NONLINEAR HYPERSPECTRAL UNMIXING USING GAUSSIAN PROCESSES

Y. Altmann, N. Dobigeon, J.-Y. Tourneret

University of Toulouse,
IRIT-ENSEEIHT,
Toulouse, France

S. McLaughlin

School of Engineering and Physical Sciences,
Heriot-Watt University,
Edinburgh, United Kingdom

ABSTRACT

This paper presents an unsupervised algorithm for nonlinear unmixing of hyperspectral images. The proposed model assumes that the pixel reflectances result from a nonlinear function of the abundance vectors associated with the pure spectral components. We assume that the spectral signatures of the pure components and the nonlinear function are unknown. The first step of the proposed method estimates the abundance vectors for all the image pixels using a Gaussian process latent variable model. The endmembers are subsequently estimated using Gaussian process regression. The performance of the unmixing strategy is compared with state-of-the-art unmixing strategies on synthetic data. One of the interesting properties of the proposed strategy is its robustness to the absence of pure pixels in the image.

Index Terms— Hyperspectral images, nonlinear spectral unmixing, unsupervised unmixing, Gaussian process regression, Bayesian estimation.

1. INTRODUCTION

Spectral unmixing (SU) consists of identifying the macroscopic materials present in an hyperspectral image and quantifying the proportions of these materials in the image pixels. Many SU strategies assume that pixel reflectances are linear mixtures of pure component spectra [1]. The resulting linear mixing model (LMM) has been widely adopted in the literature and has provided some interesting results. However, as discussed in [1], the LMM can be inappropriate for some hyperspectral images. Nonlinear mixing models provide an interesting alternative to overcome the inherent limitations of the LMM. For instance, the presence of relief can induce multiple scattering effects between the different materials present in the image. These nonlinear scattering effects typically occur in vegetation areas [2] and urban scenes. Bilinear models have been studied in [2–5] for modeling these multiple scattering effects. Conversely, the bidirectional reflectance-based model of [6] focusses on hyperspectral images including intimate mixtures. Other more flexible unmixing techniques have been also proposed to handle wider classes of nonlinearities, including radial basis function networks [7] post-nonlinear mixing models [8] and kernel-based models [9].

Most existing unmixing strategies can be decomposed into two steps referred to as endmember extraction and abundance estimation are performed consecutively. In the last decade, many linear endmember extraction algorithms (EEAs) have been developed to identify the pure spectral components contained in a hyperspectral image. However, these EEAs can be inappropriate when endmembers are nonlinearly mixed. More recently, an EEA was proposed in [10] to extract endmembers from a set of nonlinearly mixed pixels, based on the approximation of geodesic distances defined in manifolds. Similarly, our approach assumes that the data lie on a (possibly nonlinear) manifold that is parameterized by a kernel. The proposed strategy can be divided into two steps: 1) estimating the abundances using a posterior distribution obtained by marginalizing the endmembers, 2) recovering the endmembers using the prediction capacity of Gaussian processes (GPs). This approach breaks from the usual paradigm of spectral unmixing. More precisely, this paper considers a kernel-based approach for nonlinear SU based on a nonlinear dimensionality reduction using a Gaussian process latent variable model (GPLVM). A particular form of kernel is used to ensure the unmixing strategy is accurate for linear and bilinear mixing models. The algorithm proposed herein is "unsupervised" in the sense that the endmembers contained in the image and the mixing model are not known. Only the number of endmembers is assumed to be known. As a consequence, the parameters to be estimated are the kernel parameters, the endmember spectra and the abundances for all image pixels.

The paper is organized as follows. Section 2 presents the nonlinear mixing model considered in this paper for hyperspectral image unmixing. Section 3 introduces the GPLVM used for latent variable estimation. Section 4 studies the endmember estimation procedure using GP regression. Some simulation results conducted on synthetic data are shown and discussed in Section 5. Finally, conclusions are drawn in Section 6.

2. NONLINEAR MIXING MODEL

Consider a hyperspectral image of N pixels, composed of R endmembers and observed in L spectral bands. For convenience, the data are assumed to have been previously centered, i.e., the sample mean of the N original pixels has been subtracted from each observed pixel. The L-spectrum $\mathbf{y}(n) = [y_1(n), \ldots, y_L(n)]^T$ of the nth mixed pixel ($n = 1, \ldots, N$) is assumed to be a transformation of its corresponding abundance vector $\boldsymbol{a}(n) = [a_1(n), \ldots, a_R(n)]^T$ as follows

$$\begin{aligned} \mathbf{y}(n) &= \boldsymbol{g}\left[\boldsymbol{a}(n)\right] + \mathbf{e}(n), \quad n = 1, \ldots, N \\ &= \boldsymbol{W}_0 \boldsymbol{\psi}\left[\boldsymbol{a}(n)\right] + \mathbf{e}(n), \quad n = 1, \ldots, N \end{aligned} \tag{1}$$

where $\boldsymbol{g}(\cdot) = \boldsymbol{W}_0 \boldsymbol{\psi}[\cdot]$ is an unknown nonlinear function, where $\boldsymbol{W}_0$ is an $L \times D$ matrix and the dimension D is the dimension of the subspace spanned by the transformed abundance vectors $\boldsymbol{\psi}\left[\boldsymbol{a}(n)\right], n = 1, \ldots, N$. The noise vector $\mathbf{e}(n)$ is an independent, identically distributed (i.i.d.) white Gaussian noise sequence with variance σ^2, i.e., $\mathbf{e}(n) \sim \mathcal{N}\left(\mathbf{e}(n)|\mathbf{0}_L, \sigma^2 \mathbf{I}_L\right), n = 1, \ldots, N$.

This work was supported in part by the Direction Générale de l'armement, French Ministry of defence and the Madonna project. Part of this work has been funded by the Hypanema ANR Project ANR Project n°ANR-12-BS03-003.

Of course, the performance of the unmixing strategy relies on the choice of the nonlinear function $\boldsymbol{\psi}$. In this paper, we will use the following nonlinearity

$$\begin{aligned}\boldsymbol{\psi}: \quad \mathbb{R}^R &\to \mathbb{R}^D \\ \boldsymbol{a} &\mapsto \boldsymbol{\psi}(\boldsymbol{a}) = [a_1, \ldots, a_R, a_1 a_2 \ldots, a_{R-1} a_R]^T \end{aligned} \quad (2)$$

with $D = R(R+1)/2$. The primary motivation for considering this particular kind of nonlinearity is the fact that the resulting mixing model is a bilinear model with respect to each abundance $a_r, r = 1, \ldots, R$ (see [11] for details). Due to physical constraints, the abundance vector $\boldsymbol{a}(n) = [a_1(n), \ldots, a_R(n)]^T$ satisfies the following positivity and sum-to-one constraints

$$\sum_{r=1}^{R} a_r(n) = 1, \;\; a_r(n) \geq 0, \forall r \in \{1, \ldots, R\}. \quad (3)$$

Since the nonlinearity $\boldsymbol{\psi}$ is fixed, the problem of unsupervised spectral unmixing is to determine the $L \times D$ spectrum matrix $\boldsymbol{W}_0$, the $N \times R$ abundance matrix $\mathbf{A} = [\boldsymbol{a}(1), \ldots, \boldsymbol{a}(N)]^T$ satisfying (1) under the constraints (3) and the noise variance σ^2. Unfortunately, it can be shown that the solution of this constrained problem is not unique. More precisely, ina similar fashion to the linear case, solving this unsupervised unmixing problem does not ensure the estimated abundances occupy the largest volume in the simplex defined by (3). To tackle this problem, we first relax the positivity constraints for the elements of the matrix $\mathbf{A}$ and to consider only the sum-to-one constraint. For ease of understanding, we introduce *latent variable* vectors $\boldsymbol{x}(n) = [x_1(n), \ldots, x_R(n)]^T$ satisfying the sum-to-one constraint

$$\sum_{r=1}^{R} x_r(n) = 1, \quad n = 1, \ldots, N. \quad (4)$$

The positivity constraint will be handled subsequently by a scaling procedure discussed in the next section. The next section presents the Bayesian model for abundance estimation using GPLVMs.

3. BAYESIAN MODEL

GPLVMs [12] are powerful tools for probabilistic nonlinear dimensionality reduction that rewrite the nonlinear model (1) as a nonlinear mapping from a latent space to the observation space as follows

$$\begin{aligned}\mathbf{y}(n) &= \boldsymbol{W}\boldsymbol{\psi}[\boldsymbol{x}(n)] + \mathbf{e}(n) \\ &= \boldsymbol{P}\boldsymbol{U}^T\boldsymbol{\psi}[\boldsymbol{x}(n)] + \mathbf{e}(n), \quad n = 1, \ldots, N \end{aligned} \quad (5)$$

where $\boldsymbol{\psi}$ is defined in (2), $\boldsymbol{P} = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_L]^T$ is an $L \times D$ matrix with $\boldsymbol{p}_\ell = [p_{\ell,1}, \ldots, p_{\ell,D}]^T$, $\boldsymbol{U}$ a $D \times D$ matrix and $\boldsymbol{W} = \boldsymbol{P}\boldsymbol{U}^T$. This factorization of $\boldsymbol{W}$ will allow the marginalization of $\boldsymbol{P}$ (as will be shown in the next section) that is easier than the marginalization of $\boldsymbol{W}$ (see [11] for details). Note that from (1) and (5) the columns of $\boldsymbol{P}$ span the same subspace as the columns of $\boldsymbol{W}_0$. Note also that when $\boldsymbol{W}_0$ is full rank, it can be shown that the latent variables are necessarily linear combinations of the abundance vectors of interest. Fig. 1 illustrates the mapping from the abundance vectors to the observations that will be used in this paper. For brevity, the $D \times 1$ vectors $\boldsymbol{\psi}[\boldsymbol{x}(n)]$ will be denoted as $\boldsymbol{\psi}_x(n)$ in the sequel.

Assuming independence between the observations, the statistical properties of the noise lead to the following likelihood of the $N \times L$ observation matrix $\mathbf{Y} = [\mathbf{y}(1), \ldots, \mathbf{y}(N)]^T = [\mathbf{y}_1, \ldots, \mathbf{y}_L]$

$$\mathbf{Y}|\boldsymbol{P}, \boldsymbol{U}, \boldsymbol{X}, \sigma^2 \sim \prod_{\ell=1}^{L} \mathcal{N}\left(\mathbf{y}_\ell | \boldsymbol{C}\boldsymbol{p}_\ell, \sigma^2 \mathbf{I}_L\right) \quad (6)$$

where $\boldsymbol{\Psi}_x = [\boldsymbol{\psi}_x(1), \ldots, \boldsymbol{\psi}_x(N)]^T$ and $\boldsymbol{C} = \boldsymbol{\Psi}_x\boldsymbol{U}$ are $N \times D$ matrices.

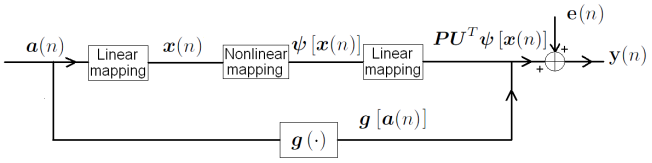


Fig. 1. Nonlinear mapping from the abundances vectors to the observed mixed pixels.

3.1. Parameter priors

GPLVMs construct a smooth mapping from the latent space to the observation space that preserves dissimilarities [13]. However, preserving local distances is also interesting: spectrally close pixels are expected to have similar abundance vectors and thus similar latent variables. In this paper, we use locally linear embedding (LLE) to assign an appropriate prior to $\boldsymbol{X}$. First, the K nearest neighbors $\{\mathbf{y}(j)\}_{j \in \nu_i}$ of each observation vector $\mathbf{y}(i)$ are computed using the Euclidean distance (ν_i is the set of integers j such that $\mathbf{y}(j)$ is a neighbor of $\mathbf{y}(i)$). The weight matrix $\boldsymbol{\Lambda}_{\text{LLE}} = [\lambda_{i,j}]$ of size $N \times N$ providing the best reconstruction of $\mathbf{y}(i)$ from its neighbors is then estimated as

$$\boldsymbol{\Lambda}_{\text{LLE}} = \arg\min_{\boldsymbol{\Lambda}} \sum_{i=1}^{N} \left\| \mathbf{y}(i) - \sum_{j \in \nu_i} \lambda_{i,j} \mathbf{y}(j) \right\|^2. \quad (7)$$

The locally linear patches obtained by the LLE can then be used to set the following prior for the latent variable matrix

$$\begin{aligned} f(\boldsymbol{X}|\boldsymbol{\Lambda}_{\text{LLE}}, \gamma) &\propto \exp\left[-\frac{\gamma}{2} \sum_{i=1}^{N} \left\| \boldsymbol{x}(i) - \sum_{j \in \nu_i} \lambda_{i,j} \boldsymbol{x}(j) \right\|^2\right] \\ &\times \prod_{n=1}^{N} \mathbf{1}_{\mathcal{D}}[\boldsymbol{x}(n)] \end{aligned} \quad (8)$$

where γ is a fixed hyperparameter to be adjusted and $\mathbf{1}_{\mathcal{D}}(\cdot)$ is the indicator function over the set $\mathcal{D}$ defined by the constraints (4). The main motivation for using an LLE-based prior for X is that spectrally close pixels are expected to be described by close latent variables.

In this paper, we propose to assign a prior to $\boldsymbol{P}$ using the standard principal component analysis (PCA) (note again that the data have been centered). Assuming prior independence between $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_L$, the following prior is considered for the matrix $\boldsymbol{P}$

$$f\left(\boldsymbol{P}|\overline{\boldsymbol{P}}, s^2\right) = \left(\frac{1}{2\pi s^2}\right)^{\frac{NL}{2}} \prod_{\ell=1}^{L} \exp\left[-\frac{1}{2s^2} \left\|\boldsymbol{p}_\ell - \bar{\boldsymbol{p}}_\ell\right\|^2\right] \quad (9)$$

where $\overline{\boldsymbol{P}} = [\bar{\boldsymbol{p}}_1, \ldots, \bar{\boldsymbol{p}}_L]^T$ is an $L \times D$ projection matrix containing the first D eigenvectors of the sample covariance matrix of the observations (provided by PCA) and s^2 is a dispersion parameter that controls the informative or non-informative nature of the prior. The main advantage of using the conjugate prior (9) is the fact that it allows $\boldsymbol{P}$ to be marginalized, as will be shown in the next section. Non-informative priors are assigned to the noise variance σ^2, the matrix $\boldsymbol{U}$ and the hyperparameter s^2, i.e,

$$\begin{aligned} f(\sigma^2) &\propto \mathbf{1}_{(0,\delta_{\sigma^2})}(\sigma^2) \\ f(u_{i,j}) &\propto \mathbf{1}_{(-\delta_{\boldsymbol{U}},\delta_{\boldsymbol{U}})}(u_{i,j}) \\ f(s^2) &\propto \mathbf{1}_{(0,\delta_{s^2})}(s^2) \end{aligned} \quad (10)$$

where the intervals $(0, \delta_{\sigma^2})$, $(0, \delta_{s^2})$ and $(-\delta_{\boldsymbol{U}}, \delta_{\boldsymbol{U}})$ cover the possible values of the parameters σ^2, s^2 and $\boldsymbol{U}$.

3.2. Marginalized posterior distribution

Assuming prior independence between $\boldsymbol{P}$, $\boldsymbol{X}$, $\boldsymbol{U}$, s^2 and σ^2, the marginalized posterior distribution of $\boldsymbol{\theta} = (\boldsymbol{X}, \boldsymbol{U}, s^2, \sigma^2)$ can be expressed as

$$f\left(\boldsymbol{\theta}|\mathbf{Y}, \boldsymbol{\Lambda}_{\text{LLE}}, \overline{\boldsymbol{P}}, \gamma\right) \propto f(\mathbf{Y}|\boldsymbol{\theta}, \overline{\boldsymbol{P}}) f(\boldsymbol{\theta}|\boldsymbol{\Lambda}_{\text{LLE}}, \gamma) \quad (11)$$

where $f(\boldsymbol{\theta}|\boldsymbol{\Lambda}_{\text{LLE}}, \gamma) = f(\boldsymbol{X}|\boldsymbol{\Lambda}_{\text{LLE}}, \gamma) f(\boldsymbol{U}) f(s^2) f(\sigma^2)$ and

$$\begin{aligned} f(\mathbf{Y}|\boldsymbol{\theta}, \overline{\boldsymbol{P}}) &= \int f(\mathbf{Y}|\boldsymbol{P}, \boldsymbol{\theta}) f\left(\boldsymbol{P}|\overline{\boldsymbol{P}}, s^2\right) \mathrm{d}\boldsymbol{P} \\ &\propto |\boldsymbol{\Sigma}|^{-\frac{L}{2}} \exp\left[-\frac{1}{2}\text{tr}(\boldsymbol{\Sigma}^{-1}\bar{\mathbf{Y}}\bar{\mathbf{Y}}^T)\right] \end{aligned} \quad (12)$$

where $\boldsymbol{\Sigma} = s^2\boldsymbol{C}\boldsymbol{C}^T + \sigma^2\mathbf{I}_N$, $\bar{\mathbf{y}}_\ell = \mathbf{y}_\ell - \boldsymbol{C}\bar{\boldsymbol{p}}_\ell$ is an $N \times 1$ vector, $\bar{\mathbf{Y}} = [\bar{\mathbf{y}}_1, \ldots, \bar{\mathbf{y}}_L] = \mathbf{Y} - \boldsymbol{C}\overline{\boldsymbol{P}}^T$ is an $N \times L$ matrix and tr$(\cdot)$ denotes the matrix trace. In this paper, we use a scaled conjugate gradient (SCG) method to maximize the marginalized log-posterior. The partial derivatives of the log-posterior w.r.t. $\boldsymbol{\theta}$ can be found in [11]. The resulting latent variable estimation procedure is referred to as locally linear GPLVM (LL-GPLVM). The MAP estimator of $\boldsymbol{P}$, conditioned upon $\hat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{X}}, \widehat{\boldsymbol{U}}, \hat{s}^2, \hat{\sigma}^2)$ (the maximum a posteriori (MAP) estimator of $\boldsymbol{\theta}$ obtained by maximizing (11)), is given by

$$\widehat{\boldsymbol{P}} = \left(\mathbf{Y}^T\widehat{\boldsymbol{C}} - \overline{\boldsymbol{P}}\right)\widehat{\boldsymbol{S}} \quad (13)$$

where $\widehat{\boldsymbol{S}}^{-1} = \hat{\sigma}^{-2}\widehat{\boldsymbol{C}}^T\widehat{\boldsymbol{C}} + \hat{s}^{-2}\mathbf{I}_D$, $\widehat{\boldsymbol{\Psi}}_x = [\boldsymbol{\psi}_{\hat{x}}(1), \ldots, \boldsymbol{\psi}_{\hat{x}}(N)]^T$, $\widehat{\boldsymbol{C}} = \widehat{\boldsymbol{\Psi}}_x\widehat{\boldsymbol{U}}$ and $\widehat{\boldsymbol{X}} = [\hat{\boldsymbol{x}}(1), \ldots, \hat{\boldsymbol{x}}(N)]^T$ (see [11] for details).

3.3. Scaling step

As mentioned above, the maximization of (11) provides a set of latent variables that represent the data but can differ from the abundance vectors of interest. However, due to the linear relation between the latent variables and the abundances, the abundance matrix (satisfying (3)) can be retrieved using a scaling procedure. More precisely, this procedure consists of estimating the $R \times R$ matrix $\boldsymbol{V}_R$ and $\mathbf{A}$ such that $\widehat{\boldsymbol{X}} \approx \mathbf{A}\boldsymbol{V}_R^T$ subject to the constraints (3) for $\mathbf{A}$. This scaling procedure can be achieved using the Bayesian algorithm presented in [14] for unsupervised SU assuming the LMM (see [11] for details). Once the final abundance matrix $\widehat{\mathbf{A}}$ and the matrix $\widehat{\boldsymbol{V}}_R$ have been estimated, we propose an endmember extraction procedure based on GP regression. This method is discussed in the next section.

4. GAUSSIAN PROCESS REGRESSION

This section studies a new endmember estimation strategy based on GP regression for nonlinear mixtures. This strategy can be used even when the scene does not contain pure pixels. It assumes that all the image abundances have been estimated using the algorithm described in Section 3. Consider the set of pixels $\{\mathbf{y}(n)\}_{n=1,\ldots,N}$ and the corresponding estimated abundance vectors $\{\hat{\boldsymbol{a}}(n)\}_{n=1,\ldots,N}$. GP regression first allows the nonlinear mapping $\boldsymbol{g}(\cdot)$ in (1) to be estimated. The estimated mapping is denoted as $\hat{\boldsymbol{g}}(\cdot)$. Then, it is possible to use the prediction capacity of GPs to predict the spectrum $\boldsymbol{z}^* = [z_1^*, \ldots, z_L^*]^T = \hat{\boldsymbol{g}}(\boldsymbol{a}^*)$ corresponding to any new abundance vector $\boldsymbol{a}^*$. In particular, the predicted spectra associated with pure pixels, i.e., the endmembers, correspond to abundance vectors that are the vertices of the simplex defined by (3). The prediction property of GPs leads to

$$z_\ell^*|\mathbf{y}_\ell \sim \mathcal{N}\left(z_\ell^*|\mu_\ell, s_l^2\right), \quad \ell = 1, \ldots, L \quad (14)$$

with

$$\begin{aligned} \mu_\ell &= \boldsymbol{\psi}_x^{*T}\boldsymbol{U}\bar{\boldsymbol{p}}_\ell + \boldsymbol{\kappa}(\boldsymbol{a}^*)^T(\boldsymbol{K} + \sigma^2\mathbf{I}_N)^{-1}(\mathbf{y}_\ell - \boldsymbol{\Psi}_x\boldsymbol{U}\hat{\boldsymbol{p}}_\ell) \\ s_l^2 &= \sigma_{\boldsymbol{a}^*}^2 - \boldsymbol{\kappa}(\boldsymbol{a}^*)^T(\boldsymbol{K} + \sigma^2\mathbf{I}_N)^{-1}\boldsymbol{\kappa}(\boldsymbol{a}^*) \end{aligned}$$

$\boldsymbol{K} = s^2\boldsymbol{\Psi}_x\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{\Psi}_x^T$, $\sigma_{\boldsymbol{a}^*}^2 = s^2\boldsymbol{\psi}_x^{*T}\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{\psi}_x^*$, $\boldsymbol{\psi}_x^* = \boldsymbol{\psi}\left[\boldsymbol{V}_R\boldsymbol{a}^*\right]$ and $\boldsymbol{\kappa}(\boldsymbol{a}^*) = s^2\boldsymbol{\psi}_x^{*T}\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{\Psi}_x$. Since the posterior distribution (14) is Gaussian, the MAP and MMSE estimators of $\boldsymbol{z}^*$ are equal to the posterior mean $\boldsymbol{\mu} = (\mu_1, ..., \mu_L)^T$. In order to estimate the rth endmember, we propose to compute the vector $\boldsymbol{\mu}$ associated with the abundance vector $\boldsymbol{a}^* = [\mathbf{0}_{r-1}^T, 1, \mathbf{0}_{R-r}^T]^T$. The next section presents some simulation results obtained for synthetic data.

5. SIMULATIONS ON SYNTHETIC DATA

The performance of the proposed GPLVM is first evaluated on three synthetic images of $N = 2500$ pixels. The $R = 3$ endmembers contained in these images have been extracted from the spectral libraries provided with the ENVI software [15] (i.e., green grass, olive green paint and galvanized steel metal). Additional simulations conducted with different endmembers are available in [11]. The first image I_1 has been generated according to the LMM. The second image I_2 is distributed according to the bilinear mixing model introduced in [2], referred to as the "Fan model" (FM). The third image I_3 has been generated according to the generalized bilinear model (GBM) studied in [5] with the nonlinearity parameters set to $\gamma_{1,2} = 0.9, \gamma_{1,3} = 0.5, \gamma_{2,3} = 0.3$ The abundance vectors $\boldsymbol{a}_n, n = 1, \ldots, N$ have been drawn from a uniform distribution in the set $\left\{\boldsymbol{a}\middle|\sum_{r=1}^R a_r = 1,\ 0.9 \geq a_r(n) \geq 0, \forall r \in \{1, \ldots, R\}\right\}$ to reflect the absence of pure pixels in the images. The noise variance has been fixed to $\sigma^2 = 10^{-4}$, which corresponds to a signal-to-noise ratio SNR $\approx$ 30dB related to the worst case for current spectrometers. The hyperparameter γ of the latent variable prior (8) has been fixed to $\gamma = 10^3$ and the number of neighbors for the LLE is $K = R$ for all the results presented in this paper. This choice for K is mainly motivated by the fact that only R neighbors are needed to perfectly reconstruct a pixel under the noise-free LMM with R endmembers.

Table 1. AREs ($\times 10^{-2}$): synthetic images.

	PCA	LL-GPLVM	SU	FCLL-GPLVM
I_1	**1.00**	**1.00**	1.14	**1.00**
I_2	1.06	**1.00**	1.57	**1.00**
I_3	1.03	**0.99**	1.12	**0.99**

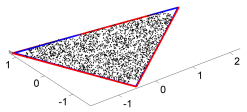
The quality of dimensionality reduction of the GPLVM can be measured by the average reconstruction error (ARE) defined as ARE $= \sqrt{\sum_{n=1}^N \|\hat{\mathbf{y}}_n - \mathbf{y}_n\|^2 / (NL)}$ where $\mathbf{y}_n$ is the nth observed pixel and $\hat{\mathbf{y}}_n$ its estimate. For the LL-GPLVM, the nth estimated pixel is given by $\hat{\mathbf{y}}_n = \widehat{\boldsymbol{P}}\widehat{\boldsymbol{U}}^T\boldsymbol{\psi}\left[\hat{\boldsymbol{x}}(n)\right]$ where $\widehat{\boldsymbol{P}}$ is estimated using (13). Table 1 compares the AREs obtained by the proposed LL-GPLVM and the projection onto the first $(R-1)$ principal vectors provided by the principal component analysis (PCA). The proposed LL-GPLVM slightly outperforms PCA for nonlinear mixtures in terms of ARE. More precisely, the AREs of the LL-GPLVM mainly consist of the noise errors ($\sigma^2 = 10^{-4}$), whereas model errors are added when applying PCA to nonlinear mixtures. The quality of unmixing pro-

Table 2. Abundance and endmember estimation: synthetic images.

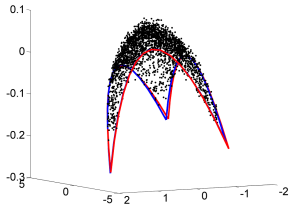
	RNMSE ($\times 10^{-3}$)			ASAM ($\times 10^{-2}$)		
	I_1	I_2	I_3	I_1	I_2	I_3
SU	49.3	86.6	47.8	2.37	6.53	4.97
FCLL-GPLVM	**4.8**	**7.2**	**7.5**	**0.64**	**0.89**	**0.80**

cedures can also be measured by comparing the estimated and actual abundances using the root normalized mean square error (RNMSE) defined by RNMSE $= \sqrt{\sum_{n=1}^{N} \|\hat{\boldsymbol{a}}_n - \boldsymbol{a}_n\|^2/(NR)}$ where $\boldsymbol{a}_n$ is the nth actual abundance vector and $\hat{\boldsymbol{a}}_n$ its estimate. Table 2 compares the RNMSEs obtained with different unmixing strategies. The endmembers have been estimated by the VCA algorithm for I_1 and the Heylen's method [10] for I_2 and I_3. The algorithms used for abundance estimation are the FCLS algorithm proposed in [16] for I_1, the LS method proposed in [2] for I_2 and the gradient-based method proposed in [5] for I_3. These procedures are referred to as "SU" in the table. These strategies are compared with the proposed FCLL-GPLVM. As mentioned above, the Bayesian algorithm [14] for the joint estimation of $\mathbf{A}$ and $\boldsymbol{V}_R$ is used in this paper for the scaling step. It can be seen that the proposed FCLL-GPLVM is general enough to accurately approximate the mixing models considered since it provides the best results in terms of abundance estimation.

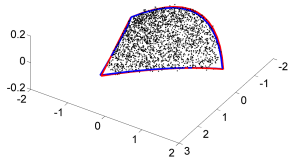
The quality of reconstruction of the unmixing procedure is also evaluated by the ARE. For the FCLL-GPLVM, the nth reconstructed pixel $\hat{\mathbf{y}}_n$ is given by $\hat{\mathbf{y}}_n = \widehat{\boldsymbol{P}}\widehat{\boldsymbol{U}}^T\boldsymbol{\psi}\left[\hat{\boldsymbol{x}}^{(c)}(n)\right]$. Table 1 shows the AREs corresponding to the different unmixing strategies. The proposed FCLL-GPLVM outperforms the other strategies in terms of ARE for these images. Finally, the performance of the FCLL-GPLVM for endmember estimation is evaluated by comparing the estimated endmembers with the actual ones using the spectral angle mapper (SAM) defined as SAM $= \arccos\left(\langle\hat{\mathbf{m}}_r, \mathbf{m}_r\rangle / (\|\hat{\mathbf{m}}_r\| \, \|\mathbf{m}_r\|)\right)$ where $\mathbf{m}_r$ is the rth actual endmember and $\hat{\mathbf{m}}_r$ its estimate. Table 2 compares the average SAMs (ASAMs) obtained by the SU procedures and the FCLL-GPLVM for the three images I_1 to I_3. Fig. 2 shows the actual (red lines) and estimated (blue lines) boundaries of the constrained set defined by the three mixing models and the three endmembers. These results show that the FCLL-GPLVM provides accurate endmember and abundance estimates for both linear and nonlinear mixtures.



(a) I_1 (LMM)



(b) I_2 (FM)



(c) I_3 (GBM)

Fig. 2. Visualization of the $N = 2500$ pixels (black dots) of I_1, I_2 and I_3 using the 3 axis provided by the PCA procedure. The actual (resp. estimated) endmembers correspond with the vertices of the red (resp. blue) lines.

6. CONCLUSIONS

We proposed a new algorithm for nonlinear spectral unmixing based on a Gaussian process latent variable model. The unmixing procedure assumed a nonlinear mapping from the constrained abundance space to the observed pixels. The abundance estimation was decomposed into two steps. Dimensionality reduction was first achieved using latent variables. A scaling procedure was then proposed to estimate the abundances. After estimating the abundance vectors of the image, a new endmember estimator based on Gaussian process regression was investigated. Simulations conducted on synthetic data illustrated the flexibility of the proposed model for linear and nonlinear spectral unmixing and provided promising results for abundance and endmember estimations even when there are few pure pixels in the image.

7. REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] J. M. P. Nascimento and J. M. Bioucas-Dias, "Nonlinear mixture model for hyperspectral unmixing," *Proc. of the SPIE*, vol. 7477, pp. 74770I–74770I–8, 2009.

[5] 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.

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

[7] 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.

[8] 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 Process.*, vol. 21, no. 6, pp. 3017–3025, 2012.

[9] J. Chen, C. Richard, and P. Honeine, "A novel kernel-based nonlinear unmixing scheme of hyperspectral images," in *Asilomar Conf. Signals, Systems, Computers*, Nov. 2011, pp. 1898–1902.

[10] R. Heylen, D. Burazerovic, and P. Scheunders, "Non-linear spectral unmixing by geodesic simplex volume maximization," *IEEE J. Sel. Top. Signal Process.*, vol. 5, no. 3, pp. 534–542, June 2011.

[11] Y. Altmann, N. Dobigeon, S. McLaughlin, and J.-Y. Tourneret, "Nonlinear spectral unmixing of hyperspectral images using Gaussian processes," *IEEE Trans. Signal Process.*, 2013, to appear.

[12] N. D. Lawrence, "The Gaussian process latent variable model," Tech. Rep., University of Sheffield, Department of Computer Science, Jan. 2006.

[13] N. D. Lawrence and J. Quiñonero Candela, "Local distance preservation in the gp-lvm through back constraints," in *Proceedings of the 23rd international conference on Machine learning*, New York, NY, USA, 2006, ICML '06, pp. 513–520, ACM.

[14] 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.

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

[16] 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.