Nonparametric Detection of Nonlinearly Mixed Pixels and Endmember Estimation in Hyperspectral Images

Tales Imbiriba, *Student Member, IEEE*, José Carlos Moreira Bermudez, *Senior Member, IEEE*
Cédric Richard, *Senior Member, IEEE*, Jean-Yves Tourneret, *Senior Member, IEEE*

Abstract—**Mixing phenomena in hyperspectral images depend on a variety of factors such as the resolution of observation devices, the properties of materials, and how these materials interact with incident light in the scene. Different parametric and nonparametric models have been considered to address hyperspectral unmixing problems. The simplest one is the linear mixing model. Nevertheless, it has been recognized that mixing phenomena can also be nonlinear. The corresponding nonlinear analysis techniques are necessarily more challenging and complex than those employed for linear unmixing. Within this context, it makes sense to detect the nonlinearly mixed pixels in an image prior to its analysis, and then employ the simplest possible unmixing technique to analyze each pixel. In this paper, we propose a technique for detecting nonlinearly mixed pixels. The detection approach is based on the comparison of the reconstruction errors using both a Gaussian process regression model and a linear regression model. The two errors are combined into a detection statistics for which a probability density function can be reasonably approximated. We also propose an iterative endmember extraction algorithm to be employed in combination with the detection algorithm. The proposed detect-then-unmix strategy, which consists of extracting endmembers, detecting nonlinearly mixed pixels and unmixing, is tested with synthetic and real images.**

Index Terms—**Hyperspectral imaging, spectral unmixing, endmember extraction, non-linearity detection, Gaussian processes.**

I. INTRODUCTION

Emerged in the 1960s with multispectral scanners, modern hyperspectral sensors produce two-dimensional hyperspectral images over a few tens to thousands of contiguous spectral bands [2]. Their high spectral resolution allows a comprehensive and quantitative analysis of materials in remotely observed data. This area has received considerable attention in the last decade. Due to historic downlink and computer processing limitations [3], hyperspectral images often trade spatial for spectral resolution [4]. Such trade-off is especially evident in remote sensing applications, and caused by the large distance between sensors and target scenes. The observed reflectances then result from spectral mixtures of several pure material signatures. As a consequence, spectral unmixing has become an important issue for hyperspectral data processing [5].

In a supervised setting, the spectral signatures of pure materials are available as vectors of reflectances of these materials at each wavelength. Such vectors are typically called endmembers due to their geometrical interpretation in the linear mixing case. Mixing phenomena depend on a variety of factors such as the resolution of observation devices, the properties of materials, and how these materials interact with incident light in the scene [6]. Therefore, different parametric and nonparametric models have been considered to address hyperspectral unmixing problems. The simplest one is the linear mixing model, which assumes linear mixing of the endmember contributions [5]. However, it has been recognized that mixing phenomena can often be nonlinear [5], [6]. Unmixing nonlinearly mixed pixels requires specific techniques, and thus nonlinear analysis of hyperspectral images has been widely explored in the past few years. See, for instance, [2], [6]–[16]. It is now acknowledged that nonlinear unmixing algorithms can lead to a better understanding of the individual spectral contributions. On the other hand, nonlinear analysis techniques are necessarily more challenging and complex than those employed for linear unmixing. As hyperspectral images tend to include both linearly and nonlinearly mixed pixels, there are two important reasons to match the unmixing method to the nature of each pixel in the image. First, nonlinear unmixing algorithms are always more complex to implement than linear unmixing algorithms. Second, unmixing linearly mixed pixels with nonlinear unmixing algorithms leads to poorer results than doing it with linear unmixing algorithms. Hence, it makes sense to detect the nonlinearly mixed pixels in an image prior to its analysis, and then employ the simplest and more accurate available unmixing technique to analyze each pixel. To this end, it is desirable to devise analysis techniques that combine endmember extraction, detection of nonlinearly mixed pixels and unmixing.

The problems of extracting endmembers, detecting nonlinearly mixed pixels and unmixing are interlaced, and addressing them jointly is not a trivial task. For instance, most nonlinear unmixing techniques assume the endmembers to be known

or to be estimated by an endmember extraction algorithm (EEA) [13]–[21]. However, most endmember extraction algorithms rely on the convex geometry associated with the linear mixing model [22]–[26], which obviously does not apply to nonlinearly mixed pixels. Endmember extraction techniques designed for situations where a significant part of the image is composed of nonlinear mixtures are rarely addressed in the literature. In fact, most of the techniques considering nonlinearly mixed pixels are part of a complete unsupervised unmixing strategy [27], [28]. Detecting nonlinearly mixed pixels in an hyperspectral image is also a complex task. Physically motivated models [7], [29] usually tend to be too complex for application in practical detection strategies. One possible approach is to consider a simplified parametric model for the nonlinearity. The parameters of this nonlinear model are then estimated from the image, and hypothesis tests are derived based on these estimates. For instance, a single-parameter polynomial post-nonlinear model is assumed in [30]. The main question regarding parametric modeling of nonlinear mixing mechanisms is whether the chosen model can capture the actual nonlinear effects present in a scene. When nothing or little is known about the nonlinear mixing mechanism, a direct strategy is to exploit the property of linear mixing models to confine the noiseless data to a simplex. The hypothesis test proposed in [31] is based on the distance between the observed pixel and this simplex. Though this test is robust to nonlinear mixing mechanisms, it conveys too little information about the nonlinearity as a tradeoff to guarantee simplicity. An alternative strategy is to use nonparametric techniques to extract information about the nonlinearity directly from the observations. A nonparametric unmixing technique based on kernel expansions is presented in [13], but this work does not address nonlinearity detection. A nonlinear mixing model for joint unmixing and nonlinearity detection is proposed in [32]. It assumes that the observed reflectances result from linear spectral mixtures corrupted by a residual nonlinear component. This model is rather similar to the model initially introduced in [13], but the estimation method relies on a computationally intensive Bayesian procedure.

All the detection methods discussed above assume known endmember spectral signatures. In most cases, the endmembers are assumed to have been estimated from the data. However, most endmember extraction algorithms exploit the convex geometry of the linear mixing model and assume the presence of pure endmember pixels in the image. They usually exploit one of the following properties: 1) the endmembers are the extreme points when projecting the data onto any subspace [22]–[24], [26], [33], [34]; 2) the volume of a simplex spanned by any subset of points in the image is maximum when these points are the endmembers [35], [36]. Other strategies deal with nonnegative matrix factorization [37]–[39]. One method of particular interest to this work is the MVES (minimum volume enclosing simplex) algorithm [25] as it does not assume the presence of pure pixels, though it still exploits the geometry of linear mixtures. It solves a constrained least-squares optimization problem with a simplex volume regularizer. Practically, MVES finds the smallest simplex circumscribing the hyperspectral data. The vertices of this simplex are defined as the endmembers. One important aspect of using one of these endmember extraction techniques when analyzing images composed of linearly and nonlinearly mixed pixels is that they tend to lead to poor endmember estimates when applied to nonlinearly mixed pixels.

In this paper, we propose a technique that combines endmember extraction and detection of nonlinearly mixed pixels in hyperspectral images. The detection approach is based on the comparison of the reconstruction errors using both a Gaussian process (GP) and a linear regression model. The two errors are combined into a detection statistics for which a probability density function can be reasonably approximated. We also propose an MVES-based iterative endmember extraction algorithm to be employed in combination with the detection algorithm to jointly detect nonlinearly mixed pixels and extract the image endmembers from the linearly mixed pixels. The proposed method is tested with synthetic and real images. This work is organized as follows. Section II reviews the linear mixing model and some nonlinear mixing models, and discusses different ways of modeling nonlinear interaction between light and endmembers. Section III discusses GP regression applied in the context of hyperspectral data unmixing. Section IV presents the application of GP to the detection of nonlinearly mixed pixels. Section V introduces a two-step iterative procedure to estimate the endmember matrix. This method combines the MVES algorithm and the nonlinear mixture detector proposed in Section IV. Simulations with synthetic and real data are presented in Section VI. Conclusions are finally presented in Section VII.

A. Contributions of the present work

The main contributions of this works are the following:

a)   a model-free detector of nonlinearly mixed pixels. The novel test statistics compares reconstruction errors of the observations modeled by a Gaussian Process and a linear regression;

b)   a novel recursive endmember estimation algorithm for scenes that are partly nonlinear;

c)   the definition of a degree of nonlinearity η_d which allows a meaningful comparison of detection results using images obtained using different mixing models.

We emphasize the need for such techniques when considering nonlinear effects in the mixing process, as conventional endmember extraction techniques strongly rely on the convex geometry associated with the linear mixing model.

II. MIXTURE MODELS

Each observed pixel can be written as a function of the endmembers plus an additive term associated with the measurement noise and the modeling error. Consider the model

$$\boldsymbol{r} = \boldsymbol{\psi}(\boldsymbol{M}) + \boldsymbol{n} \tag{1}$$

where $\boldsymbol{r} = [r_1, \ldots, r_L]^\top$ is a vector of reflectances observed in L spectral bands, $\boldsymbol{M} = [\boldsymbol{m}_1, \ldots, \boldsymbol{m}_R]$ is the $L \times R$ endmember matrix, whose i-th column $\boldsymbol{m}_i$ is an endmember,

$n \sim \mathcal{N}(\mathbf{0}, \sigma_n^2 \boldsymbol{I})$ is a white Gaussian noise (WGN) vector, and $\boldsymbol{\psi}$ is an unknown mixing function. Several models of the form (1) have been proposed in the literature, depending on the linearity or nonlinearity of $\boldsymbol{\psi}$, type of mixture, and other properties [6].

A. The linear mixing model

The linear mixing model assumes that each light ray interacts only with one material, disregarding multiple interactions between light and multiple materials [5]. The classical linear model assumes that $\boldsymbol{\psi}$ is a convex combination of the endmembers. The vector $\boldsymbol{r}$ can then be written as

$$\boldsymbol{r} = \boldsymbol{M}\boldsymbol{\alpha} + \boldsymbol{n}$$
$$\text{subject to } \mathbf{1}^\top \boldsymbol{\alpha} = 1 \text{ and } \boldsymbol{\alpha} \succeq \mathbf{0} \tag{2}$$

where $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_R]^\top$ is the vector of abundances of each endmember in $\boldsymbol{M}$, R is the number of endmembers, and $\succeq$ denotes the entrywise $\geq$ operator. Therefore, the entries of $\boldsymbol{\alpha}$ cannot be negative and should sum to one. The observation r_ℓ in the ℓ-th wavelength of (2) can be written as

$$r_\ell = \boldsymbol{m}_{\lambda_\ell}^\top \boldsymbol{\alpha} + n_\ell \tag{3}$$

where $\boldsymbol{m}_{\lambda_\ell}$ denotes the ℓ-th row of $\boldsymbol{M}$ as a column vector. In the noiseless case, namely, $n_\ell = 0$, the sum-to-one and positivity constraints over the abundances in (2) confine the data to a simplex. The vertices of this simplex are the endmembers, which justifies the terminology.

Several parametric models have been proposed in the literature to describe nonlinear mixing mechanisms of endmembers in hyperspectral images. See [6] and references therein. We shall now review two popular models that will be used later to generate synthetic data for evaluation purposes.

B. Nonlinear mixing models

The generalized bilinear model (GBM) [18] is given by

$$\boldsymbol{r} = \boldsymbol{M}\boldsymbol{\alpha} + \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} \gamma_{ij}\, \alpha_i \alpha_j\, \boldsymbol{m}_i \odot \boldsymbol{m}_j + \boldsymbol{n}$$
$$\text{subject to } \mathbf{1}^\top \boldsymbol{\alpha} = 1 \text{ and } \boldsymbol{\alpha} \succeq 0 \tag{4}$$

where the parameters $\gamma_{ij} \in [0,1]$ govern the amount of nonlinear contribution, and $\odot$ denotes the Hadamard product. In the noiseless case, data following the model (4) lie in a nonlinearly distorted simplex in $\mathbb{R}^R$ whose vertices are the endmembers as in the linear case. For simplicity, here we consider a simplified version of this model in which the nonlinear contribution is controlled by a single parameter γ such that $\gamma = \gamma_{ij}$ for all $(i \neq j)$.

The post nonlinear mixing model (PNMM) [40] is given by

$$\boldsymbol{r} = \boldsymbol{g}(\boldsymbol{M}\boldsymbol{\alpha}) + \boldsymbol{n} \tag{5}$$

where $\boldsymbol{g}$ is a nonlinear function applied to the linear mixing model. The PNMM can represent a wide range of nonlinear mixing models, depending on the definition of $\boldsymbol{g}$. For instance, the PNMM considered in [13] is given by

$$\boldsymbol{r} = (\boldsymbol{M}\boldsymbol{\alpha})^\xi + \boldsymbol{n} \tag{6}$$

where $(\boldsymbol{v})^\xi$ denotes the exponentiation applied to each entry of the vector $\boldsymbol{v}$. For $\xi = 2$, (6) becomes a bilinear model closely related to the GBM but without a linear term. The PNMM was explored in other works considering different forms for $\boldsymbol{g}$ applied to hyperspectral data unmixing [30], [41].

The GBM (4) and the PNMM (5) nonlinear mixing models mainly represent the scattering phenomenon where the light first interacts with an endmember, and then with a second one, before being captured by the hyperspectral sensor. Other models account for other kinds of interaction between light and endmembers, or consider other types of nonlinear effects. In the case of the intimate mixture model [42] for instance, the endmembers are considered to be mixed at the molecular level. Other nonlinear models can be considered depending on the characteristics of the scene [7]–[9], [11], [18], [20], [29], [40], [42]. More importantly, these informations are usually missing. Hence, it makes sense to develop nonparametric models that do not make strong assumptions about the type of nonlinearity involved in the mixture.

III. NONLINEARITY DETECTION WITH GAUSSIAN PROCESS REGRESSION MODELS

To detect nonlinearly mixed pixels in an hyperspectral image, assuming $\boldsymbol{\psi}$ in (1) is unknown, we propose to compare the reconstruction errors resulting from estimating $\boldsymbol{\psi}$ with nonlinear and linear regression methods. Gaussian process (GP) regression methods consist of defining stochastic models for functions and performing inference in functional spaces [43]. The representation is rigorous but, at the same time, lets the data speak for themselves. This characteristic is desirable when little is known about the functions to be estimated. Using some knowledge obtained from the observations about the endmember matrix, we propose a supervised learning strategy to make inference on $\boldsymbol{\psi}$.

This section describes the application of GP nonlinear regression to the problem at hand. Consider the training set $\{\boldsymbol{M}, \boldsymbol{r}\}$ with inputs $\boldsymbol{M} = [\boldsymbol{m}_{\lambda_1}, \ldots, \boldsymbol{m}_{\lambda_L}]^\top$, and outputs or observations $\boldsymbol{r} = [r_1, \ldots, r_L]^\top$. By analogy with the linear mixing model (3), we write the ℓ-th row of (1) as

$$r_\ell = \psi(\boldsymbol{m}_{\lambda_\ell}) + n_\ell, \tag{7}$$

with r_ℓ the ℓ-th entry of the observation $\boldsymbol{r}$, ψ a real-valued function in a (reproducing kernel) Hilbert space $\mathcal{H}$, and n_ℓ an additive WGN in the ℓ-th band. A Gaussian process is a collection of random variables, any finite number of which has a joint Gaussian distribution [43]. We define a Gaussian prior for ψ with mean and covariance functions given by

$$\mathbb{E}\{\psi(\boldsymbol{m}_{\lambda_\ell})\} = 0$$
$$\mathbb{E}\{\psi(\boldsymbol{m}_{\lambda_\ell})\psi(\boldsymbol{m}_{\lambda_{\ell'}})\} = \kappa(\boldsymbol{m}_{\lambda_\ell}, \boldsymbol{m}_{\lambda_{\ell'}}) \tag{8}$$

where κ is a positive definite kernel. For notational simplicity, it is common but not necessary to consider GPs with a zero mean function. This assumption is not overly restricting as the mean of the posterior distribution is not confined to be zero (as shown by (11)). The prior on the noisy observation $\boldsymbol{r}$ becomes

$$\boldsymbol{r} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{K} + \sigma_n^2 \boldsymbol{I}_L) \tag{9}$$

with $\boldsymbol{K}$ the Gram matrix whose entries $\boldsymbol{K}_{ij} = \kappa(\boldsymbol{m}_{\lambda_i}, \boldsymbol{m}_{\lambda_j})$ are given by the kernel covariance function evaluated at $\boldsymbol{m}_{\lambda_i}$ and $\boldsymbol{m}_{\lambda_j}$, σ_n^2 the noise power, and $\boldsymbol{I}_L$ the $L \times L$ identity matrix.

To obtain the predictive distribution for $\psi_* \triangleq \psi(\boldsymbol{m}_{\lambda_*})$ at any test point $\boldsymbol{m}_{\lambda_*}$, we can write the joint distribution of the observation $\boldsymbol{r}$ and $\psi(\boldsymbol{m}_{\lambda_*})$ as [43]

$$\begin{bmatrix} \boldsymbol{r} \\ \psi_* \end{bmatrix} \sim \mathcal{N}\left(\boldsymbol{0}, \begin{bmatrix} \boldsymbol{K} + \sigma_n^2 \boldsymbol{I}_L & \boldsymbol{\kappa}_* \\ \boldsymbol{\kappa}_*^\top & \kappa_{**} \end{bmatrix}\right) \quad (10)$$

with $\boldsymbol{\kappa}_* = [\kappa(\boldsymbol{m}_{\lambda_*}, \boldsymbol{m}_{\lambda_1}), \ldots, \kappa(\boldsymbol{m}_{\lambda_*}, \boldsymbol{m}_{\lambda_L})]^\top$ and $\kappa_{**} = \kappa(\boldsymbol{m}_{\lambda_*}, \boldsymbol{m}_{\lambda_*})$. The predictive distribution of ψ_*, or posterior of ψ_*, is then obtained by conditioning (10) on the observation as follows

$$\begin{aligned} \psi_* | \boldsymbol{r}, \boldsymbol{M}, \boldsymbol{m}_{\lambda_*} \sim \mathcal{N}\Big(& \boldsymbol{\kappa}_*^\top \left[\boldsymbol{K} + \sigma_n^2 \boldsymbol{I}_L\right]^{-1} \boldsymbol{r}, \\ & \kappa_{**} - \boldsymbol{\kappa}_*^\top \left[\boldsymbol{K} + \sigma_n^2 \boldsymbol{I}_L\right]^{-1} \boldsymbol{\kappa}_* \Big). \end{aligned} \quad (11)$$

The extension to a multivariate predictive distribution with test data $\boldsymbol{M}_* = [\boldsymbol{m}_{\lambda_{*1}}, \ldots, \boldsymbol{m}_{\lambda_{*L}}]^\top$ yields

$$\begin{aligned} \boldsymbol{\psi}_* | \boldsymbol{r}, \boldsymbol{M}, \boldsymbol{M}_* \sim \mathcal{N}\Big(& \boldsymbol{K}_*^\top \left[\boldsymbol{K} + \sigma_n^2 \boldsymbol{I}_L\right]^{-1} \boldsymbol{r}, \\ & \boldsymbol{K}_{**} - \boldsymbol{K}_*^\top \left[\boldsymbol{K} + \sigma_n^2 \boldsymbol{I}_L\right]^{-1} \boldsymbol{K}_* \Big) \end{aligned} \quad (12)$$

with $[\boldsymbol{K}_*]_{ij} = \kappa(\boldsymbol{m}_{\lambda_{*i}}, \boldsymbol{m}_{\lambda_j})$ and $[\boldsymbol{K}_{**}]_{ij} = \kappa(\boldsymbol{m}_{\lambda_{*i}}, \boldsymbol{m}_{\lambda_{*j}})$. Finally, we arrive at the minimum mean square error (MMSE) estimator for GP regression

$$\begin{aligned} \widehat{\boldsymbol{\psi}}_* &= \mathbb{E}\{\boldsymbol{\psi}_* | \boldsymbol{r}, \boldsymbol{M}, \boldsymbol{M}_*\} \\ &= \boldsymbol{K}_*^\top \left[\boldsymbol{K} + \sigma_n^2 \boldsymbol{I}_L\right]^{-1} \boldsymbol{r}. \end{aligned} \quad (13)$$

In order to turn GP into a practical tool for processing hyperspectral data, it is essential to derive a method for estimating free parameters such as the noise variance σ_n^2 and possible kernel parameters defining the unknown parameter vector $\boldsymbol{\theta}$. We proceed as in [43] by maximizing the marginal likelihood $p(\boldsymbol{r} | \boldsymbol{M}, \sigma_n^2, \boldsymbol{\theta})$ with respect to $(\sigma_n^2, \boldsymbol{\theta})$, which yields

$$\begin{aligned} (\hat{\sigma}_n^2, \hat{\boldsymbol{\theta}}) = \arg\max_{\sigma_n^2, \boldsymbol{\theta}} \Big(& -\frac{1}{2} \boldsymbol{r}^\top \left[\boldsymbol{K} + \sigma_n^2 \boldsymbol{I}_L\right]^{-1} \boldsymbol{r} \\ & -\frac{1}{2} \log|\boldsymbol{K} + \sigma_n^2 \boldsymbol{I}_L| \Big). \end{aligned} \quad (14)$$

This problem has to be addressed with numerical optimization methods. There is no guarantee that the cost function does not suffer from multiple local optima. However, our practical experience with hyperspectral data indicates that local optima are not a critical problem in this context. The solutions to the optimization problem (14) for all examples reported in this paper were determined using the GPML toolbox [44].

We conclude this section by introducing some kernel functions. Common examples include the linear kernel defined as

$$\kappa(\boldsymbol{m}_{\lambda_i}, \boldsymbol{m}_{\lambda_j}) = \boldsymbol{m}_{\lambda_i}^\top \boldsymbol{m}_{\lambda_j} \quad (15)$$

and radial basis function kernels, which depend on $\|\boldsymbol{m}_{\lambda_i} - \boldsymbol{m}_{\lambda_j}\|$, such as the Gaussian kernel

$$\kappa(\boldsymbol{m}_{\lambda_i}, \boldsymbol{m}_{\lambda_j}) = \exp\left(-\frac{1}{2s^2}\|\boldsymbol{m}_{\lambda_i} - \boldsymbol{m}_{\lambda_j}\|^2\right) \quad (16)$$

where $s > 0$ is the kernel bandwidth. In the sequel, we shall use the Gaussian kernel for its smoothness and non-informativeness, as we lack any knowledge about the unknown function ψ. Then, $\theta = s$ (scalar). Note that this kernel has been used successfully in many signal and image processing applications, e.g., for hyperspectral data unmixing [13], [16].

IV. DETECTION OF NONLINEARLY MIXED PIXELS

A. The detection problem

Given an observation $\boldsymbol{r}$, we formulate the nonlinear mixture detector as the following binary hypothesis test problem

$$\mathcal{H}_0 : \boldsymbol{r} = \boldsymbol{M}\boldsymbol{\alpha} + \boldsymbol{n} \quad (17\text{a})$$
$$\mathcal{H}_1 : \boldsymbol{r} = \boldsymbol{\psi}(\boldsymbol{M}) + \boldsymbol{n} \quad (17\text{b})$$

where $\boldsymbol{n}$ is a zero-mean WGN with variance σ_n^2. We assume that the endmember matrix $\boldsymbol{M}$ is available, or has been estimated from data using an endmember extraction technique [2]. We shall relax this hypothesis in Section V, and use the nonlinear mixture detector to jointly perform this task.

We propose to compare the fitting errors resulting from estimating $\boldsymbol{r}$ with a linear or a nonlinear estimator (13). Under $\mathcal{H}_0$, both estimators should provide good estimates. Under $\mathcal{H}_1$, the estimation error resulting from the linear estimator should be significantly larger than that obtained with the nonlinear estimator. We shall now evaluate these fitting errors.

B. Linear estimation error

The MMSE estimator (13) may be used with the linear kernel (15) to estimate $\boldsymbol{\alpha}$ in (17a). However, this would require to solve (14) in order to estimate σ_n^2. To save on unnecessary computational efforts, we shall limit the use of GP to nonlinear model estimation. The MMSE estimator for (17a) is given by

$$\hat{\boldsymbol{\alpha}} = (\boldsymbol{M}^\top \boldsymbol{M})^{-1} \boldsymbol{M}^\top \boldsymbol{r} \quad (18)$$

resulting in the following estimation error

$$\boldsymbol{e}_{\text{lin}} = \boldsymbol{r} - \hat{\boldsymbol{r}}_{\text{lin}} = \boldsymbol{P}\boldsymbol{r} \quad (19)$$

where $\boldsymbol{P} = \boldsymbol{I}_L - \boldsymbol{M}(\boldsymbol{M}^\top \boldsymbol{M})^{-1} \boldsymbol{M}^\top$ is an $L \times L$ projection matrix of rank $\rho = L - R$. Note that no constraint is imposed on the abundance vector $\boldsymbol{\alpha}$. The objective is to obtain the best linear estimator, since the purpose at this point is not to perform unmixing, but to decide on the linearity (or not) of the considered mixture.

Consider first the distribution for $\|\boldsymbol{e}_{\text{lin}}\|^2$. Under $\mathcal{H}_1$, we have

$$\boldsymbol{e}_{\text{lin}} | \mathcal{H}_1 = \boldsymbol{P}[\boldsymbol{\psi} + \boldsymbol{n}]. \quad (20)$$

This implies that

$$\boldsymbol{e}_{\text{lin}} | \mathcal{H}_1 \sim \mathcal{N}(\boldsymbol{P}\boldsymbol{\psi}, \sigma_n^2 \boldsymbol{P}) \quad (21)$$

where we have used that the projection matrix $\boldsymbol{P}$ is idempotent, that is, $\sigma_n^2 \boldsymbol{P}\boldsymbol{P}^\top = \sigma_n^2 \boldsymbol{P}$. Under $\mathcal{H}_0$, we have

$$\boldsymbol{e}_{\text{lin}} | \mathcal{H}_0 \sim \mathcal{N}(\boldsymbol{0}, \sigma_n^2 \boldsymbol{P}). \quad (22)$$

Proper normalization of each squared entry $e_{\mathrm{lin},i}^2$ of $\boldsymbol{e}_{\mathrm{lin}}$ yields the conditional distributions under the two hypotheses

$$\frac{e_{\mathrm{lin},i}^2}{\sigma_n^2 \boldsymbol{p}_i^\top \boldsymbol{p}_i}\bigg|\mathcal{H}_1 \sim \chi_1^2\left(\frac{[\boldsymbol{p}_i^\top\boldsymbol{\psi}]^2}{\sigma_n^2\,\boldsymbol{p}_i^\top\boldsymbol{p}_i}\right)$$

$$\frac{e_{\mathrm{lin},i}^2}{\sigma_n^2 \boldsymbol{p}_i^\top \boldsymbol{p}_i}\bigg|\mathcal{H}_0 \sim \chi_1^2(0) \tag{23}$$

where $\boldsymbol{p}_i^\top$ denotes the i-th row of matrix $\boldsymbol{P}$, and $\chi_n^2(\lambda)$ is the noncentral χ-square distribution with n degrees of freedom and centrality parameter λ [45]. As $\boldsymbol{P}$ is idempotent and of rank $\rho = L - R$, which leads to $\|\boldsymbol{e}_{\mathrm{lin}}\|^2 = \boldsymbol{r}^\top \boldsymbol{P} \boldsymbol{r}$, we conclude that [46, p. 33]

$$\frac{\|\boldsymbol{e}_{\mathrm{lin}}\|^2}{\sigma_n^2}\bigg|\mathcal{H}_0 \sim \chi_\rho^2(0). \tag{24}$$

C. Nonlinear estimation error with GP

Since our interest at this point is not to make predictions for new data, but to evaluate the fitting error between the model output and the available data, we define the GP estimation error as:

$$\boldsymbol{e}_{\mathrm{nlin}} = \boldsymbol{r} - \hat{\boldsymbol{r}}_{\mathrm{nlin}} \tag{25}$$

where $\hat{\boldsymbol{r}}_{\mathrm{nlin}}$ is given by (13) with $\boldsymbol{M}_* = \boldsymbol{M}$. Hence, using (13) in (25) yields

$$\boldsymbol{e}_{\mathrm{nlin}} = \boldsymbol{r} - \widehat{\boldsymbol{\psi}}_*\bigg|_{\boldsymbol{M}_* = \boldsymbol{M}} = \boldsymbol{H}\boldsymbol{r} \tag{26}$$

where $\boldsymbol{H} = \boldsymbol{I}_L - \boldsymbol{K}^\top\left[\boldsymbol{K} + \sigma_n^2\boldsymbol{I}_L\right]^{-1}$ is a real-valued matrix of rank L. We shall now analyze the distribution of $\|\boldsymbol{e}_{\mathrm{nlin}}\|^2$ under hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$. Under hypothesis $\mathcal{H}_1$, we have

$$\boldsymbol{e}_{\mathrm{nlin}}|\mathcal{H}_1 = \boldsymbol{H}(\boldsymbol{\psi} + \boldsymbol{n}). \tag{27}$$

This leads to the following conditional distribution

$$\boldsymbol{e}_{\mathrm{nlin}}|\mathcal{H}_1 \sim \mathcal{N}(\boldsymbol{H}\boldsymbol{\psi}, \sigma_n^2\boldsymbol{H}\boldsymbol{H}^\top). \tag{28}$$

Under hypothesis $\mathcal{H}_0$, the distribution for the error becomes

$$\boldsymbol{e}_{\mathrm{nlin}}|\mathcal{H}_0 \sim \mathcal{N}(\boldsymbol{H}\boldsymbol{M}\boldsymbol{\alpha}, \sigma_n^2\boldsymbol{H}\boldsymbol{H}^\top). \tag{29}$$

The distribution of the i-th entry of $\boldsymbol{e}_{\mathrm{nlin}}$ is thus given by

$$e_{\mathrm{nlin},i}|\mathcal{H}_0 \sim \mathcal{N}(\boldsymbol{h}_i^\top\boldsymbol{M}\boldsymbol{\alpha}, \sigma_n^2\boldsymbol{h}_i^\top\boldsymbol{h}_i). \tag{30}$$

Proper normalization of each squared entry $e_{\mathrm{nlin},i}$ of $\boldsymbol{e}_{\mathrm{nlin}}$ yields the following conditional distributions

$$\frac{e_{\mathrm{nlin},i}^2}{\sigma_n^2 \boldsymbol{h}_i^\top \boldsymbol{h}_i}\bigg|\mathcal{H}_1 \sim \chi_1^2\left(\frac{[\boldsymbol{h}_i^\top\boldsymbol{\psi}]^2}{\sigma_n^2\,\boldsymbol{h}_i^\top\boldsymbol{h}_i}\right)$$

$$\frac{e_{\mathrm{nlin},i}^2}{\sigma_n^2 \boldsymbol{h}_i^\top \boldsymbol{h}_i}\bigg|\mathcal{H}_0 \sim \chi_1^2\left(\frac{[\boldsymbol{h}_i^\top\boldsymbol{M}\boldsymbol{\alpha}]^2}{\sigma_n^2\,\boldsymbol{h}_i^\top\boldsymbol{h}_i}\right) \tag{31}$$

where $\boldsymbol{h}_i^\top$ denotes the i-th row of $\boldsymbol{H}$. Non-central χ-square distributions in (23) and (31) make the analysis of the test statistics in the next section intractable, even under $\mathcal{H}_0$. In order to proceed, we argue that it is reasonable to assume that, under $\mathcal{H}_0$, both the nonlinear GP regression method and the linear one should achieve the same level of accuracy.

Considering (24), this approximation leads to

$$\frac{\|\boldsymbol{e}_{\mathrm{nlin}}\|^2}{\sigma_n^2}\bigg|\mathcal{H}_0 = \chi_\rho^2(0). \tag{32}$$

We validated this approximation using extensive Monte Carlo simulations. Figures 2a and 2b illustrate this assumption for a representative example.

D. The test statistics

We propose to compare the squared norms of the two fitting error vectors $\boldsymbol{e}_{\mathrm{nlin}}$ and $\boldsymbol{e}_{\mathrm{lin}}$ to decide between $\mathcal{H}_0$ and $\mathcal{H}_1$. Also, the test statistics should allow for the adjustment of the detection threshold to a given probability of false alarm (PFA) for design purposes. Considering these two objectives, we propose the following statistical test

$$T = \frac{2\|\boldsymbol{e}_{\mathrm{nlin}}\|^2}{\|\boldsymbol{e}_{\mathrm{nlin}}\|^2 + \|\boldsymbol{e}_{\mathrm{lin}}\|^2} \overset{\mathcal{H}_1}{\underset{\mathcal{H}_0}{\lessgtr}} \tau \tag{33}$$

where τ is the detection threshold.

The reasoning behind the choice of T defined in (33) is as follows. Under $\mathcal{H}_0$, both $\|\boldsymbol{e}_{\mathrm{nlin}}\|^2$ and $\|\boldsymbol{e}_{\mathrm{lin}}\|^2$ are χ-square dependent random variables. Now, we write $\boldsymbol{e}_{\mathrm{lin}}$ as $\boldsymbol{e}_{\mathrm{nlin}} + \sqrt{2}\boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ is assumed to be also a zero-mean i.i.d. Gaussian vector[1], and neglect the cross-term $\boldsymbol{e}_{\mathrm{nlin}}^\top\boldsymbol{\epsilon}$ when compared to $\|\boldsymbol{\epsilon}\|^2$ in evaluating $\|\boldsymbol{e}_{\mathrm{lin}}\|^2$ under $\mathcal{H}_0$. The latter approximation is due to the lack of correlation between $\boldsymbol{e}_{\mathrm{nlin}}$ and $\boldsymbol{\epsilon}$, which can be largely attributed to mismatches resulting from the numerical optimization required to solve (14). Under these considerations, (33) can be written as $T = \|\boldsymbol{e}_{\mathrm{nlin}}\|^2/(\|\boldsymbol{e}_{\mathrm{nlin}}\|^2 + \|\boldsymbol{\epsilon}\|^2)$ with both $\|\boldsymbol{e}_{\mathrm{nlin}}\|^2$ and $\|\boldsymbol{\epsilon}\|^2$ independent and χ-square distributed. Such a statistics is known to have a beta distribution [47].

As the GP estimator tends to fit better nonlinearly mixed data, T should be less than 1 under hypothesis $\mathcal{H}_1$. Conversely, T should be close to one for linearly mixed pixels, as $\|\boldsymbol{\epsilon}\|^2$ tends to be much less than $2\|\boldsymbol{e}_{\mathrm{nlin}}\|^2$. Hence, as per (33), we accept hypothesis $\mathcal{H}_0$ if $T > \tau$ and we conclude for the nonlinear mixing hypothesis $\mathcal{H}_1$ if $T < \tau$.

E. Determining the detection threshold

Considering the assumption that the statistical test T has a beta distribution under $\mathcal{H}_0$, a decision threshold τ can be determined for a given PFA as

$$\tau = \mathcal{B}_{\alpha,\beta}^{-1}(\mathrm{PFA}) \tag{34}$$

where $\mathcal{B}_{\alpha,\beta}$ is the cumulative distribution function of the beta distribution with parameters (α, β). The parameters of this function must be estimated from the data. To this end, we initially determine an estimate $\widehat{\boldsymbol{A}}$ of the abundance matrix assuming the linear mixing model with the real observations $\boldsymbol{R} = [\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N]$ and the known endmember matrix $\boldsymbol{M}$. Then, using $\boldsymbol{M}$ and $\widehat{\boldsymbol{A}}$ we construct the synthetic image $\boldsymbol{R}_s = \boldsymbol{M}\widehat{\boldsymbol{A}}$, which satisfies $\mathcal{H}_0$. For this linearly mixed hyperspectral image, we then compute, say, N samples of the test statistics $T|\mathcal{H}_0$ defined in (33) and fit a beta distribution

---

[1] The constant factor $\sqrt{2}$ is for notation purpose only.

to these samples. The threshold τ for each PFA is then determined using (34).

This procedure requires the knowledge of the endmember matrix $\boldsymbol{M}$. The next section proposes an iterative technique to estimate $\boldsymbol{M}$ from an hyperspectral image, which we assume to contain linearly and nonlinearly mixed pixels.

V. ENDMEMBER EXTRACTION IN NONLINEARLY MIXED HYPERSPECTRAL IMAGES

The presence of nonlinearly mixed pixels in an hyperspectral image tends to degrade the estimation accuracy of endmember extraction methods based on a linear mixing model. As a consequence, nonlinearly mixed pixels also affect the performance of algorithms using the endmember matrix such as the detection method presented in this paper. There has been few papers addressing endmember estimation from nonlinearly mixed images. A nonlinear unmixing algorithm is derived in [48]. The pixel reflectances are supposed to be post-nonlinear functions of unknown pure spectral components. A Bayesian strategy is proposed to both unmix the data and estimate the endmembers. Both tasks are however mutually dependent and the unmixing model is very specific. A nonlinear endmember estimation algorithm based on the approximation of geodesic distances is introduced in [27], [49]. This algorithm can however suffer from the absence of pure pixels in the image, and the effectiveness of using manifold learning methods on real data still needs to be analyzed and confirmed. In this section, we propose an iterative technique for estimating the endmember matrix $\boldsymbol{M}$ under the reasonable assumptions that the number R of endmembers is known [50]–[52], and that these endmembers are linearly mixed within at least a small part of the image. Nonlinear mixtures may however compose a significant part of the image. The proposed technique combines the detector of nonlinearly mixed pixels presented in Section IV and the endmember estimation algorithm MVES [25].

The procedure is described in Algorithm 1. It is a two-step iterative algorithm. The first step consists of using MVES to estimate the endmembers (line 2 and 14 in Algorithm 1). The second step uses (33) to compute the detection statistics for all the L pixels in the image $\boldsymbol{R}_{\mathrm{tmp}}$ (line 7 in Algorithm 1). Then, all pixels whose detection statistic satisfies $T(i) \leq \tau_r$ are removed (line 9), where $\tau_r = r_f \times \tau$ (line 4 and 11) is the relaxed detection threshold. The use of a relaxed threshold is suggested to avoid discarding linear pixels during the first iterations, when the estimates of $\boldsymbol{M}$ are still not sufficiently accurate. The relaxing factor is initialized for $r_f = 0.9$ and is increased by a factor $r_{\mathrm{inc}} = 0.1/N_{\mathrm{max}}$ at each iteration to improve pixel selection as the estimation of the matrix $\boldsymbol{M}$ improves (line 10). The procedure is repeated until the linear and the nonlinear GP models in (33) present similar fitting errors within the limit of ε. Using this procedure, τ_r tends to the desired threshold τ as the estimation of $\boldsymbol{M}$ improves, leaving mostly linear pixels for which both models have similar performance. A maximum number of iterations N_{max} is also set to avoid discarding too much data.

Note that we have opted for the MVES algorithm for endmember extraction because it inscribes the data into a minimum-volume simplex. Thus, MVES is suitable to estimate $\boldsymbol{M}$ in the absence of pure pixels. This feature is specially interesting for our purpose since the procedure described above discards data, which may even be pure or near-pure pixels during the first iterations. Nevertheless, any other endmember estimation algorithm valid in absence of pure pixel [39], [53], [54] could be potentially used with Algorithm 1.

Algorithm 1: Iterative endmember estimation

Input : The hyperspectral image $\boldsymbol{R}$, and the number of endmembers R

Output: Estimated endmember matrix $\widehat{\boldsymbol{M}}$

1 Initialization: $T_{\mathrm{max}} = 1$, $T_{\mathrm{min}} = 0$, $\varepsilon = 0.05$, $\boldsymbol{R}_{\mathrm{tmp}} = \boldsymbol{R}$, $N_{\mathrm{max}} = 10$, $cc = 0$, $r_f = 0.9$, $r_{inc} = (1 - r_f)/N_{\mathrm{max}}$, PFA $= 0.05$;

2 $\widehat{\boldsymbol{M}} = \mathrm{MVES}(\boldsymbol{R}_{\mathrm{tmp}}, R)$;

3 Compute τ using (34);

4 $\tau_r = r_f \times \tau$;   %% (relaxed threshold)

5 **while** $T_{\mathrm{max}} - T_{\mathrm{min}} > \varepsilon$ & $cc < N_{\mathrm{max}}$ **do**

6   **for** $i = 1$ **to** N_{pixels} **do**

7 | Compute $\boldsymbol{T}(i)$ using (33);

8   **end**

9   Remove all pixels with $\boldsymbol{T}(i) \leq \tau_r$ from $\boldsymbol{R}_{\mathrm{tmp}}$;

10   $r_f = r_f + r_{inc}$;   %% (relaxing factor)

11   $\tau_r = r_f \times \tau$;

12   $T_{\mathrm{max}} = \max(\boldsymbol{T})$; $T_{\mathrm{min}} = \min(\boldsymbol{T})$;

13   $cc = cc + 1$;

14   $\widehat{\boldsymbol{M}} = \mathrm{MVES}(\boldsymbol{R}_{\mathrm{tmp}}, R)$;

15 **end**

VI. SIMULATIONS

This section presents simulation results to validate the proposed approach for detecting nonlinearly mixed pixels, with both synthetic and real images. The use of synthetic images is important as they provide a ground truth against which the performance of the detector can be verified. First, we propose a definition for a degree of nonlinearity of an hyperspectral image so that the relative performances of different detectors can be compared. This is helpful to quantify the relative energies associated with the linear and nonlinear mixing components in hyperspectral images generated with different nonlinear mixing models.

A. Degree of nonlinearity

Consider that a pixel vector can be written as the sum of a linear and a nonlinear mixing component[2] as for most existing nonlinear mixing models [8], [9], [11], [18], [41]

$$\boldsymbol{r} = \boldsymbol{r}_{\mathrm{lin}} + \boldsymbol{r}_{\mathrm{nlin}} \tag{35}$$

where $\boldsymbol{r}_{\mathrm{lin}}$ and $\boldsymbol{r}_{\mathrm{nlin}}$ are the linear and nonlinear mixing contributions to $\boldsymbol{r}$, respectively. The energy of $\boldsymbol{r}$ is given by

$$E = \|\boldsymbol{r}\|^2 = \|\boldsymbol{r}_{\mathrm{lin}}\|^2 + 2\,\boldsymbol{r}_{\mathrm{lin}}^{\top}\boldsymbol{r}_{\mathrm{nlin}} + \|\boldsymbol{r}_{\mathrm{nlin}}\|^2 \tag{36}$$

where $E_{\mathrm{lin}} = \|\boldsymbol{r}_{\mathrm{lin}}\|^2$ is the energy of the linear contribution and $E_{\mathrm{nlin}} = 2\,\boldsymbol{r}_{\mathrm{lin}}^{\top}\boldsymbol{r}_{\mathrm{nlin}} + \|\boldsymbol{r}_{\mathrm{nlin}}\|^2$ is the part of the pixel energy

---

[2]We do not account for noise contribution as it can be set by the user independently of the mixing model.

affected by the nonlinear mixing. Given a mixing model, we define the degree of nonlinearity η_d as the ratio of the energy of the nonlinear contribution E_{nlin} to the total energy E. Thus,

$$\eta_d = \frac{E_{\text{nlin}}}{E} = \frac{1}{1+A} \tag{37}$$

where $A = \|\boldsymbol{r}_{\text{lin}}\|^2/(2\,\boldsymbol{r}_{\text{lin}}^\top \boldsymbol{r}_{\text{nlin}} + \|\boldsymbol{r}_{\text{nlin}}\|^2)$. Next, we show how to apply this definition for generating synthetic samples with two different mixing models.

1) Synthetic data generation with GBM: To be able to control the relative contributions of the linear and nonlinear mixing parts of the GBM model, we introduce a new scaling factor k into the generalized bilinear model (GBM) used in [31]. For an endmember matrix $\boldsymbol{M}$ and an abundance vector $\boldsymbol{\alpha}$, we write the modified noiseless GBM model as

$$\boldsymbol{r} = k\boldsymbol{M}\boldsymbol{\alpha} + \gamma\boldsymbol{\nu} \tag{38}$$

where $0 \leq k \leq 1$, $\boldsymbol{\nu} = \sum_{i=1}^{R-1}\sum_{j=i+1}^{R} \alpha_i\alpha_j\boldsymbol{m}_i \odot \boldsymbol{m}_j$ is the nonlinear mixing term, γ is the scaling parameter for the nonlinear contribution, and $\odot$ is the Hadamard product. The degree of nonlinearity is then

$$\eta_d = \frac{2k\gamma(\boldsymbol{\nu}^\top\boldsymbol{M}\boldsymbol{\alpha}) + \gamma^2\|\boldsymbol{\nu}\|^2}{k^2\|\boldsymbol{M}\boldsymbol{\alpha}\|^2 + 2k\gamma(\boldsymbol{\nu}^\top\boldsymbol{M}\boldsymbol{\alpha}) + \gamma^2\|\boldsymbol{\nu}\|^2} = \frac{1}{1+A} \tag{39}$$

with $A = k^2\|\boldsymbol{M}\boldsymbol{\alpha}\|^2/(2k\gamma(\boldsymbol{\nu}^\top\boldsymbol{M}\boldsymbol{\alpha}) + \gamma^2\|\boldsymbol{\nu}\|^2)$. We have to determine the scaling factors k and γ so that the energy E is independent of $\eta_d \geq 0$. This condition can be expressed as $\|\boldsymbol{M}\boldsymbol{\alpha}\|^2 = k^2\|\boldsymbol{M}\boldsymbol{\alpha}\|^2 + 2k\gamma(\boldsymbol{\nu}^\top\boldsymbol{M}\boldsymbol{\alpha}) + \gamma^2\|\boldsymbol{\nu}\|^2$, leading to

$$A = \frac{k^2}{1-k^2} \tag{40}$$

or

$$k = \sqrt{\frac{A}{1+A}} = \sqrt{1-\eta_d}. \tag{41}$$

To obtain γ, note that the denominator of A can be written as $\gamma^2\|\boldsymbol{\nu}\|^2 + 2k\gamma(\boldsymbol{\nu}^\top\boldsymbol{M}\boldsymbol{\alpha}) = (1-k^2)\|\boldsymbol{M}\boldsymbol{\alpha}\|^2$. Since γ must be positive, we have

$$\begin{aligned}\gamma = \frac{1}{2\|\boldsymbol{\nu}\|^2}\Big(&-2k(\boldsymbol{\nu}^\top\boldsymbol{M}\boldsymbol{\alpha}) \\ &+ \sqrt{4k^2(\boldsymbol{\nu}^\top\boldsymbol{M}\boldsymbol{\alpha})^2 + 4\|\boldsymbol{\nu}\|^2(1-k^2)\|\boldsymbol{M}\boldsymbol{\alpha}\|^2}\Big).\end{aligned} \tag{42}$$

Once k and γ have been determined from η_d, we can generate data following the model in (38).

2) Synthetic data generation with PNMM: To match the noiseless PNMM model (6) with the proposed formulation (38), we complement it with a weighted linear mixture as follows

$$\boldsymbol{r} = k\boldsymbol{M}\boldsymbol{\alpha} + \gamma\boldsymbol{\nu} \tag{43}$$

where $\boldsymbol{\nu} = (\boldsymbol{M}\boldsymbol{\alpha})^\xi$ denotes the exponential value ξ applied to each entry of $\boldsymbol{M}\boldsymbol{\alpha}$. Model (43) reduces to (6) for $k = 0$ and $\gamma = 1$. Again, parameters k and γ are scaling factors that control the relative amounts of linear and nonlinear contributions given η_d. As for the GBM, both can be set using (41) and (42).
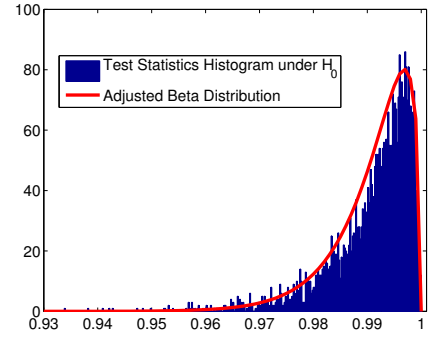

Fig. 3: Histogram of the test statistics under $\mathcal{H}_0$ and the adjusted Beta distribution.

B. Simulations with known $\boldsymbol{M}$

We now present simulations with synthetic data and a known endmember matrix $\boldsymbol{M}$. These simulations allow us to assess the detector performance disregarding estimation errors for the endmembers. Hence, they illustrate the potential of the proposed detector. To construct synthetic data, we used three materials ($R = 3$) extracted from the spectral library of the software ENVI$^{\text{TM}}$ [55]: green grass, olive green paint and galvanized steel metal. Each endmember $\boldsymbol{m}_r$ has $L = 826$ bands that were uniformly decimated by 3 to $L = 275$ bands.

To evaluate the performance of the proposed detector, we generated 8000 synthetic samples by mixing the three collected spectra. Among the 8000 pixels, 4000 were generated using the linear model in (2), and 4000 using the modified generalized bilinear model in (38). A fixed abundance vector $\boldsymbol{\alpha} = [0.3, 0.6, 0.1]^\top$ was used for all samples. Nonlinearly mixed samples were generated using different degrees of nonlinearity $\eta_d \in \{0.3, 0.5, 0.8\}$ to test the detector under different conditions. The power of the additive Gaussian noise was set to $\sigma_n^2 = 0.001$, which corresponds to SNR = 21dB.

Figure 1 shows the receiver operating characteristics (ROCs) of the proposed GP detector and the LS robust detector presented in [31] for the three values of η_d. The proposed detector performs better, especially for moderate to high degree of nonlinearity. For instance, Fig. 1c shows that the GP detector achieves a probability of detection of 1 for PFA = 0.1, while the LS robust detector yields a probability of detection of approximately 0.65 for the same PFA. Figure 2 shows the histograms of $\|\boldsymbol{e}_{\text{nlin}}\|^2$, $\|\boldsymbol{e}_{\text{lin}}\|^2$ and T for both linearly ($\mathcal{H}_0$) and nonlinearly ($\mathcal{H}_1$) mixed data. The proposed test statistics clearly leads to histograms that differ significantly under both hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, which explains the improvement in detection performance. Figure 3 compares the histogram of T under $\mathcal{H}_0$ with the fitted beta distribution, confirming that the distribution of T can be reasonably approximated by a beta distribution.

We considered two unmixing algorithms to assess the impact of the proposed detector on unmixing performance, one linear and one nonlinear. Linear unmixing was performed using the fully-constrained least-squares (FCLS) algorithm [56]. For nonlinear unmixing, we used the SK-Hype algorithm [13]. The two algorithms were employed in two unmixing strategies. First, each algorithm was used to unmix the complete

(a) Robust LS detector [31].   (b) Proposed GP detector.   (c) Comp. LS and GP ($\eta_d = 0.5$).
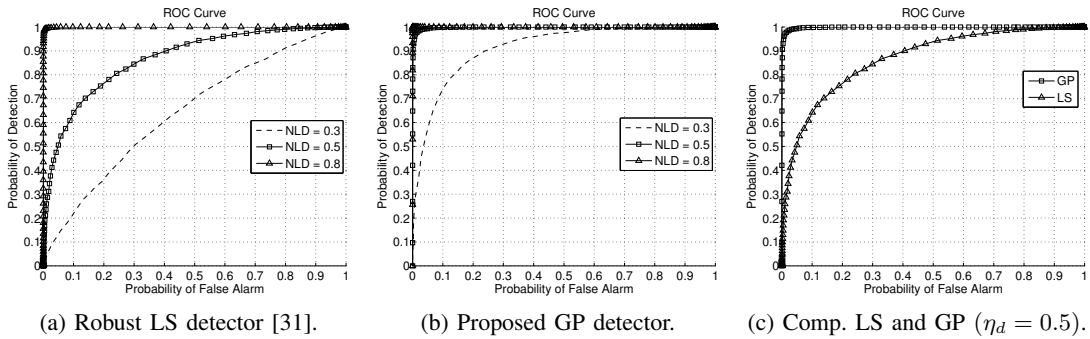
Fig. 1: Empirical ROCs for: (a) the Robust LS detector [31], (b) the proposed GP detector, (c) the two detectors for $\eta_d = 0.5$. All curves were obtained for 8000 pixels (4000 linearly mixed and 4000 nonlinearly mixed) and SNR = 21dB. Nonlinear mixtures were generated using the simplified GBM described in Section VI-A.
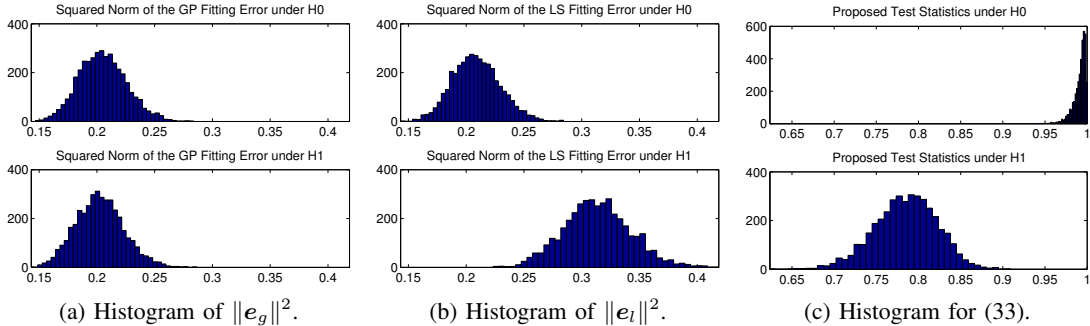

(a) Histogram of $\|e_g\|^2$.   (b) Histogram of $\|e_l\|^2$.   (c) Histogram for (33).

Fig. 2: Histograms for (a) the squared norm of the GP fitting error, (b) the least-squares fitting error, and (c) the test statistics (33).

hyperspectral image. In the second strategy called detect-then-unmix (D.+U.), the proposed detector (GP), and the detector of [31] (LS) were used as a pre-processing step. Then, FCLS was used to unmix pixels detected as linearly mixed and SK-Hype was used to unmix pixels detected as nonlinearly mixed. The detection threshold τ was determined for PFA = 0.01. Two synthetic images were considered with 1000 pixels each, 500 being linearly mixed and 500 being nonlinearly mixed. Each image was constructed with a particular nonlinear mixing model, with a fixed degree of nonlinearity $\eta_d = 0.5$ in both cases, with abundance vectors uniformly sampled in the simplex. The GBM (38) was used for the first image (Image I), while the PNMM (43) with $\xi = 3$ was considered for the second image (Image II). The SNR was 21dB in both cases, and the abundances were drawn uniformly in the simplex. Parameters k and γ were determined for each pixel to maintain the desired value of nonlinearity degree η_d for all simulations. To compare the results, we used the root mean square error (RMSE) of abundance estimation, defined as

$$\text{RMSE} = \sqrt{\frac{1}{NR} \sum_{n=1}^{N} \|\boldsymbol{\alpha}_n - \hat{\boldsymbol{\alpha}}_n\|^2} \qquad (44)$$

where N is the number of pixels in each image.

The results (RMSE $\pm$ standard deviation) are presented in Tables I and II. For each image, these tables indicate the RMSEs for the linearly mixed part (LMM), for the nonlinearly mixed part (NLM), and for the full image (Full Img.) using the three unmixing strategies. The results shown in bold blue

are those with the lowest RMSE in each row of the tables. As expected, FCLS has the best results when unmixing linearly mixed pixels. The same observation can be made for SK-Hype with nonlinearly mixed pixels. Nevertheless, we verify that the results using the detect-then-unmix strategy and the proposed detector (D.+U. GP) are very close to the best results for both types of pixels, LMM and NLM. When processing the whole image without prior information on the mixing nature of each pixel, the best results were those obtained with the detect-then-unmix GP strategy. Additionally, we present the classification error (C. E.) in percentage for the detect-then-unmix strategy in both tables. The last two columns in both tables clearly illustrate the better performance obtained using the proposed (GP) detector, as opposed to the detector of [31].

To verify the statistical significance of the results shown in Tables I and II, we performed the one-tailed left nonparametric Wilcoxon signed rank test [57]. The test was performed to compare the abundance estimation RMSEs obtained with the proposed methodology (D.+U. GP) and with each of the alternative methods listed in Tables I and II. The Wilcoxon signed rank test considers the samples to be paired, which corresponds to our case, and tests the following null hypothesis

$$\text{median}(\text{RMSE}_{\text{prop}}) = \text{median}(\text{RMSE}_{\text{alt}}) \qquad (45)$$

where $\text{RMSE}_{\text{prop}}$ and RMSE_{alt} stand for the RMSEs obtained using the proposed and the alternative methods, respectively. Tables III and IV show the results obtained for the simulations corresponding to Tables I and II. We assigned the symbol $\mathcal{A}$ if the null hypothesis was rejected with negative Z statistic, i.e.,

TABLE I: Abundance estimation RMSE for M known and using the GBM mixing model (SNR = 21dB, $\eta_d = 0.5$).

	Image I: LMM + GBM			
Model	FCLS	SK-Hype	D.+U. GP (C.E.%)	D.+U. LS (C.E.%)
LMM	**0.0095 ± 0.00010**	0.0205 ± 0.00057	0.0097 ± 0.00012 (0.6)	0.0096 ± 0.00010 (0.2)
NLM	0.0624 ± 0.00384	**0.0312 ± 0.00110**	0.0324 ± 0.00119 (5.6)	0.0509 ± 0.00314 (51.4)
F.Img	0.0446 ± 0.00332	0.0264 ± 0.00092	**0.0239 ± 0.00097** (3.1)	0.0366 ± 0.00255 (25.8)

TABLE II: Abundance estimation RMSE for M known and using the PNMM mixing model (SNR = 21dB, $\eta_d = 0.5$).

	Image II: LMM + PNMM			
Model	FCLS	SK-Hype	D.+U. GP (C.E.%)	D.+U. LS (C.E.%)
LMM	**0.0095 ± 0.00010**	0.0205 ± 0.00057	0.0099 ± 0.00013 (1.2)	**0.0095 ± 0.00010** (0)
NLM	0.0958 ± 0.00882	**0.0440 ± 0.00204**	0.0443 ± 0.00210 (0.8)	0.0483 ± 0.00276 (17)
F.Img	0.0681 ± 0.00772	0.0344 ± 0.00168	**0.0321 ± 0.00176** (1)	0.0348 ± 0.00225 (8.5)

TABLE III: One-tailed Wilcoxon signed rank test for Image I. (Significance level 0.05)

	FCLS	SK-Hype	D.+U. LS
LMM	-	$\mathcal{A}$	-
NLM	$\mathcal{A}$	-	$\mathcal{A}$
F.Img.	$\mathcal{A}$	$\mathcal{A}$	$\mathcal{A}$

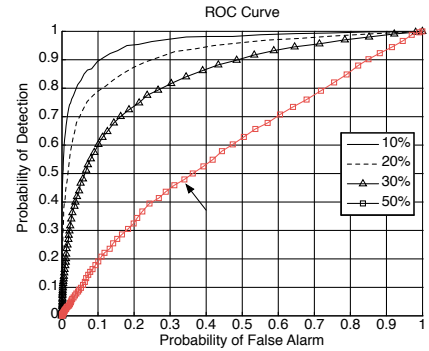TABLE IV: One-tailed Wilcoxon signed rank test for Image II. (Significance level 0.05)

	FCLS	SK-Hype	D.+U. LS
LMM	-	$\mathcal{A}$	-
NLM	$\mathcal{A}$	-	$\mathcal{A}$
F.Img.	$\mathcal{A}$	$\mathcal{A}$	$\mathcal{A}$

Fig. 4: ROCs for different proportions of nonlinearly mixed pixels and $\eta_d = 0.5$. Endmember extraction using VCA.

if there was enough evidence that $\text{RMSE}_{\text{prop}} < \text{RMSE}_{\text{alt}}$ at the 0.05 significance level. We assigned the symbol "-" if the null hypothesis could not be rejected. These results provide evidence that the improvement in abundance estimation obtained using the proposed technique is statistically consistent.

C. Simulations with an unknown endmember matrix M

The simulations conducted in Section VI-B assumed the endmember matrix M to be known. Although this study is important to quantify the potential of the proposed detector, the endmembers are rarely known in practice. Hence, in this section, we study the sensitivity of the detection performance as a function of the endmember estimation accuracy and of the degree of nonlinearity. Endmember extraction was performed with the iterative method proposed in Section V, and with VCA [23] for comparison.

Figure 4 presents the results of 4 experiments using synthetic images with 5000 samples, SNR = 21dB, abundances uniformly sampled in the simplex, a proportion of nonlinearly mixed pixels in the image varying from 10% to 50%, and $\eta_d = 0.5$. For every experiment, the endmember matrix was extracted using VCA. These results show how the detection performance can degrade as the number of nonlinear pixels

increases and as VCA loses accuracy in extracting the endmembers from the image. These results confirm the importance of devising alternatives to VCA (or to other endmember extraction algorithms specifically designed for linearly-mixed images) for images containing nonlinearly-mixed pixels. Figure 5 presents the results obtained with Algorithm 1 and classical MVSE for endmember extraction. For this experiment, we generated data with 50% of nonlinearly mixed pixels and different degrees of nonlinearity $\eta_d \in \{0.3, 0.5, 0.8\}$. The corresponding cases for $\eta_d = 0.5$ and 50% of nonlinearly mixed pixels are shown in red and pointed by arrows in Figure 4 and Figure 5. The poor results obtained using classical MVSE are also indicated. Comparing Fig. 1 and 5 shows that the results obtained with the iterative endmember extraction algorithm are very close to those obtained for a known endmember matrix M (which can be considered as the reference detector).

Figure 6 illustrates a representative example of evolution obtained with the proposed iterative endmember extraction algorithm. These plots correspond to a simulation performed using 1000 synthetic samples, 500 being linearly mixed and 500 being nonlinearly mixed. The nonlinearly mixed pixels were created using the GBM (38) with $\eta_d = 0.5$. The data were projected onto the space spanned by the columns of the current endmember matrix M. They are represented as black dots. The current endmembers are shown as green dots. The true endmembers are shown as black circles at the vertices of
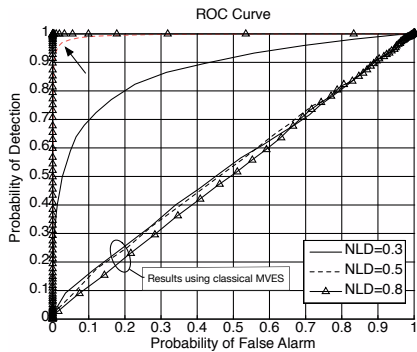
Fig. 5: ROCs for different degrees of nonlinearity η_d and 50% of nonlinearly mixed pixels in the image. Endmember extraction using Algorithm 1

TABLE V: Mean RMSE for endmember estimation.

r_f	ε	$N_{\max} = 5$	$N_{\max} = 10$	$N_{\max} = 15$
	0.01	0.0825 ± 0.0384	0.0784 ± 0.0378	$\mathbf{0.0778 \pm 0.0408}$
0.7	0.05	0.0817 ± 0.0377	0.0788 ± 0.0396	$\mathbf{0.0778 \pm 0.0410}$
	0.1	0.0832 ± 0.0383	0.0808 ± 0.0437	$\mathbf{0.0783 \pm 0.0398}$
	0.01	0.0821 ± 0.0401	0.0783 ± 0.0416	$\mathbf{0.0753 \pm 0.0393}$
0.8	0.05	0.0805 ± 0.0387	$\mathbf{0.0766 \pm 0.0400}$	0.0778 ± 0.0435
	0.1	0.0819 ± 0.0416	$\mathbf{0.0758 \pm 0.0367}$	0.0801 ± 0.0406
	0.01	0.0776 ± 0.0428	0.0738 ± 0.0399	$\mathbf{0.0702 \pm 0.0394}$
0.9	0.05	0.0764 ± 0.0379	$\mathbf{0.0741 \pm 0.0420}$	0.0744 ± 0.042
	0.1	0.0787 ± 0.0401	0.0785 ± 0.0355	$\mathbf{0.0780 \pm 0.0344}$

the true simplex drawn with black lines. The data discarded at each iteration are shown within blue circles. Figure 6a shows the first iteration of Algorithm 1. Numerous nonlinear samples are outside the simplex and endmember are poorly estimated. The situation improves in Fig. 6b, which depicts the fourth iteration. Here, much less data lie outside the simplex, and two of the endmember estimates have improved significantly. Similar improvement can be noticed in the seventh iteration in Fig. 6c. The final result obtained after 10 iterations only is shown in Fig. 6d, where most of the nonlinear data were discarded and the endmember estimates are clearly close to the true endmembers.

D. Adjusting parameters r_f, N_{max}, and ε

The implementation of Algorithm 1 requires the choice of parameters $N_{\max}$, ε, and r_f. We have found from several experiments that using $r_f \in [0.8, 09]$, $\varepsilon = 0.05$ and $N_{\max} = 10$ is a good choice for different scenarios. This section explores the sensitivity of the algorithm performance to variations of these parameter values about these choices. To this end we applied the algorithm to synthetic data with the following properties: 100 pixels, $R = 3$ endmembers, 50 pixels mixed with the LMM and 50 pixels mixed with the GBM with $\eta_d = 0.5$. The abundance vectors were sampled uniformly in the simplex, and WGN was added to produce SNR = 21dB. The spectra were the same used for the previous simulations, uniformly decimated by 5, resulting in 166 bands. The three parameters were chosen from the following sets: $N_{\max} \in [5, 10, 15]$, $\varepsilon \in [0.01, 0.05, 0.1]$, and $r_f \in [0.7, 0.8, 0.9]$. For each combination of parameters we performed $N_r = 900$ runs of Algorithm 1, and computed the RMSE of estimated endmembers using (44) with the abundance vectors replaced with the endmembers. Table V shows the obtained results. The best results were obtained for each pair (r_f, ε) are highlighted in bold blue. These results show that the performance of the algorithm is not very sensitive to different parameter choices. They also show that choosing $N_{\max} < 10$ tends to increase the RMSE. Furthermore, it is clear that choosing $r_f < 0.8$ tends to require larger values of $N_{\max}$.

The choice of the parameters should be directed to prevent the algorithm from an early convergence with elimination of a large amount of linearly mixed pixels along with the nonlinearly mixed ones. This can be achieved by setting $N_{\max}$ to a sufficiently large value, which controls both the maximum number of iterations and the increment of the detection threshold τ. From our experience with the proposed method, good results can be obtained as follows:

a) Set r_f somewhere in the range [0.8 0.9] (Remark: a larger value would probably lead to early discarding of linearly mixed pixels).

b) Set $\varepsilon \leq 0.05$, so that $\mathbf{R}_{\text{tmp}}$ would contains basically linearly mixed pixels when the condition $T_{\max} - T_{\min} > \varepsilon$ is satisfied.

c) Secure the algorithm stopping with mostly linearly mixed pixels if condition (b) cannot be satisfied by setting $N_{\max} \geq 10$.

E. Simulation with synthetic data extracted from a real scene

In this section we evaluate the performance of the proposed method using synthetic data that carries the characteristics of real data. While tests using real data are important, the use of synthetic data (for which the ground truth in known) is necessary for a more comprehensive evaluation. To conciliate both needs, we considered a scene corresponding to the alunite hill (depicted in Figure 8a) extracted from the 1997 AVIRIS scene from the Cuprite mining site in Nevada [58]. The chosen region is indicated in Figure 7. The alunite hill site has two interesting properties. First, it has a known number of endmembers ($R = 3$), i.e., alunite, muscovite, and kaolinite. Second, this scene has been accurately unmixed using linear mixing models [59]. To build the synthetic image we used the MVES to linearly unmix the pixels in the image. The reconstructed image considering the LMM is depicted in Figure 8b. The reconstructed image is clearly very similar to the original image, and thus carries its characteristics. To obtain a partly nonlinearly mixed image, we randomly selected 30% of the pixels from the reconstructed image and re-mixed them using the modified GBM model (38) with $\eta_d = 0.3$, but preserving the abundances. Finally, we added a WGN to each pixel with power adjusted to produce a 30dB SNR, which is typical for hyperspectral images. The resulting synthetic image is shown in Figure 8c. This is a partly nonlinearly mixed image for which we know the ground truth and that carries the characteristics of a real image.

(a) First iteration.   (b) Fourth iteration.   (c) Seventh iteration.   (d) Final result.
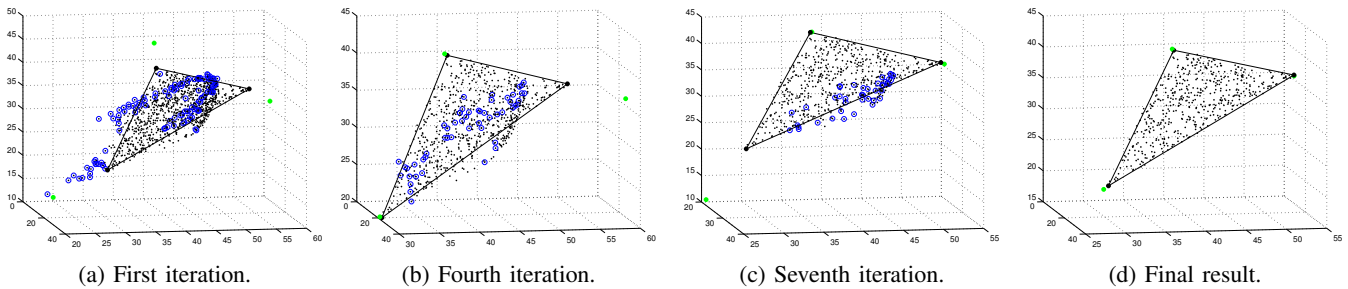
Fig. 6: Graphical illustration of the endmember estimation process using the proposed iterative algorithm. The data set consists of 2000 pixels, with a proportion of 50% nonlinearly mixed pixels obtained with the GMB model and $\eta_d = 0.5$. Green dots are the current estimated endmembers, and black dots are the data projected onto the subspace spanned by the columns of the current matrix M. The true endmembers are shown as black circles at the vertices of the true simplex drawn with black lines. The data discarded at the corresponding iteration are shown within blue circles.
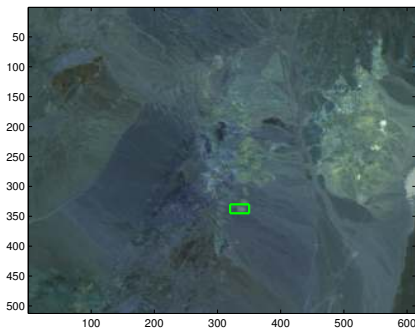
Fig. 7: Cuprite mining site. The green box corresponds to the alunite hill scene.

(a) Alunite hill.   (b) LMM.   (c) GBM + WGN.
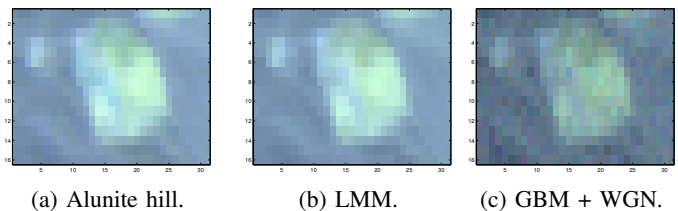
Fig. 8: (a) Plot of the alunite hill with bands 30, 70 and 100. (b) Reconstruction of the scene using the LMM. (c) Adding 30 % of nonlinearly mixed pixels and WGN to give a 30dB SNR.

We applied the proposed EEA to the image of Figure 8c and compared the endmember estimates with those obtained by applying the MVES and the VCA algorithms directly to the image. We considered $N_{\max} = 10$, $\varepsilon = 0.05$, and $r_f = 0.9$. The results are shown in Figure 9. It can be verified that the proposed method has led to the most accurate endmember estimates even for a moderate degree of nonlinearity. Figure 10 shows in black the real endmembers and the data projected into the column space of M. The endmember estimates calculated by proposed EEA after 10 iterations are shown in blue. This figure clearly shows the challenging problem posed to the algorithm, as the chosen degree of nonlinearity introduces a relatively small detachment of the nonlinearly mixed pixels from the simplex. Table VI presents the RMSE for the abundance vectors using the endmembers estimated with the proposed EEA (labelled "detect-then-unmix") and with two alternative unmixing strategies: linear with the FCLS, and nonlinear with the SK-Hype. The improvement obtained using the proposed method is of the order of 18%. For a visual evaluation, Figure 11 compares the true nonlinearity map with the detection map. The white and gray pixels where correctly classified, and the black pixels were misclassified.

F. Real Data

1) Indian Pines: To test the proposed detector using real images, we used the data set available at the Indian Pines test site in North-western Indiana [60]. This image was captured by
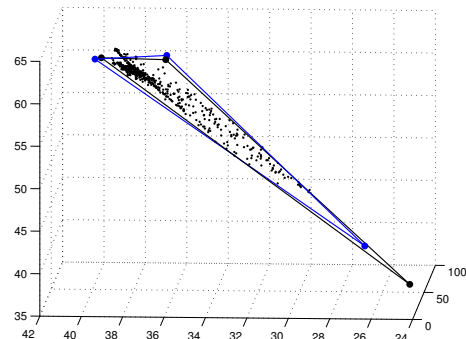

Fig. 10: The black circles are the real endmembers, the black dots are the data projected in the columns of M. The blue circles are the estimated endmembers with the proposed algorithm after 10 iterations. The simplex for the "true" and estimated endmembers are also drawn.

TABLE VI: RMSE for the abundances in the alunite hill scene.

Algorithm	RMSE $\pm$ STD (Class. Err. %)
FCLS	0.0797 $\pm$ 0.0123 (-)
SK-Hype	0.0824 $\pm$ 0.0059 (-)
detect-then-unmix	**0.0671 $\pm$ 0.0049** (3.83)

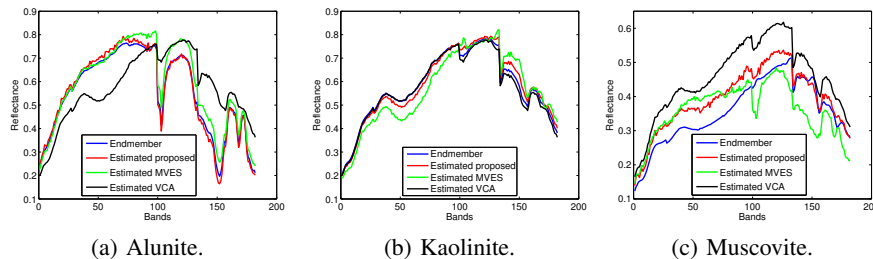(a) Alunite.   (b) Kaolinite.   (c) Muscovite.

Fig. 9: Endmember estimations for the nonlinearly mixed image with different extraction techniques.
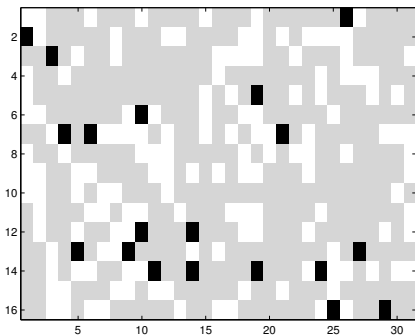

Fig. 11: Detection map and true nonlinear map. Linearly mixed pixels in gray, nonlinearly mixed pixels in white, and misclassified pixels in black.

the AVIRIS (Airborne Visible/Infrared Imaging Spectrometer). It has 145×145 samples over 220 contiguous bands with wavelengths ranging from 366 to 2497 nm. Prior to analysis, noisy and water absorption bands were removed resulting in a total of 200 bands that were uniformly decimated to 50 to speed up simulations. The data set has a ground truth map that divides the samples into 16 mutually exclusive classes. In Table VII, the classes are organized by numbers (1 to 16), and the number of samples of each class is shown. Note, however, that the number of samples in each class can vary considerably. Note also that some classes are composed of different materials. We can count 20 different materials if we consider grass as an isolated material for the whole image. We chose to count each grass (depending on the accompanying material) as a different material, leading to 22 endmembers. Figures 12a and 12b display images from the Indian Pines region constructed by selecting three different bands, while Fig. 13a presents the ground truth map for this image, where each class is represented by a different color. In Figure 13a, we also indice the class number for each area, where 0 represents the background, which is an unclassified area.

To perform the simulations, we divided the image into eight sub-images to work with smaller areas of the image and to deal with 3 to 4 endmembers at a time. To define these sub-images, we also paid attention to balance the number of samples per endmember. By looking at Figs 12a and 12b, we can note that some classes seem to have materials that are not accounted for in the available ground-truth information. For instance, this is the case for classes 5, 11 and 14. Therefore, we introduced extra endmembers for some of the sub-images. Table VIII describes how the sub-images were organized, showing the

classes, materials, numbers of pixels and endmembers chosen for each of the eight sub-images.

For each sub-image, we estimated the endmembers as discussed in Section V, with $N_{\max} = 10$, a relaxing factor initially set to $r_f = 0.8$, and incremented by $r_{\text{inc}} = (1 - r_f)/N_{\max} = 0.2$ at each of the 10 iterations. Then, we ran the detection algorithm with PFA = 0.001. Since we subdivided the real image into different sub-images, some of which have few pixels, we employed a more relaxed value of r_f when compared to previous simulations to avoid discarding too much data in the first few iterations. Moreover, natural phenomena such as endmember variability, wrong (or incomplete) ground truth and illumination factors (among others) tend to degrade the detection performance when dealing with real images, specially when considering nonlinear algorithms which are more susceptible to overfitting. Thus, we have employed a smaller PFA to minimize incorrect detections of linearly mixed samples as nonlinearly mixed. We performed the unmixing step using FCLS for pixels detected as linearly mixed and SK-Hype for pixels detected as nonlinear mixtures. Figure 13b presents the detection map superimposed to the ground-truth classes, where black dots represent pixels detected as nonlinearly mixed.

Comparing the detection map in Fig. 13b with Figs 12a and 12b, one can note similarities between the detection map and some patterns observed in the image representations. For instance, the black triangular shape in class 11 in Fig. 13b (centered at coordinate (40,80)) is just besides what seems to be a road or trail when looking to Figure 12a. Similarities can be found between contours of detected nonlinear regions in Fig. 13b and the corresponding regions in Figs 12a or 12b. Table IX reports the RMSEs for the reconstruction error for each of the eight sub-images using three approaches, namely FCLS, SK-Hype, and detect-then-unmix. The results marked in bold blue correspond to the lowest RMSEs. For almost all sub-images, we note that the use of a nonlinear mixture detector improved the image reconstruction when compared to the pure linear or pure nonlinear unmixing strategies.

2) Cuprite: This example applies the proposed EEA (Algorithm 1) to real data from a scene extracted from the Cuprite Mining site in Nevada (Figure 16a). This scene was captured by the AVIRIS instrument and has originally 224 bands. We removed the water absorption bands and decimated the data uniformly by a factor of 2, resulting in 94 bands. The decimation was carried out to speed up simulations. As reference spectra we selected 18 spectral signatures taken from

TABLE VIII: Subimages organization

Subimage	Classes	Materials	# of pixels	# of endmem.
1	9 and 7	Oats and grass-pasture-mowed	48	3
2	1, 4 and 13	Alfafa, wheat and corn	488	3
3	16	Stone-steel-towers	93	3
4	15	Buildings-grass-trees-drives	386	4
5	5	Grass-Pasture	483	3
6	8 and 12	Hay-windrowed and Soybean-clean	1071	3
7	3,6 and 10	Corn-mintill, grass-trees and soybean-notill	2532	4
8	14 2 11	Woods, corn-notill, soybean-mintill	5148	4

TABLE IX: Indian Pines recontruction error (RMSE) by subimage.

Subimage	RMSE $\pm$ STD		
	FCLS	SK-Hype	detect-then-unmix
1	**0.0028627 $\pm$ 6.6939e-06**	0.0030332 $\pm$ 6.0053e-06	0.0029083 $\pm$ 6.5229e-06
2	0.0038963 $\pm$ 1.2293e-05	0.003881 $\pm$ 9.4813e-06	**0.0038391 $\pm$ 1.1505e-05**
3	0.0044259 $\pm$ 2.9087e-05	0.0035981 $\pm$ 8.9722e-06	**0.0035537 $\pm$ 9.8622e-06**
4	0.0040145 $\pm$ 1.1417e-05	0.0039097 $\pm$ 8.0165e-06	**0.0038895 $\pm$ 8.5058e-06**
5	0.0030848 $\pm$ 7.0516e-06	0.0032353 $\pm$ 5.9761e-06	**0.0030527 $\pm$ 6.2275e-06**
6	0.0039905 $\pm$ 6.5627e-06	0.004055 $\pm$ 7.1531e-06	**0.0039644 $\pm$ 6.6603e-06**
7	0.0034804 $\pm$ 5.8657e-06	0.0035049 $\pm$ 5.9207e-06	**0.0034552 $\pm$ 5.9632e-06**
8	0.0037665 $\pm$ 7.5723e-06	0.0039314 $\pm$ 7.3092e-06	**0.0037531 $\pm$ 7.4932e-06**

TABLE VII: Indian Pines classes by region.

Class number	Class	Num. of Samples
1	Alfalfa	46
2	Corn-notill	1428
3	Corn-mintill	830
4	Corn	237
5	Grass-pasture	483
6	Grass-trees	730
7	Grass-pasture-mowed	28
8	Hay-windrowed	478
9	Oats	20
10	Soybean-notill	972
11	Soybean-mintill	2455
12	Soybean-clean	593
13	Wheat	205
14	Woods	1265
15	Buildings-Grass-Trees-Drives	386
16	Stone-Steel-Towers	93
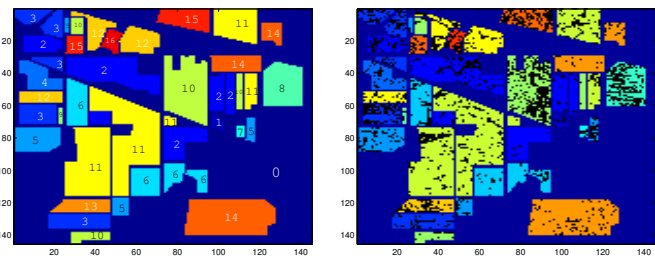
(a) Indian Pines ground truth. (b) Indian Pines detection map.

Fig. 13: Detection of nonlinearly mixed pixels in Indian Pines hyperspectral image. Black pixels were detected as nonlinearly mixed ones by the proposed detector.
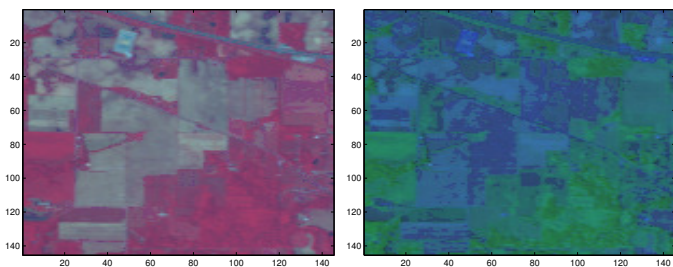
(a) Indian Pines representation (3-band combination #1) (b) Indian Pines representation (3-band combination #2)

Fig. 12: Indian Pines test site representation selecting 3 different bands in (a), and 3 other bands in (b).

the 1998 USGS spectral library. These spectral signatures were selected based on minerals reported to be present in the Cuprite Mining Field [23], [37], [61]. We estimated the number of endmembers using *Virtual Dimension* (VD) [50] with probability of false alarm $P_f = 10^{-4}$, resulting in $R = 5$ endmembers. We performed the endmember estimation using the proposed EEA (IEE), as well as VCA and MVES. We considered also a modification of Algorithm 1 where we replaced the proposed detector with the robust least-squares based detector presented in [31]. We refer to this method as LS for short. The parameter setting for the proposed EEA was $N_{max} = 10$, $\varepsilon = 0.05$, and $r_f = 0.7$.[3] We searched the 18 USGS spectra for the best match (smaller spectral angle) with the endmembers extracted. The endmembers were identified as Sphene, Montmorillonite, Kaolinite, Dumortierite, Pyrope. These endmembers have strong components in this part of the

---

[3]The parameter r_f for the LS detector case was modified to 1.2 to adjust the algorithm to the least-squares detector.

TABLE X: Spectral angles (in rad) between estimated and USGS spectra.

Endmemeber	IEE	LS	VCA	MVES
Sphene	**0.0799**	0.1498	0.3634	0.2457
Montmorillonite	**0.0615**	0.0852	0.0888	0.0762
Kaolinite	**0.1471**	0.1689	0.2022	0.2559
Dumortierite	0.1054	0.1008	**0.0942**	0.1422
Pyrope	**0.1035**	0.9792	0.1760	0.1588
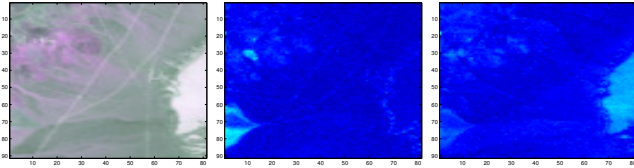
(a) Cuprite scene. (b) Alg.1 + SK-Hype (c) VCA + SK-Hype

Fig. 16: Cuprite scene and reconstruction errors.

Cuprite Mining Field [23]. Figure 14 shows the endmembers estimated with the proposed EEA (red line), with the LS (green lines), and the best matched signatures from USGS spectral library (blue lines). Table X lists the spectral angles, in radians, between the estimated and the library endmembers for the proposed EEA, LS, VCA, and MVES[4]. Clearly, the proposed method presented good estimation performance, outperforming the other methods. Figure 15 presents the abundance maps for the unmixing process using the detect-then-unmix strategy with the GP detector. These abundance maps are in good agreement with abundance maps estimated in [23]. Figure 16 presents the reconstruction error (RMSE) for the Cuprite scene using the proposed EEA (Fig. 16b) and the VCA (Fig. 16c). In both cases the unmixing procedure was carried out using the SK-Hype algorithm. The darker tone dominating Figure 16b indicates a better fitting of the model when compared with Figure 16c. This result is corroborated by the smaller RMSE obtained using the proposed method ($\text{RMSE}_{\text{prop}} = 0.0040$, $\text{RMSE}_{\text{VCA}} = 0.0051$).

VII. Conclusions

This paper proposed a nonparametric method for detecting nonlinear mixtures in hyperspectral images. The performance of the detector was studied for supervised and unsupervised unmixing problems. Additionally, an iterative algorithm was derived for endmember estimation as a pre-processing step for unsupervised unmixing problems. It was shown that the combined use of the proposed detector and endmember estimation algorithm leads to better unmixing results when compared to state-of-the-art solutions. A degree of mixture nonlinearity based on the relative energies of the linear and nonlinear contributions to the mixing process was defined to quantify the importance of the linear and nonlinear model counterparts. Such a definition is important for a proper evaluation of the

[4]Note that the mean spectral angle error used in [23] and [61] as a quality measure for the endmember estimation can be thought as a weighted mean projection of all the image vectors on the estimated endmembers, and therefore does not capture nonlinear relations between pixels and endmembers.

relative performances of different nonlinear mixture detection strategies.

REFERENCES

[1] T. Imbiriba, J.-C. M. Bermudez, J.-Y. Tourneret, and C. Richard, "Detection of nonlinear mixtures using Gaussian processes: Application to hyperspectral imaging," in *Proc. IEEE ICASSP*, Florence, Italy, 2014.

[2] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 2, pp. 6–36, 2013.

[3] D. Landgrebe, "The evolution of landsat data analysis," *Photogrammetric Engineering and Remote Sensing*, vol. LXIII, no. 7, pp. 859–867, 1997.

[4] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 17–28, 2002.

[5] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 44–57, 2002.

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

[7] T. W. Ray and B. C. Murray, "Nonlinear spectral mixing in desert vegetation," *Remote Sensing of Environment*, vol. 55, no. 1, pp. 59–64, 1996.

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

[9] 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, February 2009.

[10] 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 Transactions on Geoscience and Remote Sensing*, vol. 39, no. 10, pp. 2314–2318, 2001.

[11] 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," *International Journal of Remote Sensing*, vol. 30, no. 11, pp. 2951–2962, 2009.

[12] Y. Altmann, A. Halimi, N. Dobigeon, and J.-Y. Tourneret, "A polynomial post nonlinear model for hyperspectral image unmixing," in *Proc. IEEE IGARSS*, Vancouver, Canada, 2011.

[13] J. Chen, C. Richard, and P. Honeine, "Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model," *IEEE Transactions on Signal Processing*, vol. 61, no. 2, pp. 480–492, 2013.

[14] J. Chen, C. Richard, and P. Honeine, "Nonlinear unmixing of hyperspectral images with multi-kernel learning," in *Proc. IEEE WHISPERS*, Shanghai, China, 2012, pp. 1–4.

[15] J. Chen, C. Richard, and P. Honeine, "Estimating abundance fractions of materials in hyperspectral images by fitting a post-nonlinear mixing model," in *Proc. IEEE WHISPERS*, Jun, pp. 1–4.

[16] J. Chen, C. Richard, and P. Honeine, "Nonlinear estimation of material abundances in hyperspectral images with L1-norm spatial regularization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 5, pp. 2654–2665, 2014.

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

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

[19] J. Broadwater, R. Chellappa, A. Banerjee, and P. Burlina, "Kernel fully constrained least squares abundance estimates," in *Proc. IEEE IGARSS*, 2007, pp. 4041–4044.
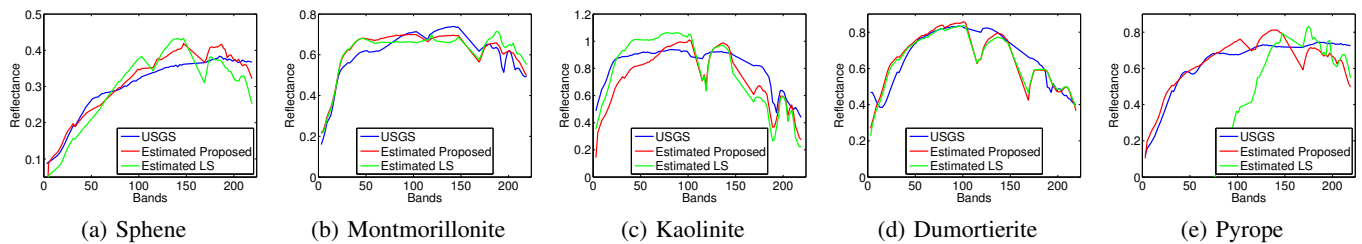
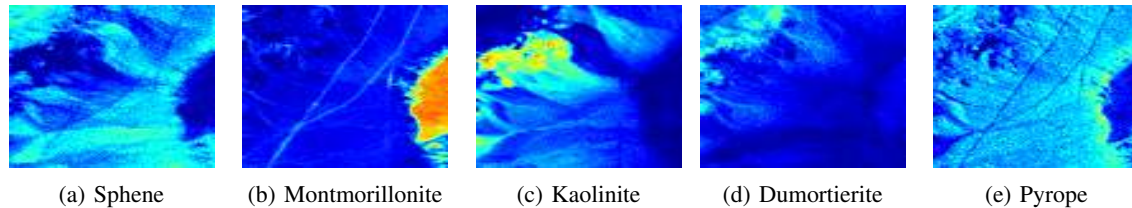Fig. 14: Estimated endmembers and USGS spectra.

Fig. 15: Abundance maps.

[20] J. Broadwater and A. Banerjee, "A comparison of kernel functions for intimate mixture models," in *Proc. IEEE WHISPERS*, Grenoble, France, 2009, pp. 1–4.

[21] Y. Altmann, N. Dobigeon, S. McLaughlin, and J.-Y. Tourneret, "Residual component analysis of hyperspectral images. application to joint nonlinear unmixing and nonlinearity detection," *IEEE Transactions on Image Processing*, vol. 23, pp. 2148–2158, 2014.

[22] J. Boardman, "Automatic spectral unmixing of AVIRIS data using convex geometry concepts," in *Proc. AVIRIS workshop*, 1993, vol. 1, pp. 11–14.

[23] J. M. P. Nascimento and J. M. Bioucas-Dias, "Vertex Component Analysis: A fast algorithm to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, April 2005.

[24] C.-I. Chang, C. C. Wu, W. Liu, and Y. C. Ouyang, "A new growing method for simplex-based endmember extraction algorithm," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 10, pp. 2804–2819, 2006.

[25] T.-H. Chan, C.-Y. Chi, Y.-M. Huang, and W.-K. Ma, "A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4418–4432, 2009.

[26] T.-H. Chan, W.-K. Ma, A. Ambikapathi, and C.-Y. Chi, "A simplex volume maximization framework for hyperspectral endmember extraction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4177–4193, May 2011.

[27] 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, 2011.

[28] Y. Altmann, N. Dobigeon, S. McLaughlin, and J.-Y. Tourneret, "Nonlinear spectral unmixing of hyperspectral images using Gaussian processes," *IEEE Transactions on Signal Processing*, vol. 61, pp. 2442–2453, May 2013.

[29] C. C. Borel and S. A. W. Gerstl, "Nonlinear spectral mixing models for vegetative and soil surfaces," *Remote Sensing of Environment*, vol. 47, no. 3, pp. 403–416, 1994.

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

[31] Y. Altmann, N. Dobigeon, J.-Y. Tourneret, and J.-C. M. Bermudez, "A robust test for nonlinear mixture detection in hyperspectral images," in *Proc. IEEE ICASSP*, Vancouver, Canada, 2013.

[32] Y. Altmann, N. Dobigeon, S. McLaughlin, and J.-Y. Tourneret, "Residual component analysis of hyperspectral images for joint nonlinear unmixing and nonlinearity detection," in *Proc. IEEE ICASSP*, Florence, Italy, 2014, pp. 3166–3170.

[33] N. Gillis and S. A. Vavasis, "Fast and robust recursive algorithms for separable nonnegative matrix factorization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 4, pp. 698–714, 2014.

[34] P. Honeine and C. Richard, "Geometric unmixing of large hyperspectral images: A barycentric coordinate approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 6, pp. 2185–2195, 2012.

[35] M. E. Winter, "N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data," in *Proc. SPIE Spectrometry V*, Denver, CO, USA, 1999, vol. 3753, pp. 266–277.

[36] R. A. Neville, K. Staenz, T. Szeredi, J. Lefebvre, and P. Hauff, "Automatic endmember extraction from hyperspectral data for mineral exploration," in *Proc. 21st Canadian Symp. Remote Sens.*, 1999, pp. 21–24.

[37] R. Ammanouil, A. Ferrari, C. Richard, and D. Mary, "Blind and fully constrained unmixing of hyperspectral images," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5510–5518, 2014.

[38] R. Ammanouil, A. Ferrari, C. Richard, and D. Mary, "GLUP: yet another algorithm for blind unmixing of hyperspectral data," in *Proc. IEEE WHISPERS*, Lausanne, Switzerland, Jun 2014, pp. 1–4.

[39] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 3, pp. 765–777, 2007.

[40] C. Jutten and J. Karhunen, "Advances in nonlinear blind source separation," in *Proc. ICA*, Nara, Japan, 2003, pp. 245–256.

[41] Y. Altmann, A. Halimi, N. Dobigeon, and J.-Y. Tourneret, "Supervised nonlinear spectral unmixing using a polynomial post nonlinear model for hyperspectral imagery," in *Proc. IEEE ICASSP*, Prague, Czech Republic, 2011, pp. 1009–1012.

[42] B. Hapke, *Theory of Reflectance and Emittance Spectroscopy*, Cambridge University Press, 1993.

[43] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, The MIT Press, 2006.

[44] C. E. Rasmussen and H. Nickisch, "Gaussian process for machine learning GPML," http://www.gaussianprocess.org/gpml/code/matlab/doc/, Accessed: 2015-08-24.

[45] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, Mc Graw Hill, fourth edition, 2006.

[46] S. M. Kay, *Fundamentals of statistical signal processing: Detection Theory*, Prentice-Hall, 2011.

[47] N. J. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, vol. 1, Wiley-Interscience, 1994.

[48] Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Unsupervised post-nonlinear unmixing of hyperspectral images using a hamiltonian monte carlo algorithm," *IEEE Transactions on Image Processing*, vol. 23, no. 6, pp. 2663–2675, 2014.

[49] N. H. Nguyen, C. Richard, P. Honeine, and C. Theys, "Hyperspectral image unmixing using manifold learning methods: derivations and comparative tests," in *Proc. IEEE IGARSS*, Munich, Germany, 2012, pp. 3086–3089.

[50] C.-I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 3, pp. 608–619, 2004.

[51] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 8, pp. 2435–2445, 2008.

[52] A. Halimi, P. Honeine, M. Kharouf, C. Richard, and J.-Y. Tourneret, "Estimating the intrinsic dimension of hyperspectral images using an eigen-gap approach," Tech. Rep., arxiv.org/abs/1501.05552, 2015.

[53] A. Ambikapathi, T.-H. Chan, W.-K. Ma, and C.-Y. Chi, "Chance-constrained robust minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4194–4209, 2011.

[54] E. M. T. Hendrix, I. García, J. Plaza, and A. Plaza, "On the minimum volume simplex enclosure problem for estimating a linear mixing model," *Journal of Global Optimization*, pp. 1–14, 2013.

[55] RSI (Research Systems Inc., "Envi user's guide version 4.0," Boulder, Colorado, Sept. 2013.

[56] D. C. Heinz and C.-I. Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 3, pp. 529–545, 2001.

[57] J. P. Marques De Sá, *Applied statistics using SPSS, STATISTICA and MATLAB*, Springer-Verlag Berlin, 2007.

[58] R. N. Clark, G. A. Swayze, K. E. Livo, R. F. Kokaly, S. J. Sutley, J. B. Dalton, R. R. McDougal, and C. A. Gent, "Imaging spectroscopy: Earth and planetary remote sensing with the usgs tetracorder and expert systems," *Journal of Geophysical Research: Planets*, vol. 108, no. E12, 2003, 5131.

[59] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A. O. Hero, "Joint bayesian endmember extraction and linear unmixing for hyperspectral imagery," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4355–4368, 2009.

[60] I. Dopido, M. Zortea, A. Villa, A. Plaza, and P. Gamba, "Unmixing prior to supervised classification of remotely sensed hyperspectral images," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 4, pp. 760–764, 2011.

[61] X. Geng, K. Sun, L. Ji, Y. Zhao, and H. Tang, "Optimizing the endmembers using volume invariant constrained model," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3441–3449, Nov 2015.

Cédric Richard (S'98–M'01–SM'07) received the Dipl.-Ing. and the M.S. degrees in 1994, and the Ph.D. degree in 1998, from Compiègne University of Technology, France, all in electrical and computer engineering. From 1999 to 2003, he was an Associate Professor at Troyes University of Technology, France, and a Full Professor from 2003 to 2009. Since 2009, he is a Full Professor at the University of Nice Sophia Antipolis, France. He is a junior member of the Institut Universitaire de France since 2010. His current research interests include statistical signal processing and machine learning. Cédric Richard is the author of over 230 papers. He was the General Co-Chair of the IEEE SSP Workshop that was held in Nice, France, in 2011. He was the Technical Co-Chair of EUSIPCO 2015 that was held in Nice, France, and of the IEEE CAMSAP Workshop 2015 that was held in Cancun, Mexico. He serves as a Senior Area Editor of the IEEE Transactions on Signal Processing and as an Associate Editor of the IEEE Transactions on Signal and Information Processing over Networks since 2015. He is also an Associate Editor of Signal Processing Elsevier since 2009. Cédric Richard is member of the Machine Learning for Signal Processing (MLSP TC) Technical Committee, and served as member of the Signal Processing Theory and Methods (SPTM TC) Technical Committee in 2009-2014.

Tales Imbiriba (S'14) received the B.E.E. and M.Sc. degrees from the Federal University of Pará (UFPA), Belém, Brazil, in 2006 and 2008 respectively. He is currently pursuing the Ph.D. degree at the Federal University of Santa Catarina (UFSC), Florianópolis, Brazil. His research interests are in digital signal processing, including audio and image processing, pattern recognition, and kernel methods, and adaptive filtering.

José C. M. Bermudez (M'85–SM'02) received the B.E.E. degree from the Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil, the M.Sc. degree in electrical engineering from COPPE/UFRJ, and the Ph.D. degree in electrical engineering from Concordia University, Montreal, Canada, in 1978, 1981, and 1985, respectively. He joined the Department of Electrical Engineering, Federal University of Santa Catarina (UFSC), Florianópolis, Brazil, in 1985. He is currently a Professor of Electrical Engineering. He was a Visiting Researcher with the Department of Electrical Engineering, with the Department of Electrical Engineering and Computer Science, University of California, Irvine (UCI), with Institut National Polytechnique de Toulouse, France, and with Université Nice Sophia-Antipolis, France, in different periods. His research interests have involved analog signal processing using continuous-time and sampled-data systems. His recent research interests are in digital signal processing, including linear and nonlinear adaptive filtering, active noise and vibration control, echo cancellation, image processing, and speech processing. Prof. Bermudez served as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1994 to 1996 and from 1999 to 2001, and of the EURASIP Journal of Advances on Signal Processing from 2006 to 2010. He is presently a Senior Area Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and Associate Editor of the GRETSI journal Traitement du Signal. He is a member of the Signal Processing Theory and Methods Technical Committee of the IEEE Signal Processing Society.

Jean-Yves TOURNERET (SM'08) received the ingénieur degree in electrical engineering from the Ecole Nationale Supérieure d'Electronique, d'Electrotechnique, d'Informatique, d'Hydraulique et des Télécommunications (ENSEEIHT) de Toulouse in 1989 and the Ph.D. degree from the National Polytechnic Institute from Toulouse in 1992. He is currently a professor in the university of Toulouse (ENSEEIHT) and a member of the IRIT laboratory (UMR 5505 of the CNRS). His research activities are centered around statistical signal and image processing with a particular interest to Bayesian and Markov chain Monte Carlo (MCMC) methods. He has been involved in the organization of several conferences including the European conference on signal processing EUSIPCO'02 (program chair), the international conference ICASSP'06 (plenaries), the statistical signal processing workshop SSP'12 (international liaisons), the International Workshop on Computational Advances in Multi-Sensor Adaptive Processing CAMSAP 2013 (local arrangements), the statistical signal processing workshop SSP'2014 (special sessions), the workshop on machine learning for signal processing MLSP'2014 (special sessions). He has been the general chair of the CIMI workshop on optimization and statistics in image processing hold in Toulouse in 2013 (with F. Malgouyres and D. Kouamé) and of the International Workshop on Computational Advances in Multi-Sensor Adaptive Processing CAMSAP 2015 (with P. Djuric). He has been a member of different technical committees including the Signal Processing Theory and Methods (SPTM) committee of the IEEE Signal Processing Society (2001-2007, 2010-present). He has been serving as an associate editor for the IEEE Transactions on Signal Processing (2008-2011, 2015-present) and for the EURASIP journal on Signal Processing (2013-present).