

# UNSUPERVISED NONLINEAR UNMIXING OF HYPERSPECTRAL IMAGES USING GAUSSIAN PROCESSES

Yoann Altmann<sup>(1)</sup>, Nicolas Dobigeon<sup>(1)</sup>, Steve McLaughlin<sup>(2)</sup> and Jean-Yves Tourneret<sup>(1)</sup>

<sup>(1)</sup> University of Toulouse, IRIT-ENSEEIH, France

<sup>(2)</sup> School of Engineering and Electronics, University of Edinburgh, U.K.

{yoann.altmann, nicolas.dobigeon, jean-yves.tourneret}@enseeiht.fr, steve.mclaughlin@ee.ed.ac.uk

## ABSTRACT

This paper describes a Gaussian process based method for nonlinear hyperspectral image unmixing. The proposed model assumes a nonlinear mapping from the abundance vectors to the pixel reflectances contaminated by an additive white Gaussian noise. The parameters involved in this model satisfy physical constraints that are naturally expressed within a Bayesian framework. The proposed abundance estimation procedure is applied simultaneously to all pixels of the image by maximizing an appropriate posterior distribution which does not depend on the endmembers. After determining the abundances of all image pixels, the endmembers contained in the image are estimated by using Gaussian process regression. The performance of the resulting unsupervised unmixing strategy is evaluated through simulations conducted on synthetic data.

**Index Terms**— Nonlinear unmixing, hyperspectral images, Gaussian Processes.

## 1. INTRODUCTION

Spectral unmixing (SU) is a major issue that arises when analyzing hyperspectral images. SU consists of identifying the macroscopic materials present in an hyperspectral image and the proportions of these materials for all image pixels. Most SU strategies assume that pixel reflectances are linear combinations of pure component spectra [1]. The resulting linear mixing model (LMM) has been widely adopted in the literature and has provided interesting results. However, as discussed in [1], 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. Various nonlinear mixing models have been recently studied in the literature. More specifically, the bidirectional reflectance-based model proposed in [2] has been introduced for hyperspectral images defined by intimate mixtures. Bilinear models recently studied in [3,4] have shown interesting properties for images subjected to scattering effects, i.e., observed in wooded areas. Other more flexible unmixing techniques have also been proposed to handle wider class of nonlinearities, including radial basis function networks [5,6] and kernel-based models [7]. This paper considers a kernel-based approach for unsupervised SU based on a nonlinear dimension reduction method referred to as Gaussian process latent variable model (GP-LVM). The main advantage of the GP-LVM is its capacity to accurately model any nonlinearity. Note that “unsupervised” means that the endmembers contained in the image are not known *a priori*.

Part of this work has been supported by Direction Generale de l’Armement, French Ministry of Defence.

As a consequence, the model parameters to be estimated are the endmembers and the abundances of all pixels of the image. In the last decades, many endmember extraction algorithms (EEAs) have been developed to identify the pure spectral components contained in a hyperspectral image. However, most of these algorithms are based on the LMM and can be inefficient in handling nonlinear mixtures of endmembers. Conversely, the algorithm studied in this paper estimates the image endmembers using the regression capacity of Gaussian processes (GPs) in addition to the image pixel abundances.

The paper is organized as follows. Section 2 introduces the GP-LVM for dimensionality reduction. Section 3 introduces a constrained latent model for abundance estimation in hyperspectral imagery. Section 4 studies a new endmember estimation procedure using Gaussian process regression. Some simulation results conducted on synthetic data are shown and discussed in Section 5. Conclusions are reported in Section 6.

## 2. GAUSSIAN PROCESS LATENT VARIABLE MODEL

The GP-LVM is a powerful approach for probabilistic nonlinear dimensionality reduction [8] which generalizes probabilistic PCA by introducing nonlinearities through kernel functions. The linear GP-LVM assumes the following linear mapping

$$\mathbf{y}(n) = \mathbf{W}\mathbf{x}(n) + \mathbf{e}(n), \quad n = 1, \dots, N$$

between the  $N$  observation vectors  $\mathbf{y}(1), \dots, \mathbf{y}(N)$  of size  $L \times 1$  and  $N$  so-called *latent variables*  $\mathbf{x}(1), \dots, \mathbf{x}(N)$  of size  $d \times 1$ , where  $d$  is the dimension of the latent space and  $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_L]^T$  is an  $L \times d$  mixing matrix. We assume that the observations  $\mathbf{y}(n)$  have been centered and whitened (i.e.,  $E[\mathbf{y}(n)] = 0$  and  $E[\mathbf{y}^2(n)] = 1$ ) and that  $\mathbf{e}(n)$  is a Gaussian noise vector such that  $\mathbf{e}(n) \sim \mathcal{N}(\mathbf{0}_L, \sigma^2 \mathbf{I}_L)$  for  $n = 1, \dots, N$ . Assuming independence between the noise vectors, the distribution of the  $N \times L$  observation matrix  $\mathbf{Y} = [\mathbf{y}(1), \dots, \mathbf{y}(N)]^T$  conditioned on  $\mathbf{W}, \mathbf{X}, \sigma^2$  is

$$\mathbf{Y} | \mathbf{W}, \mathbf{X}, \sigma^2 \sim \prod_{n=1}^N \mathcal{N}(\mathbf{W}\mathbf{x}(n), \sigma^2 \mathbf{I}_L) \quad (1)$$

where  $\mathbf{X} = [\mathbf{x}(1), \dots, \mathbf{x}(N)]^T$  is an  $N \times d$  latent variable matrix. Assigning independent Gaussian  $\mathcal{N}(\mathbf{0}_d, \sigma^2 \mathbf{I}_d)$  prior distributions to  $\mathbf{x}(1), \dots, \mathbf{x}(N)$ , marginalizing the joint distribution  $f(\mathbf{Y}, \mathbf{X} | \mathbf{W}, \sigma^2)$  over  $\mathbf{X}$  lead to the probabilistic PCA [9]. The idea of GP-LVMs is different and consists of assigning a prior to  $\mathbf{W}$  and integrating out  $\mathbf{W}$  from the posterior distribution of interest. More precisely, consider independent Gaussian priors for  $\mathbf{w}_1, \dots, \mathbf{w}_L$ , the joint prior of  $\mathbf{W}$  is

$$f(\mathbf{W}) = \left(\frac{1}{2\pi}\right)^{\frac{dL}{2}} \prod_{\ell=1}^L \exp\left[-\frac{1}{2}\|\mathbf{w}_\ell\|^2\right]. \quad (2)$$

Straightforward computations lead to the following result

$$\begin{aligned} f(\mathbf{Y}|\mathbf{X}, \sigma^2) &= \int f(\mathbf{Y}|\mathbf{W}, \mathbf{X}, \sigma^2) f(\mathbf{W}) d\mathbf{W} \\ &\propto |\mathbf{C}|^{-\frac{L}{2}} \exp\left[-\frac{1}{2}\text{tr}(\mathbf{C}^{-1}\mathbf{Y}\mathbf{Y}^T)\right] \end{aligned} \quad (3)$$

where  $\propto$  means ‘‘proportional to’’ and  $\mathbf{C} = \mathbf{X}\mathbf{X}^T + \sigma^2\mathbf{I}_N$ . The main advantage of this so-called dual representation of probabilistic PCA is that it can be easily generalized to nonlinear mapping with respect to (w.r.t.)  $\mathbf{X}$ . More precisely, the sample covariance matrix of the latent variables  $\mathbf{X}\mathbf{X}^T$  can be replaced by an  $N \times N$  matrix  $\mathbf{K} = [k_{i,j}]$  such that  $k_{i,j} = k(\mathbf{x}(i), \mathbf{x}(j))$  for  $i, j = 1, \dots, N$ , where  $k(\cdot, \cdot)$  is a kernel function. In this paper, we propose to use the Gaussian kernel defined by

$$k_{i,j} = \alpha \exp\left(-\frac{1}{2}\|\mathbf{x}(i) - \mathbf{x}(j)\|^2\right) \quad (4)$$

where the kernel bandwidth has been set to 1 without loss of generality and where  $\alpha$  is the variance of the the nonlinearly transformed variables. Note that in contrast to kernel-PCA which applies the kernel function to the data space, the GP-LVM assumes a nonlinear transformation of the latent space into a higher dimensional space. When using kernels, the following results can be obtained

$$\begin{aligned} f(\mathbf{Y}|\mathbf{X}, \alpha, \sigma^2) &\propto |\boldsymbol{\Sigma}|^{-\frac{L}{2}} \exp\left[-\frac{1}{2}\text{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{Y}\mathbf{Y}^T)\right] \\ &\propto \prod_{\ell=1}^L |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\mathbf{y}_\ell^T \boldsymbol{\Sigma}^{-1} \mathbf{y}_\ell\right] \end{aligned} \quad (5)$$

where  $\boldsymbol{\Sigma} = \mathbf{K} + \sigma^2\mathbf{I}_N$ ,  $\mathbf{y}_\ell$  contains all observations associated with the  $\ell$ th spectral band and  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_L]$ . It has been shown that the GP-LVM summarized in this section allows the data set to be represented using latent variables belonging to a subspace whose dimension can be lower than the dimension of a PCA for a given precision of representation [8]. The next section introduces a new abundance estimation method based on the latent variables of the GP-LVM used for nonlinear dimensionality reduction.

### 3. ABUNDANCE ESTIMATION USING THE GP-LVM

A fundamental assumption in hyperspectral imagery is that each observation vector  $\mathbf{y}(n)$  results from a mixture of  $R$  pure spectral components (endmembers)  $\mathbf{m}_1, \dots, \mathbf{m}_R$  of size  $L \times 1$  whose abundance vector  $\mathbf{a}(n) = [a_1(n), \dots, a_R(n)]^T$  satisfies the following positivity and sum-to-one constraints

$$a_r(n) \geq 0, \forall r = 1, \dots, R, \quad \text{and} \quad \sum_{r=1}^R a_r(n) = 1. \quad (6)$$

Because of these constraints, the data  $\mathbf{y}(1), \dots, \mathbf{y}(N)$  live in an  $(R-1)$ -dimensional manifold. Thus it seems reasonable to assume that the simplest latent variables  $\mathbf{x}_1, \dots, \mathbf{x}_N$  that can represent the data  $\mathbf{y}(1), \dots, \mathbf{y}(N)$  also belong to a simplex of dimension  $R-1$ , i.e.,  $d = R-1$ . However, these latent variables are not exactly the abundances since they do not necessarily satisfy positivity and sum-to-one constraints<sup>1</sup>. Thus, we propose to express the latent variables as a linear transformation of the abundances as follows

$$\mathbf{x}(n) = \mathbf{V}\mathbf{a}(n) \quad (7)$$

<sup>1</sup>Note that in the linear case, the latent variables provided by PCA also belong to a simplex of dimension  $R-1$  and differ from the abundances.

where  $\mathbf{V}$  is an  $(R-1) \times R$  matrix and  $\mathbf{a}(n)$  is the  $n$ th abundance vector satisfying the constraints (6). Replacing (7) in (4), we obtain

$$k_{i,j} = \alpha \exp\left(-\frac{1}{2}\|\mathbf{V}(\mathbf{a}(i) - \mathbf{a}(j))\|^2\right). \quad (8)$$

Using the results of Section 2, the distribution of  $\mathbf{Y}|\mathbf{A}, \mathbf{V}, \alpha, \sigma^2$  can be expressed as in (5) with  $\boldsymbol{\Sigma} = \mathbf{K} + \sigma^2\mathbf{I}_N$  and the elements of  $\mathbf{K}$  are defined in (8). We denote as  $\mathbf{A} = [\mathbf{a}(1), \dots, \mathbf{a}(N)]^T$  the  $N \times R$  abundance matrix and as  $\boldsymbol{\theta} = \{\mathbf{A}, \mathbf{V}, \alpha, \sigma^2\}$  the unknown parameters associated with the GP-LVM for abundance estimation. It has been shown in [10] that GP-LVMs preserve dissimilarities. In the SU context, it means that pixels that are spectrally different have different abundance vectors. However, preserving local distances is also interesting: spectrally close pixels are expected to have similar abundance vectors. Several approaches have been considered to preserve local distances using GP-LVMs. In this paper, we propose as in [11] to include the locally linear embedding (LLE) in the proposed GP-LVM. First, the  $K$  nearest neighbors  $\{\mathbf{y}(j)\}_{j \in \nu_i}$  of each observation vector  $\mathbf{y}(i)$  are computed using the Euclidian distance ( $\nu_i$  is the set of integers  $j$  such that  $\mathbf{y}(j)$  is a neighbor of  $\mathbf{y}(i)$ ). The weight matrix  $\boldsymbol{\Lambda} = [\lambda_{i,j}]$  of size  $N \times N$  leading to the best reconstruction of  $\mathbf{y}(i)$  from its neighbors is then estimated as

$$\hat{\boldsymbol{\Lambda}} = \arg \min_{\boldsymbol{\Lambda}} \sum_{i=1}^N \left\| \mathbf{y}(i) - \sum_{j \in \nu_i} \lambda_{i,j} \mathbf{y}(j) \right\|^2. \quad (9)$$

Note that the solution of (9) is easy to obtain in closed form since the criterion to optimize is a quadratic function of  $\boldsymbol{\Lambda}$ . Note also that the matrix  $\boldsymbol{\Lambda}$  is sparse since each pixel is only described by its  $K$  nearest neighbors. The locally linear patches obtained by the LLE can then be used to set the following prior for the abundance matrix

$$f(\mathbf{A}|\hat{\boldsymbol{\Lambda}}) \propto \exp\left[-\frac{1}{2s^2} \sum_{i=1}^N \left\| \mathbf{a}(i) - \sum_{j \in \nu_i} \hat{\lambda}_{i,j} \mathbf{a}(j) \right\|^2\right] \prod_{i=1}^N \mathbf{1}_{\mathcal{S}}(\mathbf{a}(i))$$

where  $s^2$  is an hyperparameter to be adjusted and  $\mathbf{1}_{\mathcal{S}}(\cdot)$  is the indicator function over the simplex  $\mathcal{S}$  defined by the constraints (6). The following non-informative priors are also assigned to  $\mathbf{V}$ ,  $\alpha$  and  $\sigma^2$

$$\begin{aligned} f(\mathbf{V}_{i,j}) &\propto \mathbf{1}_{(-\delta_{\mathbf{V}}, \delta_{\mathbf{V}})}(\mathbf{V}_{i,j}) \\ f(\alpha) &\propto \mathbf{1}_{(0, \delta_{\alpha})}(\alpha) \\ f(\sigma^2) &\propto \mathbf{1}_{(0, \delta_{\sigma^2})}(\sigma^2) \end{aligned}$$

where the intervals  $(-\delta_{\mathbf{V}}, \delta_{\mathbf{V}})$ ,  $(0, \delta_{\alpha})$  and  $(0, \delta_{\sigma^2})$  cover the possible values of the parameters  $\mathbf{V}_{i,j}$ ,  $\alpha$  and  $\sigma^2$ . Assuming a priori independence between  $\mathbf{A}$ ,  $\mathbf{V}$ ,  $\alpha$  and  $\sigma^2$ , the joint posterior distribution of  $\boldsymbol{\theta}$  can be expressed as

$$f(\boldsymbol{\theta}|\mathbf{Y}, \hat{\boldsymbol{\Lambda}}) \propto f(\mathbf{Y}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\hat{\boldsymbol{\Lambda}}) \quad (10)$$

where  $f(\boldsymbol{\theta}|\hat{\boldsymbol{\Lambda}}) = f(\mathbf{A}|\hat{\boldsymbol{\Lambda}}) f(\mathbf{V}) f(\alpha) f(\sigma^2)$ . Mainly due to the nonlinearity introduced through the kernel, a closed form expression for the Bayesian estimators of  $\boldsymbol{\theta}$  (minimum mean square error (MMSE) or maximum a posteriori (MAP) estimators) is difficult to obtain. Consequently, we propose to use a gradient-based method for maximizing the posterior (10) w.r.t.  $\boldsymbol{\theta}$  and computing the MAP estimator<sup>2</sup> of  $\boldsymbol{\theta}$ . To handle the constraints for  $\mathbf{A}$ , we pro-

<sup>2</sup>Note that for a given  $\mathbf{X}$ , there is an intrinsic ambiguity in (8) (and possibly in the posterior (10)) between the scales of the matrices  $\mathbf{A}$  and  $\mathbf{V}$ . However, this ambiguity can be removed by assuming that the actual abundances occupy the largest volume in  $\mathcal{S}$ . Thus, after estimating the unknown parameters  $\boldsymbol{\theta}$  by maximization of (10), the estimated abundance matrix is mapped via a linear transformation into a simplex with maximum volume whose vertices have abundance vectors  $[\mathbf{0}_{r-1}^T, 1, \mathbf{0}_{R-r}^T]^T$  for  $r = 1, \dots, R$ . The simplex with maximum volume containing the abundances is determined using minimum volume simplex analysis (MVSA) [12].

pose to use a subgradient-based algorithm updating sequentially the abundance vector  $\mathbf{a}(n)$ , the matrix  $\mathbf{V}$ , the scale parameter  $\alpha$  and the noise variance  $\sigma^2$ . The partial derivatives of the posterior (10) w.r.t.  $\mathbf{a}(n), n = 1, \dots, N$ ,  $\mathbf{V}$  and  $\sigma^2$  can be derived from the partial derivatives given in [10] and the usual chain rules. To satisfy the positivity and sum-to-one constraints, the partial derivatives w.r.t.  $\mathbf{a}(n)$  are projected onto the simplex defined by (6). Moreover, we have considered a bounded line search step size to ensure the abundance vector satisfies the positivity constraints. Note that the abundance estimation procedure studied in this section does not require the knowledge of the image endmembers. The next section shows that these endmembers can be estimated using GP regression.

#### 4. ENDMEMBER ESTIMATION USING GAUSSIAN PROCESS REGRESSION

This section studies a new endmember estimation strategy based on GP regression. This strategy assumes that all the image abundances have been estimated (e.g., by using the algorithm described in Section 3). Eq. (5) shows that  $f(\mathbf{Y}|\mathbf{X}, \alpha, \sigma^2)$  is the product of  $L$  independent GPs associated with each spectral band of the data space

$$\mathbf{y}_\ell \sim \mathcal{N}(\mathbf{0}_N, \mathbf{K} + \sigma^2 \mathbf{I}_N). \quad (11)$$

Looking carefully at the covariance matrix of  $\mathbf{y}_\ell$ , one can write

$$\mathbf{y}_\ell = \mathbf{z}_\ell + e_\ell \quad (12)$$

where  $e_\ell$  is the  $N \times 1$  white Gaussian noise vector associated with the  $\ell$ th spectral band (having covariance matrix  $\sigma^2 \mathbf{I}_N$ ) and

$$\mathbf{z}_\ell \sim \mathcal{N}(\mathbf{0}, \mathbf{K}) \quad (13)$$

with  $\mathbf{K}$  the  $N \times N$  covariance matrix of  $\mathbf{z}_\ell$  which is constructed from  $\boldsymbol{\theta}$  and (8). Thus the observed vector  $\mathbf{y}_\ell$  is related to a hidden vector  $\mathbf{z}_\ell$  associated with the  $\ell$ th components of all the training data such that

$$\mathbf{y}_\ell | \mathbf{z}_\ell \sim \mathcal{N}(\mathbf{z}_\ell, \sigma^2 \mathbf{I}_N). \quad (14)$$

Consider an  $L \times 1$  test data with hidden vector  $\mathbf{t} = [t(1), \dots, t(L)]^T$  and associated abundance vector<sup>3</sup>  $\mathbf{u} = [u(1), \dots, u(R)]^T$ . We assume that the test data share the same statistical properties as the training data  $\mathbf{y}_1, \dots, \mathbf{y}_L$  in the sense that  $[\mathbf{z}_\ell^T, t(\ell)]$  is a Gaussian vector such that

$$\begin{bmatrix} \mathbf{z}_\ell \\ t(\ell) \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \mathbf{0}_N \\ 0 \end{bmatrix}, \begin{bmatrix} \mathbf{K} & \boldsymbol{\kappa}(\mathbf{u}) \\ \boldsymbol{\kappa}(\mathbf{u})^T & \alpha \end{bmatrix} \right) \quad (15)$$

where  $\alpha$  is the variance of  $t(\ell)$  (note that according to (8), the variance of any element of  $\mathbf{z}_\ell$  also equals  $\alpha$ ) and  $\boldsymbol{\kappa}(\mathbf{u})$  contains the covariances between the training inputs and the test inputs, i.e.,

$$\boldsymbol{\kappa}_i(\mathbf{u}) = \alpha \exp \left( -\frac{1}{2} \|\mathbf{V}(\mathbf{a}(i) - \mathbf{u})\|^2 \right), \quad i = 1, \dots, N. \quad (16)$$

Note that the abundances  $\mathbf{u}$  have to be known to compute  $\boldsymbol{\kappa}(\mathbf{u})$ . Combining (15) with (14) leads to

$$t(\ell) | \mathbf{y}_\ell \sim \mathcal{N}(\mu_\ell, s_\ell^2) \quad (17)$$

with

$$\begin{aligned} \mu_\ell &= \boldsymbol{\kappa}(\mathbf{u})^T (\mathbf{K} + \sigma^2 \mathbf{I}_N)^{-1} \mathbf{y}_\ell \\ s_\ell^2 &= \alpha - \boldsymbol{\kappa}(\mathbf{u})^T (\mathbf{K} + \sigma^2 \mathbf{I}_N)^{-1} \boldsymbol{\kappa}(\mathbf{u}). \end{aligned}$$

<sup>3</sup> $\mathbf{u}$  will be set to  $[\mathbf{0}_{r-1}^T, 1, \mathbf{0}_{R-r}^T]^T$  for the estimation of the  $r$ th endmember.

Since the posterior distribution (17) is Gaussian, the MAP and MMSE estimators of  $\mathbf{t}$  equal the posterior mean  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_L)^T$ .

In order to estimate the endmembers, we propose to compute the estimated hidden vectors associated with the abundance vectors  $[\mathbf{0}_{r-1}^T, 1, \mathbf{0}_{R-r}^T]^T$  for  $r = 1, \dots, R$ . For each value of  $r$ , the  $r$ th estimated hidden vector will be the  $r$ th estimated endmember. Indeed, for the LMM and the nonlinear models considered in this paper, the endmembers are obtained by setting  $\mathbf{u} = [\mathbf{0}_{r-1}^T, 1, \mathbf{0}_{R-r}^T]^T$  in the model relating the observations to the abundances. Note that the proposed endmember estimation procedure provides the posterior distribution of each endmember via (17) which can be used to derive confidence intervals for the estimates.

#### 5. SIMULATIONS

We first evaluate the performance of GPs for abundance estimation by unmixing 3 synthetic images of 200 pixels. The  $R = 3$  endmembers contained in these images have been extracted from the spectral libraries provided with the ENVI software (i.e., green grass, olive green paint and galvanized steel metal). The first image  $I_1$  has been generated according to the LMM. The second image  $I_2$  has been generated according to the bilinear mixing model introduced in [3], referred to as ‘‘Fan model’’ (FM). The third image  $I_3$  has been constructed by using the generalized bilinear model (GBM) introduced in [4]. The abundance vectors  $\mathbf{a}(n), n = 1, \dots, 200$  have been randomly generated according to a uniform distribution concentrated on the admissible set defined by the positivity and sum-to-one constraints. The nonlinearity parameters of the GBM have been randomly generated according to a uniform distribution on the interval  $(0, 1)^R$ . The noise variance and the variance involved in the abundance prior  $f(\mathbf{A}|\hat{\boldsymbol{\Lambda}})$  have been fixed to  $\sigma^2 = s^2 = 10^{-4}$ . The number of nearest neighbors required to define the neighborhoods  $\nu_i$  in  $f(\mathbf{A}|\hat{\boldsymbol{\Lambda}})$  has been fixed to  $K = 3$ . Finally, the hyperparameters in (10) have been fixed to  $\delta_V = \delta_\alpha = \delta_{\sigma^2} = 10^4$  in order to cover all possible values of  $\mathbf{V}_{i,j}, \alpha$  and  $\sigma^2$ . The gradient algorithm used to maximize the posterior distribution (10) requires endmembers and abundances to be initialized. The initial endmembers have been computed using the vertex component analysis (VCA) algorithm [13] which provides good results even in the case of nonlinear mixtures. The initial abundance matrix has then been computed using the fully constrained least-squares (FCLS) algorithm. The number of iterations for the optimization algorithm has been set to  $N_{\text{iter}} = 200$  (which has ensured algorithm convergence in all experiments). After the optimization step, the abundances have been rescaled using the MVSA method as discussed in Section 3.

The performance of GPs for abundance estimation has been first observed by comparing the reconstructed pixels (computed from the estimated abundance vectors) with the observed spectra. More precisely, the quality of reconstruction has been evaluated from the average reconstruction error (ARE) defined as  $\text{ARE} = \sqrt{\sum_{n=1}^N \|\hat{\mathbf{y}}(n) - \mathbf{y}(n)\|^2 / LN}$  where  $\mathbf{y}(n)$  is the  $n$ th observed pixel and  $\hat{\mathbf{y}}(n)$  the reconstructed pixel obtained using GP regression by setting  $\mathbf{u} = \hat{\mathbf{a}}(n)$  in (17). Table 1 compares the AREs obtained using the proposed algorithm (referred to as GP-LVM) and the VCA algorithm for the three images  $I_1, I_2$  and  $I_3$ . These results show that the GP-LVM is flexible enough to approximate the different mixing models since it provides either the first (bold underlined) or second (bold) best results in term of ARE. The quality of the unmixing procedures can also be measured by comparing the estimated and actual abundances using the root mean square error (RMSE) defined by  $\text{RMSE} = \sqrt{\sum_{n=1}^N \|\hat{\mathbf{a}}(n) - \mathbf{a}(n)\|^2 / NR}$

where  $\mathbf{a}(n)$  is the  $n$ th actual abundance vector and  $\hat{\mathbf{a}}(n)$  its estimate. Table 2 compares the RMSEs obtained with the proposed algorithm (GP-LVM) and other existing algorithms (referred to as FCLS [14], Fan [3] and Halimi [4]). Note that the GP-LVM algorithm does not require to estimate the endmembers. The other algorithms (FCLS, FAN, GBM) have been run with endmembers estimated with the VCA algorithm. Table 2 shows that the proposed algorithm is general enough to accurately approximate the considered mixing models since it provides either the first or second best results for abundance estimation.

The last experiments evaluate the performance of GPs for endmember estimation. The quality of endmember estimation has been evaluated by the mean square error (MSE) defined as  $MSE = \sqrt{\|\hat{\mathbf{m}}_r - \mathbf{m}_r\|^2 / L}$  where  $\mathbf{m}_r$  is the  $r$ th actual endmember and  $\hat{\mathbf{m}}_r$  its estimate. Table 3 compares the REs obtained for each endmember using the GP-LVM and the VCA algorithms for the three images  $I_1$  to  $I_3$ . These results show that the GP-LVM provides accurate endmember estimates for both linear and nonlinear mixtures. Fig. 1 shows the MMSE estimates of the three endmembers obtained using VCA and the proposed GP-LVM algorithm. These results are very promising.

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

	FCLS [14]	Fan [3]	Halimi [4]	GP-LVM
$I_1$ (LMM)	<b>0.99</b>	3.06	1.05	<b>1.00</b>
$I_2$ (FM)	2.36	1.83	<b>1.05</b>	<b>0.98</b>
$I_3$ (GBM)	1.38	2.31	<b>1.05</b>	<b>0.98</b>

**Table 2.** RMSEs ( $\times 10^{-2}$ ).

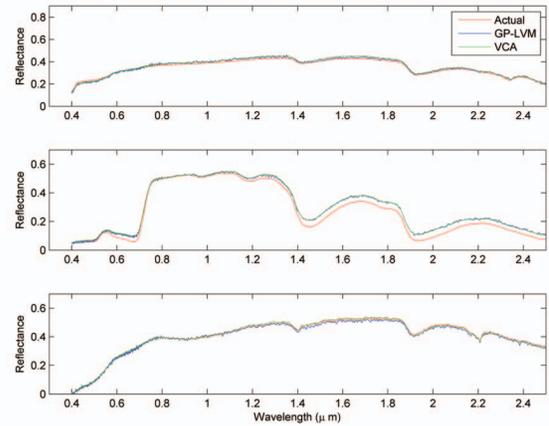
	FCLS [14]	Fan [3]	Halimi [4]	GP-LVM
$I_1$ (LMM)	<b>1.93</b>	4.22	2.04	<b>0.67</b>
$I_2$ (FM)	4.22	<b>1.63</b>	2.77	<b>1.95</b>
$I_3$ (GBM)	2.80	2.72	<b>1.73</b>	<b>1.35</b>

**Table 3.** MSEs ( $\times 10^{-2}$ ).

		VCA [14]	MVSA [12]	GP-LVM
$I_1$ (LMM)	$\mathbf{m}_1$	<b>0.29</b>	<b>0.21</b>	<b>0.29</b>
	$\mathbf{m}_2$	<b>1.65</b>	<b>0.40</b>	<b>1.65</b>
	$\mathbf{m}_3$	0.48	<b>0.25</b>	<b>0.41</b>
$I_2$ (FM)	$\mathbf{m}_1$	<b>1.25</b>	2.79	<b>1.09</b>
	$\mathbf{m}_2$	<b>2.96</b>	3.19	<b>2.85</b>
	$\mathbf{m}_3$	<b>0.60</b>	3.76	<b>0.92</b>
$I_3$ (GBM)	$\mathbf{m}_1$	1.61	<b>0.97</b>	<b>0.72</b>
	$\mathbf{m}_2$	<b>0.98</b>	1.02	<b>0.66</b>
	$\mathbf{m}_3$	<b>0.56</b>	2.13	<b>0.61</b>

## 6. CONCLUSIONS

This paper has proposed a new unsupervised algorithm based on Gaussian processes for nonlinear spectral unmixing. The main idea of the proposed strategy was to express the hyperspectral image pixels as functions of latent variables whose covariances depend on abundances via appropriate kernels. Positivity and sum-to-one constraints for the abundance vectors were included in the estimation procedure. A new endmember estimator based on Gaussian process regression was finally introduced. Simulations conducted on synthetic images provided very promising results. Future works include the design of appropriate kernels for nonlinear spectral unmixing. Considering other forms of spatial correlation between adjacent pixels of the image is also an interesting field of research.



**Fig. 1.** Estimated endmembers using VCA (green) and GP-LVM (blue) compared with the actual endmembers (red) for the image  $I_2$ .

## 7. REFERENCES

- [1] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 44–57, Jan. 2002.
- [2] B. W. Hapke, "Bidirectional reflectance spectroscopy. I. Theory," *J. Geophys. Res.*, vol. 86, pp. 3039–3054, 1981.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] Y. Altmann, N. Dobigeon, S. McLaughlin, and J.-Y. Tourneret, "Non-linear unmixing of hyperspectral images using radial basis functions and orthogonal least squares," in *Proc. IEEE Int. Conf. Geosci. and Remote Sensing (IGARSS)*, Vancouver, Canada, 2011.
- [7] 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)*, July 2007, pp. 4041–4044.
- [8] N. D. Lawrence, "Gaussian process latent variable models for visualization of high dimensional data," in *NIPS*, Vancouver, Canada, 2003.
- [9] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *Journal of the Royal Statistical Society, Series B*, vol. 61, pp. 611–622, 1999.
- [10] N. D. Lawrence, "The Gaussian process latent variable model," University of Sheffield, Department of Computer Science, Tech. Rep., Jan. 2006. [Online]. Available: <http://staffwww.dcs.shef.ac.uk/people/N.Lawrence/>
- [11] R. Urtasun, D. J. Fleet, and N. D. Lawrence, "Modeling human locomotion with topologically constrained latent variable models," in *Conf. Human motion: understanding, modeling, capture and animation*, Rio de Janeiro, Brazil, 2007, pp. 104–118.
- [12] J. Li and J. M. Bioucas-Dias, "Minimum volume simplex analysis: a fast algorithm to unmix hyperspectral data," in *Proc. IEEE Int. Conf. Geosci. and Remote Sensing (IGARSS)*, vol. 3, Boston, USA, July 2008, pp. 250–253.
- [13] J. M. P. 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.
- [14] 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.