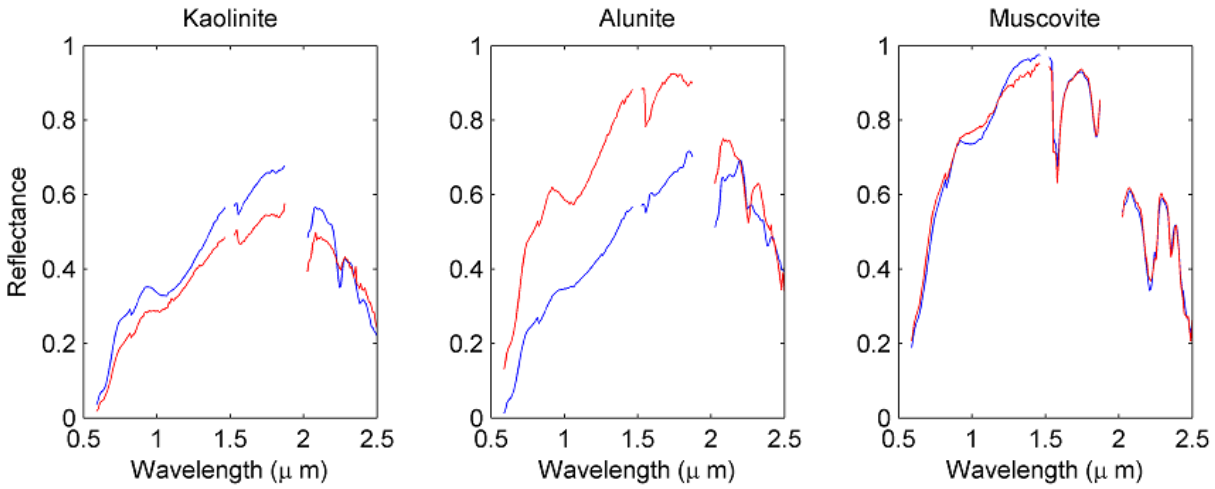

Fig. 3. The $R = 3$ endmembers estimated by VCA (blue lines) and Heylen (red lines) for the Cuprite scene.

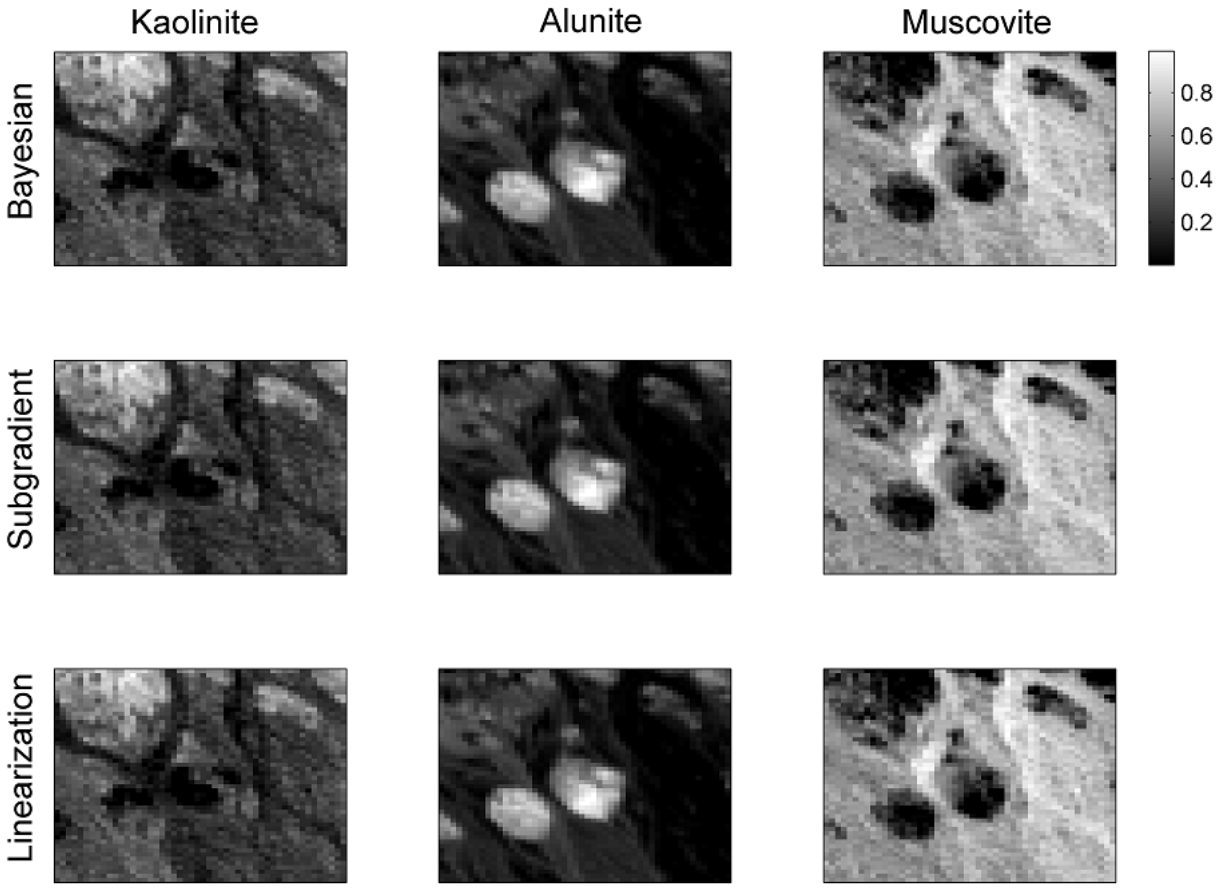

Fig. 4. Abundance maps estimated by the Bayesian, linearization and subgradient methods for the Cuprite scene using Heylen's method.

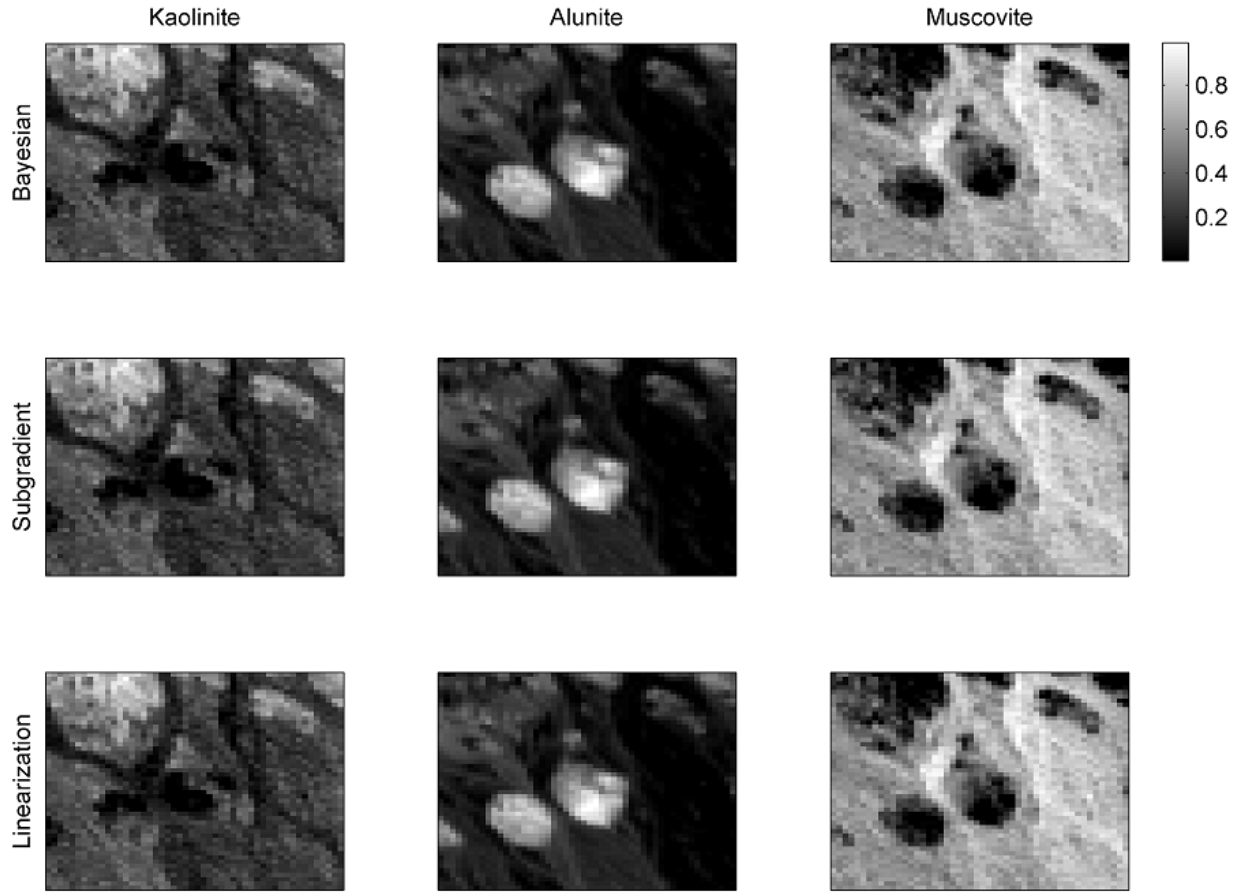

Fig. 5. Abundance maps estimated by the Bayesian, linearization and subgradient methods for the Cuprite scene using the VCA algorithm.

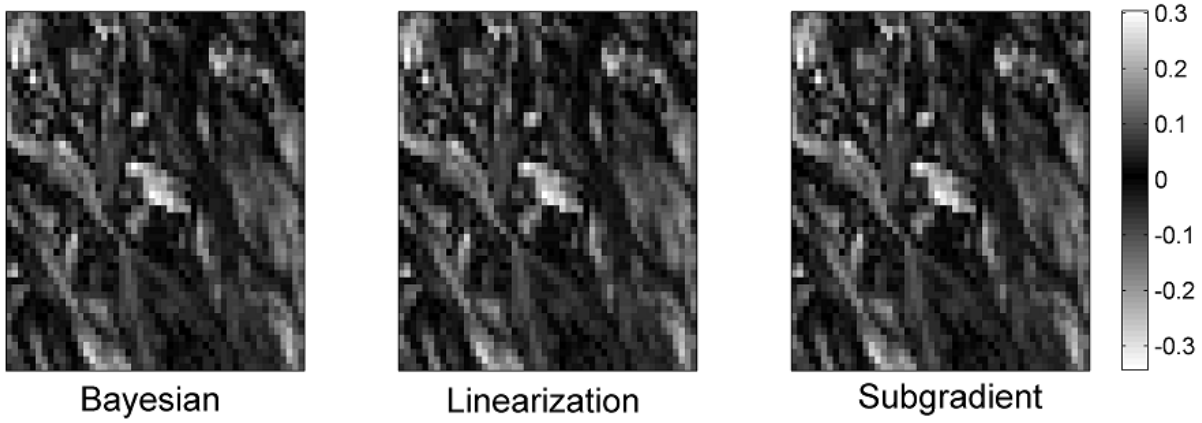

Fig. 6. Maps of the nonlinearity parameter b estimated by the Bayesian, linearization and subgradient methods for the Cuprite scene.

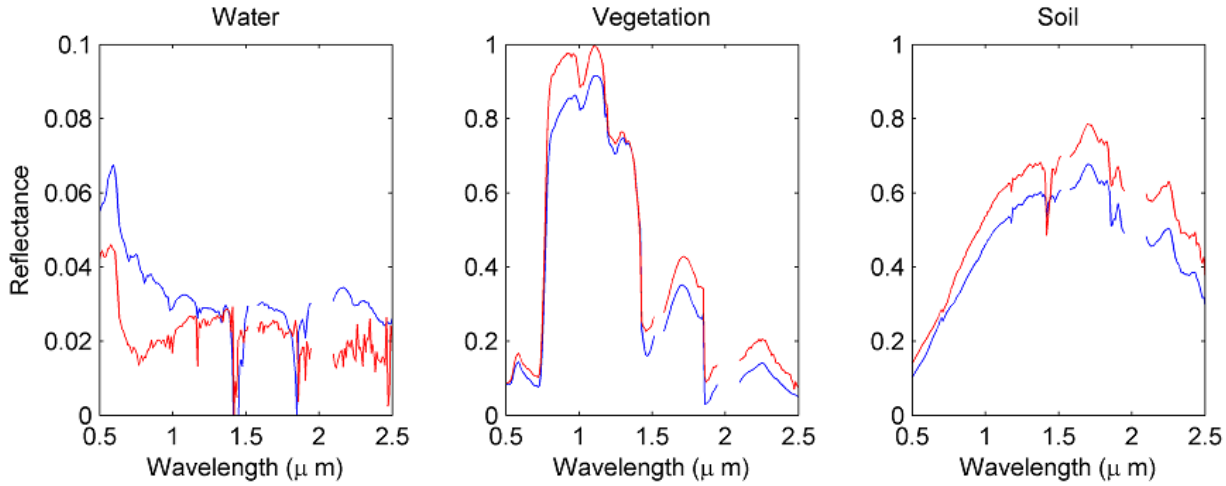

Fig. 7. The $R = 3$ endmembers estimated by VCA (blue lines) and Heylen (red lines) for the Moffett scene.

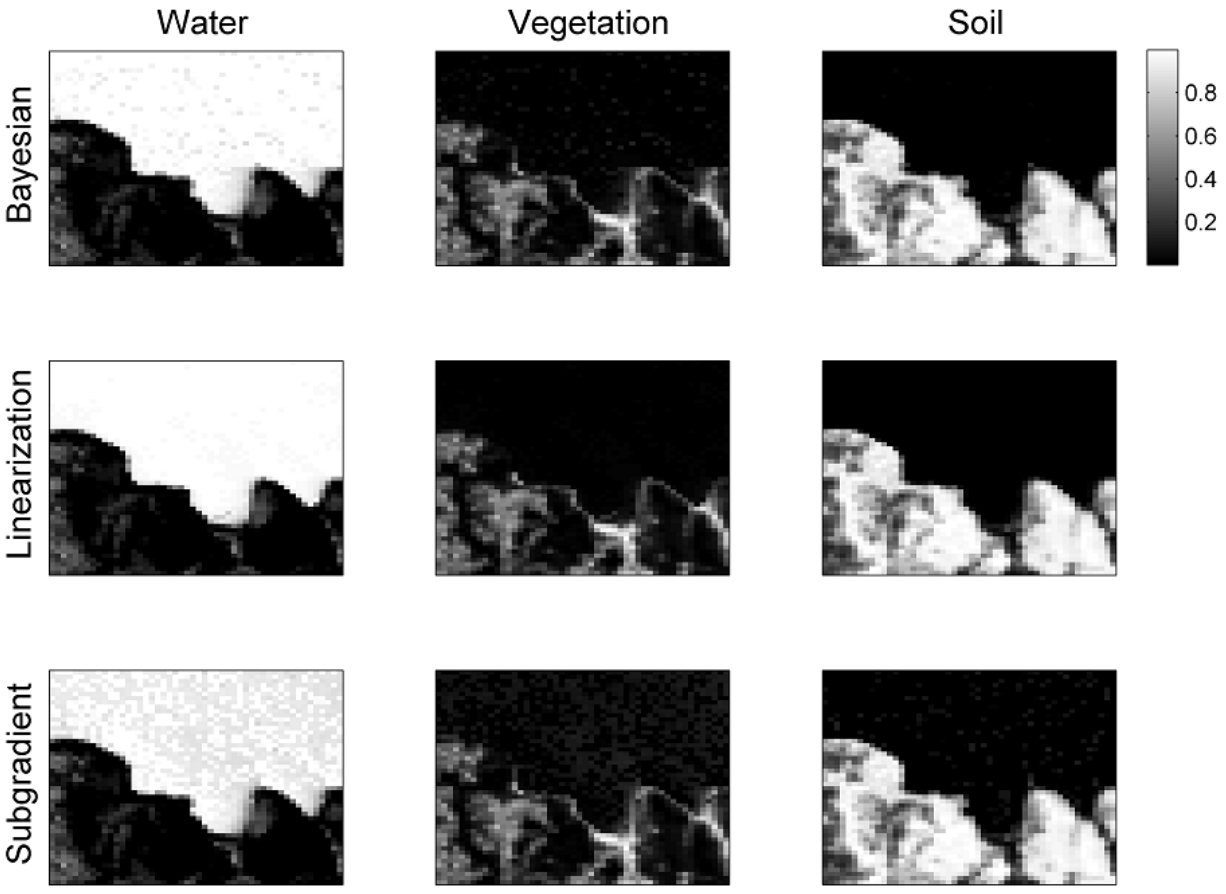

Fig. 8. Abundance maps estimated by the Bayesian, linearization and subgradient methods for the Moffett scene.

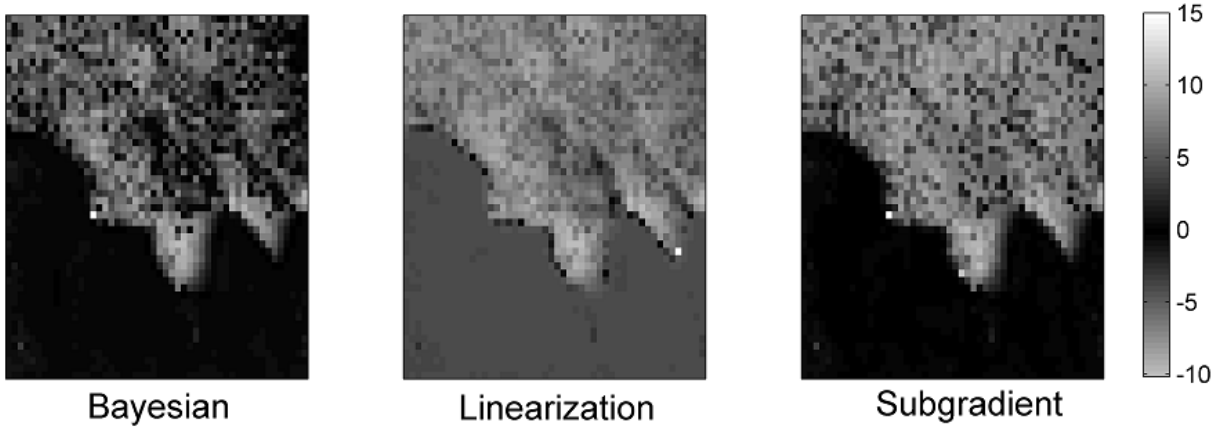

Fig. 9. Maps of the nonlinearity parameter b estimated by the Bayesian, linearization and subgradient methods for the Moffett scene.

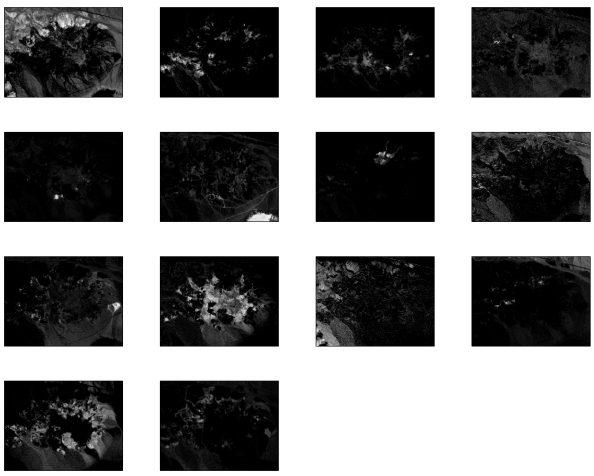

Fig. 10. Fourteen abundance maps estimated with the gradient-based algorithm for the Cuprite scene.

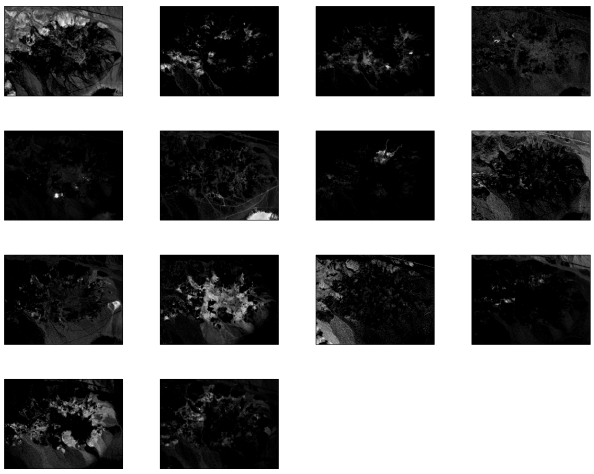

Fig. 11. Fourteen abundance maps estimated with the FCLS algorithm for the Cuprite scene.