Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model

Jie Chen, Cédric Richard, Paul Honeine

Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model

Jie Chen, *Student Member, IEEE*, Cédric Richard, *Senior Member, IEEE*, Paul Honeine, *Member, IEEE*

***Abstract*—Spectral unmixing is an important issue to analyze remotely sensed hyperspectral data. Although the linear mixture model has obvious practical advantages, there are many situations in which it may not be appropriate and could be advantageously replaced by a nonlinear one. In this paper, we formulate a new kernel-based paradigm that relies on the assumption that the mixing mechanism can be described by a linear mixture of endmember spectra, with additive nonlinear fluctuations defined in a reproducing kernel Hilbert space. This family of models has clear interpretation, and allows to take complex interactions of endmembers into account. Extensive experiment results, with both synthetic and real images, illustrate the generality and effectiveness of this scheme compared with state-of-the-art methods.**

I. Introduction

Hyperspectral imaging is a continuously growing area of remote sensing, which has received considerable attention in the last decade. Hyperspectral data provide a wide spectral range, coupled with a high spectral resolution. These characteristics are suitable for detection and classification of surfaces and chemical elements in the observed images. Applications include land use analysis, pollution monitoring, wide-area reconnaissance, and field surveillance, to cite a few. Due to multiple factors, including the possible low spatial resolution of some hyperspectral-imaging devices, the diversity of materials in the observed scene, the reflections of photons onto several objects, etc., mixed-pixel problems can occur and be critical for proper interpretation of images. Indeed, assigning mixed pixels to a single pure component, or *endmember*, inevitably leads to a loss of information.

Spectral unmixing is an important issue to analyze remotely sensed hyperspectral data. This involves the decomposition of each mixed pixel into its pure endmember spectra, and the estimation of the abundance value for each endmember [1]. Several approaches have been developed for endmember extraction [2]. On the one hand, methods with pure pixel assumption have been proposed to extract the endmembers from pixels in the scene, such as the pixel purity index algorithm [3], the vertex component analysis (VCA) [4], and the N-FINDR algorithm [5], among others [6], [7]. On the other hand, some methods have been proposed to overcome the absence of pure pixels, by generating virtual endmembers, such as the minimum volume simplex analysis (MVSA) [8], the minimum volume enclosing simplex algorithm (MVES) [9], and the minimum volume constrained nonnegative matrix factorization (MVC-NMF) [10]. Endmember identification and abundance estimation can be conducted either in a sequential or collaborative manner. Under the assumption that the endmembers have been identified, hyperspectral image unmixing then reduces to estimating the fractional abundances. The term unmixing in the paper represents the abundance estimation step, which is referred to as the *supervised unmixing* in some literature.

The linear mixture model is widely used to identify and quantify pure components in remotely sensed images due to its simple physical interpretation and trackable estimation process. To be physically interpretable, the driving abundances are often required to satisfy two constraints: all abundances must be nonnegative, and their sum must be equal to one. In addition to the extremely low-complexity method that has been recently proposed [7], which is based on geometric considerations, at least two classes of approaches can be distinguished to determine abundances. On the one hand, there are estimation methods that lead to an optimization problem which must be solved subject to non-negativity and sum-to-one constraints [11]. On the other hand, following the principles of Bayesian inference, there are simulation techniques that define prior distributions for abundances, and estimate unknown parameters based on the resulting joint posterior distribution [12], [13], [14], [15]. Some recent works also take sparsity constraints into account in the unmixing process [2], [15], [16], [17], [18].

Although the linear mixture model has obvious practical advantages, there are many situations in which it may not be appropriate (e.g., involving multiple light scattering effects) and could be advantageously replaced by a nonlinear one. For instance, multiple scattering effects can be observed on complex vegetated surfaces [19] where it is assumed that incident solar radiation is scattered by the scene through multiple bounces involving several endmembers. Some nonlinear mixture models, such as the *generalized bilinear model* studied in [20], account for presence of multi-photon interactions by introducing additional interaction terms in the linear model. Another typical situation is the case where the components of interest are in an intimate association, and the photons interact with all the materials simultaneously as they are multiply scattered. A bidirectional reflectance model based on the fundamental principles of radiative transfer theory was proposed in [21] to describe these interactions. It is usually referred to as the *intimate mixture model*. Obviously, the mixture mechanism in a real scene may be much more complex than the above models and often relies on scene parameters that are difficult to obtain.

Nonlinear unmixing has generated considerable interest among researchers, and different methods have been pro-

J. Chen and P. Honeine are with the Institut Charles Delaunay, CNRS, Université de Technologie de Troyes, France (e-mail: chenjieg@sina.com, paul.honeine@utt.fr)

C. Richard is with the Université de Nice Sophia-Antipolis, CNRS, Observatoire de la Côte d'Azur, France (e-mail: cedric.richard@unice.fr)

posed to account for nonlinear effects. Using training-based approaches is a way to bypass difficulties with unknown mixing mechanism and parameters. In [22], a radial basis function neural network was used to unmix intimate mixtures. In [23], the authors designed a multi-layer perceptron neural network combined with a Hopfield neural network to deal with nonlinear mixtures. In [6], [24], the authors discussed methods for automatic selection and labeling of training samples. These methods require the networks to be trained using pixels with known abundances, and the quality of the training data may affect the performance notably. Moreover, for a new set of spectra in a scene, or different embedded parameters, a new neural network should be trained again before unmixing can be performed. Approaches that do not require training samples were also studied in the literature. In [20], a nonlinear unmixing algorithm for the general bilinear mixture model was proposed. Based on Bayesian inference, this method however has a high computational complexity and is dedicated to the bilinear model. In [25], [26], the authors extended the collection of endmembers by adding artificial cross-terms of pure signatures to model light scattering effects on different materials. However, it is not easy to identify which cross-terms should be selected and added to the endmember dictionary. If all the possible cross-terms were considered, the set of endmembers would expand dramatically. In [27], the authors addressed the nonlinear unmixing problem with an intimate mixture model. The proposed method first converts observed reflectances into albedo using a look-up table, then a linear algorithm estimates the endmember albedos and the mass fractions for each sample. This method is based on the hypothesis that all the parameters of the intimate mixture model are known. Nonlinear algorithms operating in reproducing kernel Hilbert spaces (RKHS) have been a topic of considerable interest in the machine learning community, and have proved their efficiency in solving nonlinear problems. Kernel-based methods have already been considered for detection and classification in hyperspectral images [28], [29]. Kernel-based nonlinear unmixing approaches have also been investigated [30], [31], [32]. These algorithms were mainly obtained by replacing each inner product between endmember spectra, in the cost functions to be optimized, by a kernel function. This can be viewed as a nonlinear distortion map applied to the spectral signature of each material, independently of their interactions. This principle may be extremely efficient in solving detection and classification problems as a proper distortion can increase the detectability or separability of some patterns. It is however of little physical interest in solving the unmixing problem because the nonlinear nature of mixing is not only governed by individual spectral distortions, but also by nonlinear interactions of the materials.

In this paper, we formulate the problem of estimating abundances of a nonlinear mixture of hyperspectral data. This new kernel-based paradigm allows to take nonlinear interactions of the endmembers into account. It leads to a more meaningful interpretation of the unmixing mechanism than existing kernel based methods. The abundances are determined by solving an appropriate kernel-based regression problem under constraints. This paper is organized as follows. Section II introduces the basic concepts of our modeling approach. Section III presents a new kernel-based hyperspectral mixture model, called K-Hype, and the associated identification algorithm to extract the abundances within this nonlinear context. The balance between linear and nonlinear contributions is unfortunately fixed in K-Hype. In order to overcome this drawback, a natural generalization called SuperK-Hype or SK-Hype, is then largely described. It is based on the concept of Multiple Kernel Learning, and allows to automatically adapt the balance between linear spectral interactions and nonlinear ones. Finally, major differences with some existing works on kernel-based processing of hyperspectral images are also pointed out. In Section IV, experiments are conducted using both synthetic and real images. Performance comparisons with other popular methods are also provided. Section V concludes this paper and gives a short outlook onto future work.

II. A KERNEL-BASED NONLINEAR UNMIXING PARADIGM

Let $\boldsymbol{r} = [r_1, r_2, \ldots, r_L]^\top$ be an observed column pixel, supposed to be a mixture of R endmember spectra, with L the number of spectral bands. Assume that $\boldsymbol{M} = [\boldsymbol{m}_1, \boldsymbol{m}_2, \ldots, \boldsymbol{m}_R]$ is the $(L \times R)$ target endmember matrix, where each column $\boldsymbol{m}_i$ is an endmember spectral signature. For the sake of convenience, we shall denote by $\boldsymbol{m}_{\lambda_\ell}^\top$ the ℓ-th $(1 \times R)$ row of $\boldsymbol{M}$, that is, the vector of the R endmember signatures at the ℓ-th wavelength band. Let $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_R]^\top$ be the abundance column vector associated with the pixel $\boldsymbol{r}$.

We first consider the linear mixing model where any observed pixel is a linear combination of the endmembers, weighted by the fractional abundances, that is,

$$\boldsymbol{r} = \boldsymbol{M}\boldsymbol{\alpha} + \boldsymbol{n} \tag{1}$$

where $\boldsymbol{n}$ is a noise vector. Under the assumption that the endmember matrix $\boldsymbol{M}$ is known, the vector $\boldsymbol{\alpha}$ of fractional abundances is usually determined by minimizing a cost function of the form

$$\begin{aligned} J(\boldsymbol{\alpha}) &= J_{\text{reg}}(\boldsymbol{\alpha}) + \frac{1}{2\mu}\|\boldsymbol{r} - \boldsymbol{M}\boldsymbol{\alpha}\|^2 \\ &= J_{\text{reg}}(\boldsymbol{\alpha}) + \frac{1}{2\mu}\sum_{\ell=1}^{L}(r_\ell - \boldsymbol{\alpha}^\top \boldsymbol{m}_{\lambda_\ell})^2 \end{aligned} \tag{2}$$

under the non-negativity and sum-to-one constraints[1]

$$\begin{aligned} &\alpha_i \geq 0, \qquad \forall i \in 1, \ldots, R \\ &\sum_{i=1}^{R} \alpha_i = 1, \end{aligned} \tag{3}$$

where $J_{\text{reg}}(\cdot)$ is a regularization function, and μ is a small positive parameter that controls the trade-off between regularization and fitting. The above analysis assumes that the relationship between $\boldsymbol{m}_{\lambda_\ell}$ and r_ℓ is dominated by a linear function. There are however many situations, involving multiple scattering effects, in which model (1) may be inappropriate and could be advantageously replaced by a nonlinear one.

[1]For ease of notation, these two constraints will be denoted by $\boldsymbol{\alpha} \succeq \mathbf{0}$ and $\mathbf{1}^\top\boldsymbol{\alpha} = 1$, where $\mathbf{1}$ is a vector of ones.

Consider the general mixing mechanism

$$\boldsymbol{r} = \boldsymbol{\Psi}(\boldsymbol{M}) + \boldsymbol{n} \tag{4}$$

with $\boldsymbol{\Psi}$ an unknown nonlinear function that defines the interactions between the endmembers in matrix $\boldsymbol{M}$. This requires us to consider a more general problem of the form

$$\psi^* = \arg\min_{\psi\in\mathcal{H}} \frac{1}{2}\|\psi\|^2_{\mathcal{H}} + \frac{1}{2\mu}\sum_{\ell=1}^{L}(r_\ell - \psi(\boldsymbol{m}_{\lambda_\ell}))^2 \tag{5}$$

with $\mathcal{H}$ a given functional space, and μ a positive parameter that controls the trade-off between regularity of the function ψ and fitting. Clearly, this basic strategy may fail if the functionals ψ of $\mathcal{H}$ cannot be adequately and finitely parameterized. Kernel-based methods rely on mapping data from the original input space into a feature space by means of a nonlinear function, and then solving a linear problem in that new space. They lead to efficient and accurate resolution of the inverse problem (5), as it has been showed in the literature. See, e.g., [33], [34]. Our paper exploits the central idea of this research area, known as the *kernel trick*, to investigate new nonlinear unmixing algorithms. We shall now review the main definitions and properties related to reproducing kernel Hilbert spaces [35] and Mercer kernels [36].

Let $\mathcal{H}$ denote a Hilbert space of real-valued functions ψ on a compact $\mathcal{M}$, and let $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ be the inner product in the space $\mathcal{H}$. Suppose that the evaluation functional δ_{m_λ} defined by $\delta_{m_\lambda}[\psi] = \psi(\boldsymbol{m}_\lambda)$ is linear with respect to ψ and bounded, for all $\boldsymbol{m}_\lambda$ in $\mathcal{M}$. By virtue of the Riesz representation theorem, there exists a unique positive definite function $\boldsymbol{m}_\lambda \mapsto \kappa(\boldsymbol{m}_\lambda, \boldsymbol{m}_{\lambda_j})$ in $\mathcal{H}$, denoted by $\kappa(\cdot, \boldsymbol{m}_{\lambda_j})$ and called *representer of evaluation* at $\boldsymbol{m}_{\lambda_j}$, which satisfies [35]

$$\psi(\boldsymbol{m}_{\lambda_j}) = \langle\psi, \kappa(\cdot, \boldsymbol{m}_{\lambda_j})\rangle_{\mathcal{H}}, \quad \forall\psi\in\mathcal{H} \tag{6}$$

for every fixed $\boldsymbol{m}_{\lambda_j} \in \mathcal{M}$. A proof of this may be found in [35]. Replacing ψ by $\kappa(\cdot, \boldsymbol{m}_{\lambda_i})$ in (6) yields

$$\kappa(\boldsymbol{m}_{\lambda_i}, \boldsymbol{m}_{\lambda_j}) = \langle\kappa(\cdot, \boldsymbol{m}_{\lambda_i}), \kappa(\cdot, \boldsymbol{m}_{\lambda_j})\rangle_{\mathcal{H}} \tag{7}$$

for all $\boldsymbol{m}_{\lambda_i}$, $\boldsymbol{m}_{\lambda_j} \in \mathcal{M}$. Equation (7) is the origin of the generic term *reproducing kernel* to refer to κ. Denoting by $\boldsymbol{\Phi}$ the map that assigns the kernel function $\kappa(\cdot, \boldsymbol{m}_{\lambda_j})$ to each input data $\boldsymbol{m}_{\lambda_j}$, equation (7) immediately implies that $\kappa(\boldsymbol{m}_{\lambda_i}, \boldsymbol{m}_{\lambda_j}) = \langle\boldsymbol{\Phi}(\boldsymbol{m}_{\lambda_i}), \boldsymbol{\Phi}(\boldsymbol{m}_{\lambda_j})\rangle_{\mathcal{H}}$. The kernel thus evaluates the inner product of any pair of elements of $\mathcal{M}$ mapped to the space $\mathcal{H}$ without any explicit knowledge of $\boldsymbol{\Phi}$ and $\mathcal{H}$. Within the machine learning area, this key idea is known as the *kernel trick*.

The kernel trick has been widely used to transform linear algorithms expressed only in terms of inner products into nonlinear ones. Considering again (5), the optimum function ψ^* can be obtained by solving the following least squares support vector machines (LS-SVM) problem [37]

$$\begin{aligned} \psi^* = &\arg\min_{\psi\in\mathcal{H}} \frac{1}{2}\|\psi\|^2_{\mathcal{H}} + \frac{1}{2\mu}\sum_{\ell=1}^{L} e_\ell^2 \\ &\text{subject to} \quad e_\ell = r_\ell - \psi(\boldsymbol{m}_{\lambda_\ell}), \quad \ell\in\{1,\ldots,L\} \end{aligned} \tag{8}$$

We introduce the Lagrange multipliers β_ℓ, and consider the Lagrange function

$$\mathcal{L} = \frac{1}{2}\|\psi\|^2_{\mathcal{H}} + \frac{1}{2\mu}\sum_{\ell=1}^{L} e_\ell^2 - \sum_{\ell=1}^{L}\beta_\ell\,(e_\ell - r_\ell + \psi(\boldsymbol{m}_{\lambda_\ell})). \tag{9}$$

The conditions for optimality with respect to the primal variables are given by

$$\begin{cases} \psi^* = \sum_{\ell=1}^{L}\beta_\ell^*\,\kappa(\cdot, \boldsymbol{m}_{\lambda_\ell}) \\ e_\ell^* = \mu\,\beta_\ell^* \end{cases} \tag{10}$$

We then derive the dual optimization problem

$$\boldsymbol{\beta}^* = \arg\max_{\boldsymbol{\beta}} -\frac{1}{2}\,\boldsymbol{\beta}^\top(\boldsymbol{K} + \mu\boldsymbol{I})\,\boldsymbol{\beta} + \boldsymbol{\beta}^\top\boldsymbol{r}, \tag{11}$$

where $\boldsymbol{K}$ is the so-called Gram matrix whose (ℓ, p)-th entry is defined by $\kappa(\boldsymbol{m}_{\lambda_\ell}, \boldsymbol{m}_{\lambda_p})$. Classic examples of kernels are the radially Gaussian kernel $\kappa(\boldsymbol{m}_{\lambda_\ell}, \boldsymbol{m}_{\lambda_p}) = \exp\big(-\|\boldsymbol{m}_{\lambda_\ell} - \boldsymbol{m}_{\lambda_p}\|^2/2\sigma^2\big)$, and the Laplacian kernel $\kappa(\boldsymbol{m}_{\lambda_\ell}, \boldsymbol{m}_{\lambda_p}) = \exp\big(-\|\boldsymbol{m}_{\lambda_\ell} - \boldsymbol{m}_{\lambda_p}\|/\sigma\big)$, with $\sigma \geq 0$ the kernel bandwidth. Another example of interest is the q-th degree polynomial kernel $\kappa(\boldsymbol{m}_{\lambda_\ell}, \boldsymbol{m}_{\lambda_p}) = (1 + \boldsymbol{m}_{\lambda_\ell}^\top\boldsymbol{m}_{\lambda_p})^q$, with $q\in\mathbb{N}^*$.

The kernel function κ maps $\boldsymbol{m}_{\lambda_\ell}$ into a very high, even infinite, dimensional space $\mathcal{H}$ without any explicit knowledge of the associated nonlinear function. The vector $\boldsymbol{\beta}^*$ and κ then describe the relation between the endmembers and the observation. The goal of the analysis is however to estimate the abundance vector, and there is no direct relation between $\boldsymbol{\alpha}^*$ and $\boldsymbol{\beta}^*$ in the general case. In what follows, we shall focus attention on the design of specific kernels that enable us to determine abundance fractions within this context.

III. Kernel design and unmixing algorithms

The aim of this section is to propose kernel design methods and the corresponding algorithms to estimate abundances. The two approaches described hereafter are flexible enough to capture wide classes of nonlinear relationships, and to reliably interpret a variety of experimental measurements. Both have clear interpretation.

A. A preliminary approach for kernel-based hyperspectral unmixing: the K-Hype algorithm

In order to extract the mixing ratios of the endmembers, we define the function ψ in (5) by a linear trend parameterized by the abundance vector $\boldsymbol{\alpha}$, combined with a nonlinear fluctuation term, namely,

$$\begin{aligned} \psi(\boldsymbol{m}_{\lambda_\ell}) &= \boldsymbol{\alpha}^\top\boldsymbol{m}_{\lambda_\ell} + \psi_{\text{nlin}}(\boldsymbol{m}_{\lambda_\ell}) \\ &\text{subject to} \quad \boldsymbol{\alpha}\succeq\boldsymbol{0} \quad \text{and} \quad \boldsymbol{1}^\top\boldsymbol{\alpha} = 1 \end{aligned} \tag{12}$$

where ψ_{nlin} can be any real-valued functions on a compact $\mathcal{M}$, of a reproducing kernel Hilbert space denoted by $\mathcal{H}_{\text{nlin}}$. Let κ_{nlin} be its reproducing kernel. It can be shown [38] that, as the direct sum $\mathcal{H}_{\text{lin}} \oplus \mathcal{H}_{\text{nlin}}$ of the RKHS of kernels $\kappa_{\text{lin}}(\boldsymbol{m}_{\lambda_\ell}, \boldsymbol{m}_{\lambda_p}) = \boldsymbol{m}_{\lambda_\ell}^\top\boldsymbol{m}_{\lambda_p}$ and $\kappa_{\text{nlin}}(\boldsymbol{m}_{\lambda_\ell}, \boldsymbol{m}_{\lambda_p})$ defined on

$\mathcal{M}$, the space $\mathcal{H}$ of functions of the form (12) is also a RKHS with kernel function

$$\begin{aligned}\kappa(\boldsymbol{m}_{\lambda_\ell}, \boldsymbol{m}_{\lambda_p}) &= (\kappa_{\text{lin}} \oplus \kappa_{\text{nlin}})(\boldsymbol{m}_{\lambda_\ell}, \boldsymbol{m}_{\lambda_p}) \\ &= \boldsymbol{m}_{\lambda_\ell}^\top \boldsymbol{m}_{\lambda_p} + \kappa_{\text{nlin}}(\boldsymbol{m}_{\lambda_\ell}, \boldsymbol{m}_{\lambda_p}).\end{aligned} \tag{13}$$

The corresponding Gram matrix $\boldsymbol{K}$ is given by

$$\boldsymbol{K} = \boldsymbol{M}\boldsymbol{M}^\top + \boldsymbol{K}_{\text{nlin}} \tag{14}$$

where $\boldsymbol{K}_{\text{nlin}}$ is the Gram matrix associated with the nonlinear map ψ_{nlin}, with (ℓ, p)-th entry $\kappa_{\text{nlin}}(\boldsymbol{m}_{\lambda_\ell}, \boldsymbol{m}_{\lambda_p})$.

We propose to conduct hyperspectral data unmixing by solving the following convex optimization problem

$$\begin{aligned}&\psi^* = \arg\min_{\psi} \frac{1}{2}\left(\|\psi_{\text{lin}}\|^2_{\mathcal{H}_{\text{lin}}} + \|\psi_{\text{nlin}}\|^2_{\mathcal{H}_{\text{nlin}}}\right) + \frac{1}{2\mu}\sum_{\ell=1}^{L} e_\ell^2 \\ &\quad \text{where} \quad \psi = \psi_{\text{lin}} + \psi_{\text{nlin}} \quad \text{with} \quad \psi_{\text{lin}}(\boldsymbol{m}_{\lambda_\ell}) = \boldsymbol{\alpha}^\top \boldsymbol{m}_{\lambda_\ell} \\ &\quad \text{subject to} \quad e_\ell = r_\ell - \psi(\boldsymbol{m}_{\lambda_\ell}) \\ &\qquad\qquad \boldsymbol{\alpha} \succeq \boldsymbol{0} \quad \text{and} \quad \boldsymbol{1}^\top \boldsymbol{\alpha} = 1\end{aligned} \tag{15}$$

By the strong duality property, we can derive a dual problem that has the same solution as the above primal problem. Let us introduce the Lagrange multipliers β_ℓ, γ_r and λ. The Lagrange function associated with the problem (15) can be written as

$$\begin{aligned}G = &\frac{1}{2}\left(\|\boldsymbol{\alpha}\|^2 + \|\psi_{\text{nlin}}\|^2_{\mathcal{H}_{\text{nlin}}}\right) + \frac{1}{2\mu}\sum_{\ell=1}^{L} e_\ell^2 \\ &- \sum_{\ell=1}^{L} \beta_\ell \left(e_\ell - r_\ell + \psi(\boldsymbol{m}_{\lambda_\ell})\right) - \sum_{r=1}^{R} \gamma_r \alpha_r + \lambda(\boldsymbol{1}^\top \boldsymbol{\alpha} - 1)\end{aligned} \tag{16}$$

with $\gamma_r \geq 0$. We have used that $\|\psi_{\text{lin}}\|^2_{\mathcal{H}_{\text{lin}}} = \|\boldsymbol{\alpha}\|^2$ because the functional space $\mathcal{H}_{\text{lin}}$, parametrized by $\boldsymbol{\alpha}$, contains all the function of the variable $\boldsymbol{m}_{\lambda_\ell}$ of the form $\psi_{\text{lin}}(\boldsymbol{m}_{\lambda_\ell}) = \boldsymbol{\alpha}^\top \boldsymbol{m}_{\lambda_\ell}$. It is characterized by the norm

$$\|\psi_{\text{lin}}\|^2_{\mathcal{H}_{\text{lin}}} = \langle \kappa_{\text{lin}}(\boldsymbol{\alpha}, \cdot), \kappa_{\text{lin}}(\boldsymbol{\alpha}, \cdot)\rangle_{\mathcal{H}} = \kappa_{\text{lin}}(\boldsymbol{\alpha}, \boldsymbol{\alpha}) = \|\boldsymbol{\alpha}\|^2. \tag{17}$$

The conditions for optimality of G with respect to the primal variables are given by

$$\begin{cases} \boldsymbol{\alpha}^* = \sum_{\ell=1}^{L} \beta_\ell^* \, \boldsymbol{m}_{\lambda_\ell} + \boldsymbol{\gamma}^* - \lambda^* \boldsymbol{1} \\ \psi^*_{\text{nlin}} = \sum_{\ell=1}^{L} \beta_\ell^* \, \kappa_{\text{nlin}}(\cdot, \boldsymbol{m}_{\lambda_\ell}) \\ e_\ell^* = \mu \, \beta_\ell^* \end{cases} \tag{18}$$

By substituting (18) into (16), we get the dual problem (19) (next page).

Provided that the coefficient vector $\boldsymbol{\beta}^*$ has been determined, the measured pixel can be reconstructed using

$$\begin{aligned}\boldsymbol{r}^* &= [\psi^*(\boldsymbol{m}_{\lambda_1}), \ldots, \psi^*(\boldsymbol{m}_{\lambda_L})]^\top \\ &= \boldsymbol{M}(\boldsymbol{M}^\top \boldsymbol{\beta}^* + \boldsymbol{\gamma}^* - \lambda^* \boldsymbol{1}) + \boldsymbol{K}_{\text{nlin}} \, \boldsymbol{\beta}^*\end{aligned} \tag{20}$$

as indicated by (18). Comparing the above expression with (12), we observe that the first and the second term of the r.h.s. of equation (20) correspond to the linear trend and the nonlinear fluctuations, respectively. Finally, the abundance vector $\boldsymbol{\alpha}^*$ can be estimated as follows

$$\boldsymbol{\alpha}^* = \boldsymbol{M}^\top \boldsymbol{\beta}^* + \boldsymbol{\gamma}^* - \lambda^* \boldsymbol{1} \tag{21}$$

Problem (19) is a quadratic program (QP). Numerous candidate methods exist to solve it, such as interior point, active set and projected gradient, as presented in [39], [40]. These well known numerical procedures lie beyond the scope of this paper.

B. Some remarks on kernel selection

Selecting an appropriate kernel is of primary importance as it captures the nonlinearity of the mixture model. Though an infinite variety of possible kernels exists, it is always desirable to select a kernel that is closely related to the application context. The following example justifies the combination (12), which associates a linear model with a nonlinear fluctuation term. It also allows us to define a possible family of appropriate kernels for data unmixing.

Consider the generalized bilinear mixing model presented in [20], at first, limited to three endmember spectra for the sake of clarity

$$\begin{aligned}\boldsymbol{r} = &\boldsymbol{M}\,\boldsymbol{\alpha} + \gamma_{12}\,\alpha_1\,\alpha_2\,(\boldsymbol{m}_1 \otimes \boldsymbol{m}_2) + \gamma_{13}\,\alpha_1\,\alpha_3\,(\boldsymbol{m}_1 \otimes \boldsymbol{m}_3) \\ &+ \gamma_{23}\,\alpha_2\,\alpha_3\,(\boldsymbol{m}_2 \otimes \boldsymbol{m}_3) + \boldsymbol{n} \\ &\text{with} \quad \alpha_1, \alpha_2, \alpha_3 \geq 0 \\ &\qquad \alpha_1 + \alpha_2 + \alpha_3 = 1\end{aligned} \tag{22}$$

where γ_{12}, γ_{13} and γ_{23} are attenuation parameters, and $\otimes$ the Hadamard product. It can be observed that the nonlinear term with respect to $\boldsymbol{\alpha}$, in the r.h.s. of (22), is closely related to the homogeneous polynomial kernel of degree 2, that is, $\kappa_{\text{nlin}}(\boldsymbol{m}_{\lambda_\ell}, \boldsymbol{m}_{\lambda_p}) = (\boldsymbol{m}_{\lambda_\ell}^\top \boldsymbol{m}_{\lambda_p})^2$. Indeed, with a slight abuse of notation, the latter can be written in an inner product form as follows

$$\kappa_{\text{nlin}}(\boldsymbol{m}_{\lambda_\ell}, \boldsymbol{m}_{\lambda_p}) = \boldsymbol{\Phi}_{\text{nlin}}(\boldsymbol{m}_{\lambda_\ell})^\top \boldsymbol{\Phi}_{\text{nlin}}(\boldsymbol{m}_{\lambda_p}) \tag{23}$$

with

$$\begin{aligned}\boldsymbol{\Phi}_{\text{nlin}}(\boldsymbol{m}_{\lambda_\ell}) = &(m^2_{\lambda_{\ell,1}},\, m^2_{\lambda_{\ell,2}},\, m^2_{\lambda_{\ell,3}},\, \sqrt{2} m_{\lambda_{\ell,1}} m_{\lambda_{v,2}}, \\ &\sqrt{2} m_{\lambda_{\ell,1}} m_{\lambda_{\ell,3}},\, \sqrt{2} m_{\lambda_{\ell,2}} m_{\lambda_{\ell,3}})^\top\end{aligned} \tag{24}$$

where $m_{\lambda_{\ell,i}}$ is the i-th entry of $\boldsymbol{m}_{\lambda_\ell}$. This means that, in addition to the linear mixture term $\boldsymbol{M}\boldsymbol{\alpha}$, the auto and interaction terms considered by the kernel-based model are of the form $\boldsymbol{m}_i \otimes \boldsymbol{m}_j$ for all $i, j = 1, \ldots, R$.

By virtue of the reproducing kernel machinery, endmember spectra do not need to be explicitly mapped into the feature space. This allows to consider complex interaction mechanisms by changing the kernel κ_{nlin}, without having to modify the optimization algorithm described in the previous subsection. As an illustration, consider the polynomial kernel $\kappa_{\text{nlin}}(\boldsymbol{m}_{\lambda_\ell}, \boldsymbol{m}_{\lambda_p}) = (1 + \boldsymbol{m}_{\lambda_\ell}^\top \boldsymbol{m}_{\lambda_p})^q$. Making use of the binomial theorem yields

$$\kappa_{\text{nlin}}(\boldsymbol{m}_{\lambda_\ell}, \boldsymbol{m}_{\lambda_p}) = \sum_{k=0}^{q} \binom{q}{k} (\boldsymbol{m}_{\lambda_\ell}^\top \boldsymbol{m}_{\lambda_p})^k. \tag{25}$$

We observe that each component $(\boldsymbol{m}_{\lambda_\ell}^\top \boldsymbol{m}_{\lambda_p})^k = (m_{\lambda_{\ell,1}}\, m_{\lambda_{p,1}} + \ldots + m_{\lambda_{\ell,R}}\, m_{\lambda_{p,R}})^k$ of the above expression

$$\max_{\boldsymbol{\beta},\boldsymbol{\gamma},\lambda} G'(\boldsymbol{\beta},\boldsymbol{\gamma},\lambda) = -\frac{1}{2}\begin{pmatrix}\boldsymbol{\beta}\\ \hline \boldsymbol{\gamma}\\ \hline \lambda\end{pmatrix}^\top \left(\begin{array}{c|c|c} \boldsymbol{K}+\mu\boldsymbol{I} & \boldsymbol{M} & -\boldsymbol{M}\boldsymbol{1} \\ \hline \boldsymbol{M}^\top & \boldsymbol{I} & -\boldsymbol{1} \\ \hline -\boldsymbol{1}^\top\boldsymbol{M}^\top & -\boldsymbol{1}^\top & R \end{array}\right)\begin{pmatrix}\boldsymbol{\beta}\\ \hline \boldsymbol{\gamma}\\ \hline \lambda\end{pmatrix} + \begin{pmatrix}\boldsymbol{r}\\ \hline \boldsymbol{0}\\ \hline -1\end{pmatrix}^\top\begin{pmatrix}\boldsymbol{\beta}\\ \hline \boldsymbol{\gamma}\\ \hline \lambda\end{pmatrix} \quad (19)$$
$$\text{subject to} \quad \boldsymbol{\gamma} \succeq \boldsymbol{0}$$
$$\text{with } \boldsymbol{K} = \boldsymbol{M}\boldsymbol{M}^\top + \boldsymbol{K}_{\text{nlin}}.$$

can be expanded into a weighted sum of k-th degree monomials of the form

$$(m_{\lambda_{\ell,1}}\, m_{\lambda_{p,1}})^{k_1}(m_{\lambda_{\ell,2}}\, m_{\lambda_{p,2}})^{k_2}\, \ldots\, (m_{\lambda_{\ell,R}}\, m_{\lambda_{p,R}})^{k_R} \quad (26)$$

with $\sum_{r=1}^R k_r = k$. This means that, in addition to the linear mixture term $\boldsymbol{M}\boldsymbol{\alpha}$, the auto and interaction terms considered by the kernel-based model are of the form $\boldsymbol{m}_1^{k_1} \otimes \boldsymbol{m}_2^{k_2} \otimes \ldots \otimes \boldsymbol{m}_R^{k_R}$ for every set of exponents in the Hadamard sense satisfying $0 \leq \sum_{r=1}^R k_r \leq q$. Note that it would be computationally expensive to explicitly form these interaction terms. Their number is indeed very large: there are p^k monomials (26) of degree k, and then $\frac{1-R^{q+1}}{1-R}$ components in the entire q-th order representation. Compared with the methods introduced in [25], [26], which insert products of pure material signatures as new endmembers, we do not need to extend the endmember matrix by adding such terms. The kernel trick makes the computation much more tractable. In the experimentations reported hereafter, the following 2-nd degree polynomial kernel was used

$$\kappa_{\text{nlin}}(\boldsymbol{m}_{\lambda_\ell}, \boldsymbol{m}_{\lambda_p}) = \left(1 + \frac{1}{R^2}(\boldsymbol{m}_{\lambda_\ell} - 1/2)^\top(\boldsymbol{m}_{\lambda_p} - 1/2)\right)^2 \quad (27)$$

where the constants $1/R^2$ and $1/2$ serve the purpose of normalization.

C. Nonlinear unmixing by multiple kernel learning: the SK-Hype algorithm

The proposed model relies on the assumption that the mixing mechanism can be described by a linear mixture of endmember spectra, with additive nonlinear fluctuations ψ_{nlin} defined in a RKHS. This justifies the use of a Gram matrix of the form $\boldsymbol{K} = \boldsymbol{M}\boldsymbol{M}^\top + \boldsymbol{K}_{\text{nlin}}$ in the algorithm presented previously. Model (12) however has some limitations in that the balance between the linear component $\boldsymbol{\alpha}^\top\boldsymbol{m}_{\lambda_\ell}$ and the nonlinear component $\psi_{\text{nlin}}(\boldsymbol{m}_{\lambda_\ell})$ cannot be tuned. This should however be made possible as recommended by physically-inspired models such as model (22). In addition, kernels κ_{nlin} with embedded linear component such as the inhomogeneous polynomial kernel (25) introduces a bias into the estimation of $\boldsymbol{\alpha}$, unless correctly estimated and removed. Another difficulty is that the model (12) cannot captures the dynamic of the mixture, which requires that $\boldsymbol{r}$ or the $\boldsymbol{m}_{\lambda_\ell}$'s be locally normalized. This unlikely situation occurs, e.g., if a library of reflectance signatures is used for the unmixing process. In order to address problems such as the above, we shall now consider Gram matrices of the form

$$\boldsymbol{K}_u = u\boldsymbol{M}\boldsymbol{M}^\top + (1-u)\boldsymbol{K}_{\text{nlin}} \quad (28)$$

with u in $[0, 1]$ in order to ensure positiveness of $\boldsymbol{K}_u$. Learning both the parameter u and the mixing coefficients $\boldsymbol{\beta}$ in a single optimization problem is known as the multiple kernel learning problem. See [41] and references therein. The rest of this section is devoted to the formulation and resolution of a convex optimization problem.

In order to tune the balance between $\psi_{\text{lin}}(\boldsymbol{m}_{\lambda_\ell}) = \boldsymbol{\alpha}^\top\boldsymbol{m}_{\lambda_\ell}$ and $\psi_{\text{nlin}}(\boldsymbol{m}_{\lambda_\ell})$, it might seem tempting to substitute matrix $\boldsymbol{K}$ with matrix $\boldsymbol{K}_u$ in the dual problem (19). Unfortunately, a primal problem must be first formulated in order to identify, in the spirit of equation (18), explicit expressions for ψ_{lin} and ψ_{nlin}. We propose to conduct hyperspectral data unmixing by solving the following primal problem

$$\begin{aligned}\psi^*, u^* = \arg\min_{\psi,u} \frac{1}{2}\left(\frac{1}{u}\|\psi_{\text{lin}}\|^2_{\mathcal{H}_{\text{lin}}} + \frac{1}{1-u}\|\psi_{\text{nlin}}\|^2_{\mathcal{H}_{\text{nlin}}}\right)\\ + \frac{1}{2\mu}\sum_{\ell=1}^L e_\ell^2\\ \text{subject to} \quad e_\ell = r_\ell - \psi(\boldsymbol{m}_{\lambda_\ell}) \quad \text{and} \quad 0 \leq u \leq 1\\ \text{where} \quad \psi = \psi_{\text{lin}} + \psi_{\text{nlin}}\end{aligned} \quad (29)$$

where u allows to adjust the balance between ψ_{lin} and ψ_{nlin} via their norms. The spaces $\mathcal{H}'_{\text{lin}}$ and $\mathcal{H}'_{\text{nlin}}$ are RKHS of the general form

$$\mathcal{H}' = \left\{\psi \in \mathcal{H} : \frac{\|\psi\|_{\mathcal{H}}}{u} < \infty\right\} \quad (30)$$

with the convention $\frac{x}{0} = 0$ if $x = 0$, and ∞ otherwise. This implies that, if $u = 0$, then ψ belongs to space $\mathcal{H}'$ if and only if $\psi = 0$. By continuity consideration via this convention, it can be shown that problem (29) is a convex optimization problem by virtue of the convexity of the so-called perspective function $f(u, \psi) = \|\psi\|^2_{\mathcal{H}'}/u$ over $\mathbb{R}_+ \times \mathcal{H}'$. This has been shown in [42, Chapter 3] in the finite-dimensional case, and extended in [41] to the infinite-dimensional case. This allows to formulate the two-stage optimization procedure, with respect to ψ and u successively, in order to solve problem (29).

$$\min_u J(u) \qquad \text{subject to} \qquad 0 \leq u \leq 1 \quad (31)$$

where $J(u)$ is defined by (32) (next page).

The connection between (29) and this problem is as follows. We always have [42, p. 133]

$$\min_{u,\psi} F(u, \psi) = \min_u J(u) \quad (33)$$

where $J(u) = \min_\psi F(u, \psi)$, subject to all the constraints over u and ψ defined in (31)-(32). In addition, as proven in

$$J(u) = \begin{cases} \min\limits_{\psi} F(u,\psi) = \dfrac{1}{2}\left(\dfrac{1}{u}\|\psi_{\text{lin}}\|^2_{\mathcal{H}_{\text{lin}}} + \dfrac{1}{1-u}\|\psi_{\text{nlin}}\|^2_{\mathcal{H}_{\text{nlin}}}\right) + \dfrac{1}{2\mu}\displaystyle\sum_{\ell=1}^{L} e_\ell^2 \\ \qquad\qquad \text{subject to} \quad e_\ell = r_\ell - \psi(\boldsymbol{m}_{\lambda_\ell}) \quad \text{with} \quad \psi = \psi_{\text{lin}} + \psi_{\text{nlin}} \\ \qquad\qquad \text{and} \quad \psi_{\text{lin}}(\boldsymbol{m}_{\lambda_\ell}) = \boldsymbol{h}^\top \boldsymbol{m}_{\lambda_\ell} \quad \text{with} \quad \boldsymbol{h} \succeq \boldsymbol{0} \end{cases} \tag{32}$$

textbooks [42, p. 87], because F is convex in (u,ψ) subject to convex constraints over ψ, it turns out that $J(u)$ is convex in u and, as a consequence, that the constrained optimization problem (31) is convex.

Compared to the preliminary algorithm described in Section III-A, it is important to note that the sum-to-one constraint $\mathbf{1}^\top \boldsymbol{h} = 1$ has been given up. We have adopted an alternative strategy, which consists of relaxing this constraint, and normalizing $\boldsymbol{h}$ afterward in order to be correctly interpreted as a vector $\boldsymbol{\alpha}$ of fractional abundances. The main reasons are as follows. The weight vector $\boldsymbol{h}$ is optimally scaled by the factor u in order to achieve the best balance between ψ_{lin} and ψ_{nlin}. As a consequence, there is no significant advantage of satisfying the sum-to-one constraint $\mathbf{1}^\top \boldsymbol{h} = 1$ along the optimization process, which is consequently simplified. On the contrary, we clearly observed that relaxing the ℓ_1-norm $\mathbf{1}^\top \boldsymbol{h}$ of the weight vector $\boldsymbol{h}$ acts as an additional degree of freedom to adjust the balance between ψ_{lin} and ψ_{nlin}, which would otherwise be limited to the factors u and $1-u$. This has been confirmed by experiments, which have revealed a significant improvement in performance.

By the strong duality property, we shall now derive a dual problem that has the same solution $J(u) = F(u,\psi^*)$ as the primal problem (32). Let us introduce the Lagrange multipliers β_ℓ and γ_r. The Lagrange function associated with the problem (32) can be written as

$$\begin{aligned} G =& \frac{1}{2}\left(\frac{1}{u}\|\boldsymbol{h}\|^2 + \frac{1}{1-u}\|\psi_{\text{nlin}}\|^2_{\mathcal{H}_{\text{nlin}}}\right) + \frac{1}{2\mu}\sum_{\ell=1}^{L} e_\ell^2 \\ & - \sum_{\ell=1}^{L} \beta_\ell \left(e_\ell - r_\ell + \psi(\boldsymbol{m}_{\lambda_\ell})\right) - \sum_{r=1}^{R} \gamma_r\, h_r \end{aligned} \tag{34}$$

with $\gamma_r \geq 0$, where we have used that $\|\psi_{\text{lin}}\|^2_{\mathcal{H}_{\text{lin}}} = \|\boldsymbol{h}\|^2$. The conditions for optimality of G with respect to the primal variables are given by

$$\begin{cases} \boldsymbol{h}^* = u\left(\sum_{\ell=1}^{L} \beta_\ell^*\, \boldsymbol{m}_{\lambda_\ell} + \boldsymbol{\gamma}^*\right) \\ \psi^*_{\text{nlin}} = (1-u)\sum_{\ell=1}^{L} \beta_\ell^*\, \kappa_{\text{nlin}}(\cdot, \boldsymbol{m}_{\lambda_\ell}) \\ e_\ell^* = \mu\, \beta_\ell^* \end{cases} \tag{35}$$

By substituting (35) into (34), we get the dual problem (36) (next page).

Pixel reconstruction can be performed using

$$\boldsymbol{r}^* = [\psi^*(\boldsymbol{m}_{\lambda_1}), \ldots, \psi^*(\boldsymbol{m}_{\lambda_L})]^\top$$

with $\psi^*(\boldsymbol{m}_{\lambda_\ell}) = \boldsymbol{m}_{\lambda_\ell}^\top \boldsymbol{h}^* + \psi^*_{\text{nlin}}(\boldsymbol{m}_{\lambda_\ell})$ defined in equation (35). Finally, the estimated abundance vector is given by

$$\boldsymbol{\alpha}^* = \frac{\boldsymbol{M}^\top \boldsymbol{\beta}^* + \boldsymbol{\gamma}^*}{\mathbf{1}^\top (\boldsymbol{M}^\top \boldsymbol{\beta}^* + \boldsymbol{\gamma}^*)} \tag{37}$$

Let us briefly address the differentiability issue of the problem (31)-(36). The existence and computation of the derivatives of supremum functions such as $J(u)$ have been largely discussed in the literature. As pointed out in [41], [43], the differentiability of J at any point u_0 is ensured by the unicity of the corresponding minimizer $(\boldsymbol{\beta}_0^*, \boldsymbol{\gamma}_0^*)$, and by the differentiability of the cost function $F(u,\psi)$ in (32). The derivative of J at u_0 can be calculated as if the minimizer $(\boldsymbol{\beta}_0^*, \boldsymbol{\gamma}_0^*)$ was independent of u_0, namely, $\frac{dJ(u)}{du}\big|_{u=u_0} = \frac{\partial G'(u,\boldsymbol{\beta}_0^*,\boldsymbol{\gamma}_0^*)}{\partial u}\big|_{u=u_0}$. This yields

$$\left.\frac{dJ(u)}{du}\right|_{u=u_0} = -\frac{1}{2}\left(\|\boldsymbol{M}^\top \boldsymbol{\beta}_0^* + \boldsymbol{\gamma}_0^*\|^2 - \boldsymbol{\beta}_0^{*\top} \boldsymbol{K}_{\text{nlin}}\, \boldsymbol{\beta}_0^*\right) \tag{38}$$

Table I summarizes the proposed algorithm. Note that (31) is a very small-size problem. Indeed, it involves a one-dimension optimization variable and can thus be solved with an ad-hoc procedure. Using a gradient projection method, e.g., based on Armijo rule along the feasible direction, makes practical sense in this case [39, Chapter 2]. See also [44]. Moreover, both problems can benefit of warm-starting between successive solutions to speed-up the optimization procedure. The algorithm can be stopped based on conditions for optimality in convex optimization framework. In particular, the KKT conditions and the duality gap should be equal to zero, within a numerical error tolerance specified by the user. The variation of the cost $J(u)$ between two successive iterations should also be considered as a potential stopping criterion.

Before testing our algorithms, and comparing their performance with state-of-the-art approaches, we shall now explain how they differ from existing kernel-based techniques for hyperspectral data processing.

D. Comparison with existing kernel-based methods in hyperspectral imagery

Some kernel-based methods have already been proposed to process hyperspectral images, with application to classification, supervised or unsupervised unmixing, etc. By taking advantage of capturing nonlinear data dependences, some of them have been shown to achieve better performance than their linear counterpart. Let us now briefly discuss the main difference between our kernel-based model and those presently existing. The central idea underlying most of state-of-the-art methods is to nonlinearly transform hyperspectral pixel-vectors prior to applying a linear algorithm, simply by replacing inner products with kernels in the cost function. This basic principle is fully justified in detection/classification problems because a proper nonlinear distortion of spectral signatures can increase the detectability/separability of materials. Within

$$J(u) = \begin{cases} \max\limits_{\boldsymbol{\beta},\boldsymbol{\gamma}} G'(u,\boldsymbol{\beta},\boldsymbol{\gamma}) = -\frac{1}{2}\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\gamma}\end{pmatrix}^\top \left(\begin{array}{c|c} \boldsymbol{K}_u + \mu\boldsymbol{I} & u\boldsymbol{M} \\ \hline u\boldsymbol{M}^\top & u\boldsymbol{I}\end{array}\right)\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\gamma}\end{pmatrix} + \begin{pmatrix}\boldsymbol{r}\\ \boldsymbol{0}\end{pmatrix}^\top \begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\gamma}\end{pmatrix} \\ \text{subject to} \quad \boldsymbol{\gamma} \succeq \boldsymbol{0} \\ \text{with} \quad \boldsymbol{K}_u = u\boldsymbol{M}\boldsymbol{M}^\top + (1-u)\boldsymbol{K}_{\text{nlin}} \end{cases} \tag{36}$$

TABLE I
NONLINEAR UNMIXING BY MULTIPLE KERNEL LEARNING: THE SK-HYPE ALGORITHM

Initialization
Choose the kernel κ_{nlin} and the regularization constant $\mu \geq 0$.
Calculate the kernel matrix $\boldsymbol{K}_{\text{nlin}}$.
Repeat
Calculate $J(u)$ in (36) by using a generic QP solver
Solve the optimization problem
$\min_u J(u)$ subject to $0 \leq u \leq 1$
by performing one iteration of the gradient projection algorithm, using (38)
until stopping criterion is satisfied
Estimate the abundances by (37)

the context of hyperspectral unmixing, this leads to consider mixtures of the form

$$\Phi(\boldsymbol{r}) = [\Phi(\boldsymbol{m}_1), \Phi(\boldsymbol{m}_2), \ldots, \Phi(\boldsymbol{m}_R)]\,\boldsymbol{\alpha} + \boldsymbol{n}. \tag{39}$$

This model is inherent in the KFCLS algorithm [30], [31], which optimizes the following mean-square error criterion where all the inner products have been replaced by kernels

$$J(\boldsymbol{\alpha}) = \boldsymbol{\alpha}^\top \boldsymbol{K}_{\boldsymbol{m}}\,\boldsymbol{\alpha} - 2\,\boldsymbol{\alpha}^\top \boldsymbol{\kappa}_{rm} + \kappa(\boldsymbol{r},\boldsymbol{r}), \tag{40}$$

where $\boldsymbol{K}_{\boldsymbol{m}}$ is the Gram matrix with (i,j)-th entry $\kappa(\boldsymbol{m}_i,\boldsymbol{m}_j)$, and $\boldsymbol{\kappa}_{rm}$ is a vector with i-th entry $\kappa(\boldsymbol{r},\boldsymbol{m}_i)$. Unfortunately, even though model (39) allows distortions of spectral signatures, it does not explicitly include nonlinear interactions of the endmember spectra. The analysis in Section III-B has shown strong connections between our kernel-based model and well-characterized models, e.g., the generalized bilinear mixture model. The experimental comparison on simulated and real data reported in the next section confirms this view.

IV. EXPERIMENTAL RESULTS

We shall now conduct some simulations to validate the proposed unmixing algorithms, and to compare them with state-of-the-art methods, using both synthetic and real images.

A. Experiments on synthetic images

Let us first report some experimental results on synthetic images, generated by linear and nonlinear mixing of several endmember signatures. The materials we have considered are alunite, calcite, epidote, kaolinite, buddingtonite, almandine, jarosite and lepidolite. They were selected from the ENVI software library. These spectra consist of 420 contiguous bands, covering wavelengths ranging from 0.3951 to 2.56 micrometers.

In the first scene, only three materials were selected to generate images: epidote, kaolinite, buddingtonite. In the second scene, five materials were used: alunite, calcite, epidote, kaolinite, buddingtonite. In the third scene, the eight materials were used. For each scene, three 50-by-50 hyperspectral images were generated with different mixture models, each providing $N = 2500$ pixels for evaluating and comparing the performance of several algorithms. These three models were the linear model (1), the bilinear mixture model defined as

$$\boldsymbol{r} = \boldsymbol{M}\boldsymbol{\alpha} + \sum_{i=1}^{R-1}\sum_{j=i+1}^{R} \alpha_i\,\alpha_j\,(\boldsymbol{m}_i \otimes \boldsymbol{m}_j) + \boldsymbol{n}, \tag{41}$$

and a post-nonlinear mixing model (PNMM) [45] defined by

$$\boldsymbol{r} = (\boldsymbol{M}\boldsymbol{\alpha})^\xi + \boldsymbol{n} \tag{42}$$

where $(\,\cdot\,)^\xi$ denotes the exponential value ξ applied to each entry of the input vector. Parameter ξ was set to 0.7. The abundance vectors $\boldsymbol{\alpha}_n$, with $n = 1,\ldots,2500$, were uniformly generated in the simplex defined by non-negative and sum-to-one constraints. Finally, all these images were corrupted with an additive white Gaussian noise $\boldsymbol{n}$ with two levels of SNR, 30 dB and of 15 dB.

The following algorithms were considered

- **The so-called Fully Constrained Least Square method (FCLS)**, [11]: This technique was derived based on linear mixture model. It provides the optimal solution in the least-mean-square sense, subject to non-negativity and sum-to-one constraints. A relaxation parameter ν has to be tuned to specify a compromise between the residual error and the sum-to-one constraint.
- **The extended endmember-matrix method (ExtM)**, [25]: This method consists of extending the endmember matrix $\boldsymbol{M}$ artificially with cross-spectra of pure materials in order to model light scatter effects. In the experiments, all the second-order cross terms $\boldsymbol{m}_i \otimes \boldsymbol{m}_j$ were inserted so that it would correspond to the generalized bilinear model. This approach also has a relaxation parameter ν for the sum-to-one constraint.

- **The so-called Kernel Fully Constrained Least Square method (KFCLS)**, [30]: This is a kernel method, directly derived from FCLS, in which all the inner products are replaced by kernel functions. As for all the other kernel-based algorithms considered in this paper, the Gaussian kernel was used for simulations. This algorithm has two parameters, the bandwidth σ of the Gaussian kernel, and a relaxation parameter ν for the sum-to-one constraint.
- **The Bayesian algorithm derived for generalized bilinear model (BilBay)**, [20]: This method is based on appropriate prior distributions for the unknown parameters, which must satisfy the non-negativity and sum-to-one constraints, and then derives joint posterior distribution of these parameters. A Metropolis-within-Gibbs algorithm is used to estimate the unknown model parameters. The MMSE estimates of the abundances were computed by averaging the 2500 generated samples obtained after 500 burn-in iterations.
- **The first algorithm proposed in this paper (K-Hype)**: This is the preliminary algorithm described in Section III-A. The Gaussian kernel (G) with bandwidth σ, and the polynomial kernel (P) defined by (27) were considered. The Matlab optimization function Quadprog was used to solve the QP problem.
- **The second algorithm proposed in this paper (SK-Hype)**: This is the main algorithm described in Section III-C and Table I. As for K-Hype, the Gaussian kernel and the polynomial kernel were considered. In order to simplify the experiments, the weight vector $\boldsymbol{u}$ was constrained to be of the form $\boldsymbol{u} = [u, 1-u]^\top$ with $0 \leq u \leq 1$. Obviously, this allows to reduce the number of variables but does not change the structure of the algorithm. The variable u was initially set to$\frac{1}{2}$. A gradient projection method, based on the Armijo rule to compute the optimal step size along the feasible direction, was used to determined u. The algorithm was stopped when the relative variation of u between two successive iterations became less than $\zeta_{\max} = 10^{-3}$, or the maximum number of iterations $\text{It}_{\max} = 10$ was reached. The Matlab optimization function Quadprog was used to solve the QP problem.

The root mean square error defined by

$$\text{RMSE} = \sqrt{\frac{1}{NR}\sum_{n=1}^{N}\|\boldsymbol{\alpha}_n - \boldsymbol{\alpha}_n^*\|^2} \tag{43}$$

was used to compare these six algorithms. In order to tune their parameters, preliminary runs were performed on 100 independent test pixels for each experiment. The bandwidth σ of the Gaussian kernel in the algorithms ExtM, K-Hype and SK-Hype was varied within $\{1, \ldots, 3\}$ with increment of $\frac{1}{2}$. The parameter μ of K-Hype and SK-Hype algorithms was varied within $\{1, 10^{-1}, 10^{-2}, 5 \cdot 10^{-3}\}$. The parameter ν in algorithms FCLS, ExtM, KFCLS was chosen within $\{1, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$. All the parameters used in the experiments are reported in the Appendix.

Results for Scene 1 to Scene 3 unmixing, with three, five and eight endmember materials, are reported in Table II, Table III and Table IV respectively. Because the FCLS method was initially derived for the linear mixing model, it achieves a very low RMSE for linearly-mixed images, and produces a relatively large RMSE with nonlinearly-mixed images. With second-order cross terms that extend the endmember matrix $\boldsymbol{M}$, the ExtM algorithm notably reduces the RMSE when dealing with bilinearly-mixed images when compared with FCLS. However, it marginally improves the performance in PNMM image ummixing. BilBay algorithm was derived for the bilinear mixing model, and thus achieves very good performance with bilinearly-mixed images. Nevertheless, the performance of BilBay clearly degrades when dealing with a nonlinear mixing model for which it was not originally designed. KFCLS with Gaussian kernel performs worse than FCLS, even with nonlinearly-mixed images as it does not clearly investigate nonlinear interactions between materials.

For the less noisy scenes (30 dB), our algorithms K-Hype and SK-Hype exhibit significantly reduced RMSE when dealing with nonlinearly-mixed images. In the case of the bilinear model, K-Hype and SK-Hype achieve very good performance compared to the other algorithms. Indeed, they are the best performers except in a few cases. In the case of the PNMM model, they outperform all the other algorithms, and it can be observed that SK-Hype outperforms K-Hype in several scenarios. For the noisiest scenes (15 dB), although the increase in the noise level significantly degrades the performance of all the algorithms, K-Hype and SK-Hype still maintain an advantage. Last but not least, the margin of performance over the other approaches becomes larger as the number of endmembers increases.

To give a more meaningful comparison of the performance of these algorithms, one-tailed Welch's t-tests with significance level 0.05 were used to test the hypothesis

$$\mathcal{H}_0: \quad \text{RMSE}_{\text{proposed}} < \text{RMSE}_{\text{literature}}$$

where $\text{RMSE}_{\text{proposed}}$ denotes the RMSE of the K-Hype and SK-Hype algorithms, with Gaussian and polynomial kernels, and $\text{RMSE}_{\text{literature}}$ is the RMSE of the algorithms of the literature selected in this paper. Due to limited space, only the results for Scene 2 and the SNR level 30 dB are reported here, in Table V to VII. The letter $\mathcal{A}$ means that the hypothesis $\mathcal{H}_0$ is accepted. Without ambiguity, these results confirm the advantage of our algorithms.

TABLE V
WELSH'S t-TESTS FOR SCENE 2 WITH SNR = 30 DB (LINEAR MODEL)

	FCLS	ExtM	KFCLS	BilBay
K-Hype (G)				$\mathcal{A}$
K-Hype (P)				$\mathcal{A}$
SK-Hype (G)				$\mathcal{A}$
SK-Hype (P)				$\mathcal{A}$

The computational time of these algorithms mainly depends on the constrained optimization problem to be solved. FCLS and KFLCS minimize a quadratic cost function of dimension$R$, under inequality constraints of the same dimension. ExtM solves a similar problem but with an increased

TABLE II
SCENE 1 (THREE MATERIALS): RMSE COMPARISON

	SNR = 30 dB			SNR = 15 dB		
	linear	bilinear	PNMM	linear	bilinear	PNMM
FCLS	$0.0037\pm2\times10^{-5}$	0.0758±0.0019	0.0604±0.0017	0.0212±0.0005	0.0960±0.0060	0.0886±0.0063
ExtM	0.0079±0.0001	0.0312±0.0013	0.0601±0.0016	0.0404±0.0031	0.0991±0.096	0.0869±0.0066
KFCLS	$0.0054\pm3\times10^{-5}$	0.2711±0.0516	0.2371±0.0197	0.0296±0.0009	0.2694±0.0498	0.2372±0.0235
BilBay	0.0384±0.0013	0.0285±0.0006	0.1158±0.0058	0.1135±0.0098	0.1059±0.0085	0.1191±0.0091
K-Hype (G)	0.0208±0.0004	0.0349±0.0013	0.0446±0.0020	0.0562±0.0041	0.0611±0.0048	0.0786±0.0067
K-Hype (P)	0.0346±0.0011	0.0281±0.0011	0.0569±0.0031	0.0589±0.0041	0.0628±0.0053	0.0794±0.0066
SK-Hype (G)	0.0104±0.0001	0.0315±0.0012	0.0230±0.0007	0.0562±0.0044	0.0598±0.0048	0.0757±0.0073
SK-Hype (P)	0.0106±0.0002	0.0310±0.0011	0.0245±0.0007	0.0561±0.0043	0.0602±0.0048	0.0742±0.0075

TABLE III
SCENE 2 (FIVE MATERIALS): RMSE COMPARISON

	SNR = 30 dB			SNR = 15 dB		
	linear	bilinear	PNMM	linear	bilinear	PNMM
FCLS	0.0134±0.0002	0.1137±0.0032	0.1428±0.0039	0.0657±0.0047	0.1444±0.0116	0.1611±0.0134
ExtM	0.0157±0.0003	0.0575±0.0024	0.1427±0.0040	0.0761±0.0060	0.1207±0.0160	0.1678±0.0139
KFCLS	0.0200±0.0004	0.2051±0.0148	0.1955±0.0115	0.0890±0.0080	0.1884±0.0113	0.1572±0.0114
BilBay	0.0585 ±0.0017	0.0441±0.0010	0.1741±0.0082	0.1465±0.0109	0.1007±0.0063	0.1609±0.0124
K-Hype (G)	0.0231±0.0004	0.0307±0.0008	0.0398±0.0012	0.1076±0.0093	0.0748±0.0046	0.0823±0.0053
K-Hype (P)	0.0218±0.0004	0.0465±0.0012	0.0386±0.0011	0.0738±0.0043	0.0847±0.0052	0.0828±0.0054
SK-Hype (G)	0.0196±0.0004	0.0288±0.0007	0.0346±0.0010	0.0675±0.0040	0.0778±0.0043	0.0942±0.0065
SK-Hype (P)	0.0195±0.0004	0.0349±0.0008	0.0346±0.0010	0.0673±0.0040	0.0830±0.0046	0.0965±0.0071

TABLE IV
SCENE 3 (EIGHT MATERIALS): RMSE COMPARISON

	SNR = 30 dB			SNR = 15 dB		
	linear	bilinear	PNMM	linear	bilinear	PNMM
FCLS	0.0148±0.0002	0.0930±0.0024	0.1079±0.0018	0.0652±0.0031	0.1177±0.0068	0.1252±0.0065
ExtM	0.0173±0.0003	0.0560±0.0017	0.1126±0.0019	0.0743±0.0038	0.1066±0.0062	0.1322±0.0063
KFCLS	0.0216±0.0004	0.1431±0.0059	0.1274±0.0039	0.0647±0.0032	0.1270±0.0038	0.2250±0.0220
BilBay	0.0448±0.0007	0.0369±0.0004	0.1159±0.0029	0.0745±0.0020	0.0792±0.0026	0.1040±0.0430
K-Hype (G)	0.0203±0.0003	0.0202±0.0003	0.0300±0.0006	0.0562±0.0020	0.0548±0.0018	0.0642±0.0024
K-Hype (P)	0.0195±0.0003	0.0330±0.0006	0.0297±0.0006	0.0585±0.0021	0.0646±0.0024	0.0657±0.0026
SK-Hype (G)	0.0185±0.0003	0.0221±0.0003	0.0291±0.0006	0.0561±0.0019	0.0573±0.0020	0.0696±0.0027
SK-Hype (P)	0.0184±0.0002	0.0247±0.0004	0.0313±0.0007	0.0571±0.0021	0.0620±0.0021	0.0736±0.0031

TABLE VI
WELSH'S t-TESTS FOR SCENE 2 WITH SNR = 30 DB (BILINEAR MODEL)

	FCLS	ExtM	KFCLS	BilBay
K-Hype (G)	$\mathcal{A}$	$\mathcal{A}$	$\mathcal{A}$	$\mathcal{A}$
K-Hype (P)	$\mathcal{A}$	$\mathcal{A}$	$\mathcal{A}$	
SK-Hype (G)	$\mathcal{A}$	$\mathcal{A}$	$\mathcal{A}$	$\mathcal{A}$
SK-Hype (P)	$\mathcal{A}$	$\mathcal{A}$	$\mathcal{A}$	$\mathcal{A}$

TABLE VII
WELSH'S t-TESTS FOR SCENE 2 WITH SNR = 30 DB (PNMM)

	FCLS	ExtM	KFCLS	BilBay
K-Hype (G)	$\mathcal{A}$	$\mathcal{A}$	$\mathcal{A}$	$\mathcal{A}$
K-Hype (P)	$\mathcal{A}$	$\mathcal{A}$	$\mathcal{A}$	$\mathcal{A}$
SK-Hype (G)	$\mathcal{A}$	$\mathcal{A}$	$\mathcal{A}$	$\mathcal{A}$
SK-Hype (P)	$\mathcal{A}$	$\mathcal{A}$	$\mathcal{A}$	$\mathcal{A}$

dimension due to the cross-spectra that are artificially inserted. In the case where only the second-order cross spectra are added, the dimension of the optimization problem is $R+\binom{R}{2}$ with $R = 3$, 5 and 8 in this study. BilBay has to generate numerous samples to estimate the model parameters, and suffers from the large computational cost of this sampling strategy. K-Hype solves a quadratic programming problem of dimension $L + R + 1$. It is interesting to note that the computational cost is independent of the complexity of the unmixing model. A sparsification strategy as described in [46] should be advantageously used to greatly reduce the computational complexity with negligible effect on the quality of the results. SK-Hype has similar advantages as K-Hype except that the alternating optimization scheme requires more time. The average computational times per pixel of all these algorithms are listed in Table VIII.[2]

B. Experiment with AVIRIS image

This section illustrates the performance of the proposed algorithms, and several other algorithms, when applied to real hyperspectral data. The scene that was used for our experiment is the well-known image captured on the Cuprite mining district (NV, USA) by AVIRIS. A sub-image of 250×191 pixels was chosen to evaluate the algorithms. This area of interest has $L = 188$ spectral bands. The number of endmembers was first estimated via the virtual dimensionality, and R was accordingly set to 12 [4]. VCA algorithm was then used to

[2]Using Matlab R2008a on a iMac with 3.06GHz Intel Core 2 Duo and 4 Go Memory

TABLE VIII
AVERAGED COMPUTATIONAL TIME PER PIXEL (IN SECONDS)

	$L = 420$			$L = 210$		
	$R = 3$	$R = 5$	$R = 8$	$R = 3$	$R = 5$	$R = 8$
FCLS	7.6×10^{-4}	1.2×10^{-3}	2.4×10^{-3}	6.4×10^{-4}	1.1×10^{-3}	2.1×10^{-3}
ExtM	1.6×10^{-3}	5.5×10^{-3}	0.020	1.4×10^{-3}	4.0×10^{-3}	0.014
KFCLS	6.0×10^{-4}	1.3×10^{-3}	2.5×10^{-3}	5.6×10^{-4}	8.4×10^{-4}	1×10^{-3}
BilBay	6.6	15	40	6.4	14	33
K-Hype	0.17	0.29	0.47	0.038	0.065	0.11
SK-Hype	4.3	7.5	13.4	0.92	1.7	3.1

extract the endmembers. Both our algorithms were compared with all the state-of-the-art algorithms considered previously. After preliminary experiments, the regularization parameters of FCLS and ExtM algorithms were set to $\nu = 0.01$. K-Hype algorithm and SK-Hype algorithm were run with the polynomial kernel (27), and the Gaussian kernel. The bandwidth of the Gaussian kernel was set to $\sigma = 2$. The regularization parameter μ was fixed to $2 \cdot 10^{-3}$. To evaluate the performance, the averaged spectral angle between original $\boldsymbol{r}$ and reconstructed $\boldsymbol{r}^*$ pixel vectors was used

$$\Theta = \frac{1}{N} \sum_{n=1}^{N} \theta(\boldsymbol{r}_n, \boldsymbol{r}_n^*)$$

where N is the number of processed pixels and $\theta(\boldsymbol{r}, \boldsymbol{r}^*) = \cos^{-1}\left(\frac{\langle \boldsymbol{r}, \boldsymbol{r}^* \rangle}{\|\boldsymbol{r}\| \|\boldsymbol{r}^*\|}\right)$. It is important to note that the quality of reconstruction, estimated by the averaged spectral angle or mean-square error for instance, is not necessarily in proportion to the the quality of unmixing, especially for real images where the nonlinear mixing mechanism can be complex. In particular, more complicated model may better fit the data. Parameter Θ is only reported here as complementary information. The averaged spectral angle of each approach is reported in Table IX. Note that KFCLS was not considered in these tests as there is no possible direct reconstruction of pixels. Clearly, our algorithms have much lower reconstruction errors than the other approaches. Six typical estimated abundance maps out of twelve available are shown in Figure 1. It can be observed that the estimated locations of the different materials are quite similar for the four methods, except the US Highway 95 in the last column which is much more accurately depicted by our methods. Finally, the distributions of reconstruction errors $\|\boldsymbol{r}_n - \boldsymbol{r}_n^*\|/L$ associated to these methods are shown in Figure2.

TABLE IX
SPECTRAL ANGLES COMPARISON

Algorithms	Θ
FCLS	0.0136
ExtM	0.0123
BilBay	0.0182
K-Hype (G)	0.0070
K-Hype (P)	0.0098
SK-Hype (G)	0.0078
SK-Hype (P)	0.0104

V. CONCLUSION

Spectral unmixing is an important issue to analyze remotely sensed hyperspectral data. This involves the decomposition of each mixed pixel into its pure endmember spectra, and the estimation of the abundance value for each endmember. To be physically interpretable, the abundances are often required to be nonnegative, and their sum must be equal to one. Although the linear mixture model has many practical advantages, there are many situations in which it may not be appropriate. In this paper, we formulated a new kernel-based paradigm that relies on the assumption that the mixing mechanism can be described by a linear mixture of endmember spectra, with additive nonlinear fluctuations defined in a RKHS. This family of models has a clear physical interpretation, and allows to take complex interactions of endmembers into account. Two kernel-based algorithms for estimating the abundances were proposed. The second one, based on the concept of multiple kernel learning, is the natural generalization of the first one as it allows to automatically adapt the balance between linear spectral interactions and nonlinear ones. Future work will include studying the feasibility and constraints of designing physically meaningful kernels, possibly based on manifold learning as in [47], [48]. We will also focus our attention on adaptive kernel-based algorithms, in the spirit of [46], to unmix neighboring pixel-vectors sequentially and thus speed up processing.

REFERENCES

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

[2] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 354– 379, 2012.

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

[4] 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.

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

[6] A. Plaza, G. Martín, J. Plaza, M. Zortea, and S. Sánchez, "Recent developments in endmember extraction and spectral unmixing," in *Optical Remote Sensing: Advances in Signal Processing and Exploitation Techniques*, S. Prasad, L. Bruce, and J. Chanussot, Eds. 2011, pp. 235–267, Springer.

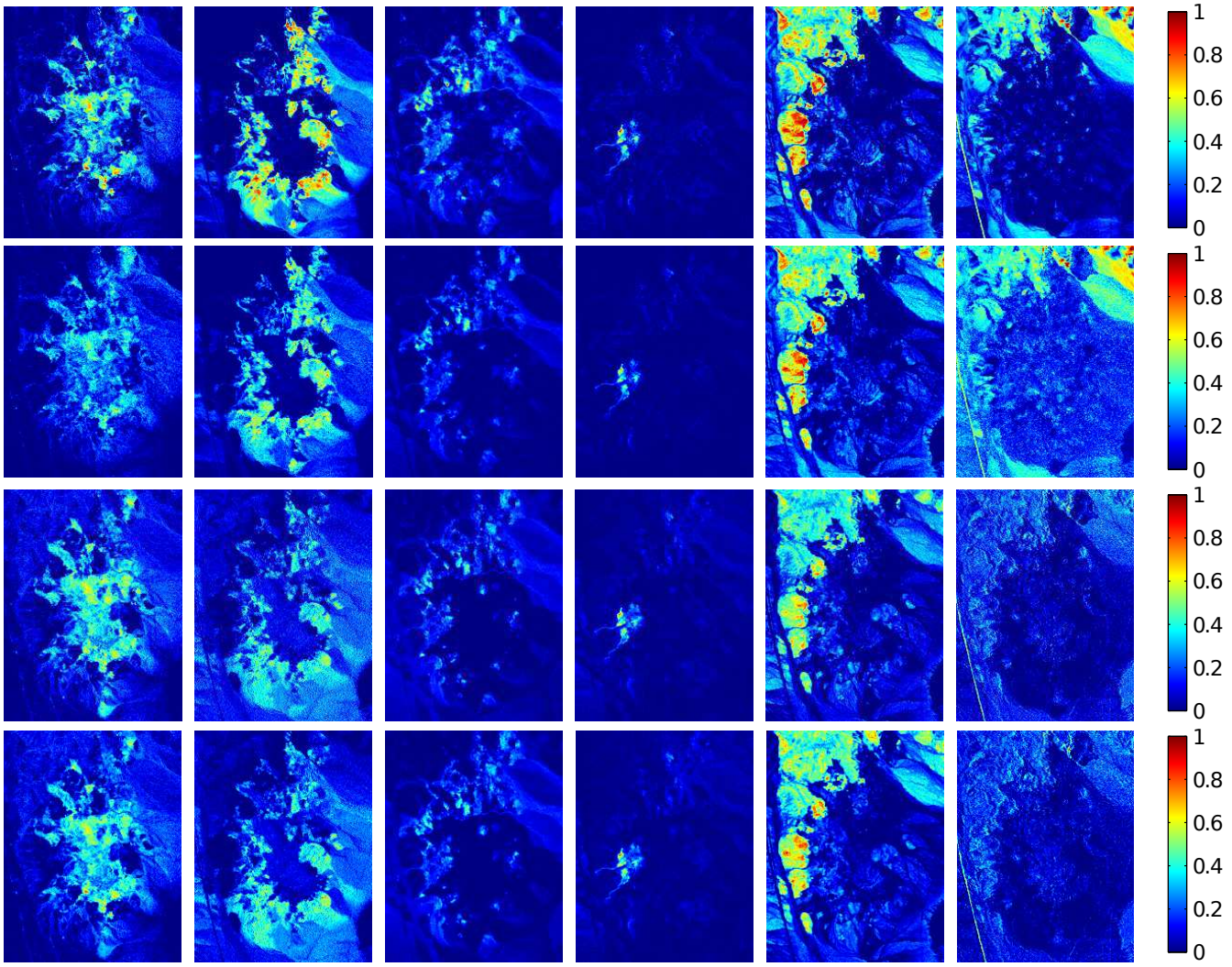

Fig. 1. Abundances maps of selected materials. From top to bottom: FCLS, BilBay, K-Hype (G), SK-Hype (G). From left to right: calcedony, alunite, kaolinite, buddingtonite, sphene, US highway 95.

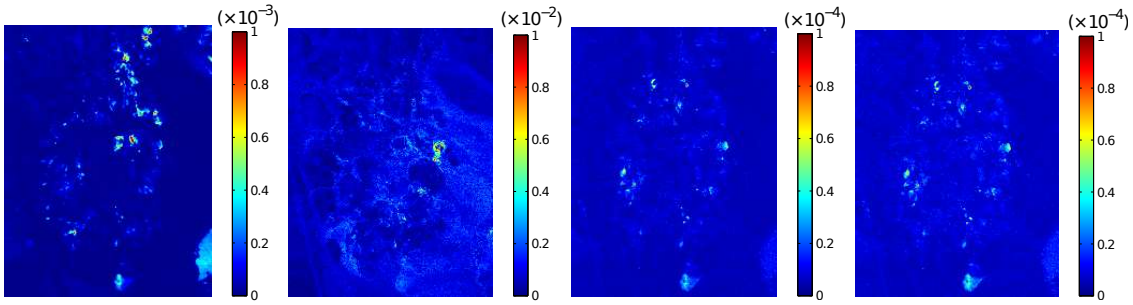

Fig. 2. Maps of reconstruction error. From left to right: FCLS, BilBay, K-Hype (G), SK-Hype (G).

[7] 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.

[8] J. Li and J. M. Bioucas-Dias, "Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data," in *Proc. IEEE IGARSS*, 2008, pp. III–250–III–253.

[9] 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.

[10] 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.

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

[12] S. Moussaoui, H. Hauksdottir, F. Schmidt, C. Jutten, J. Chanussot, D. Brie, S. Douté, and J. A. Benediktsson, "On the decomposition of Mars hyperspectral data by ICA and bayesian positive source separation," *Neurocomputing*, vol. 71, no. 10, pp. 2194–2208, 2008.

[13] 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.
[14] O. Eches, N. Dobigeon, C. Mailhes, and J.-Y. Tourneret, "Bayesian estimation of linear mixtures using the normal compositional model. application to hyperspectral imagery," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1403–1413, 2010.
[15] K. E. Themelis, A. A. Rontogiannis, and K. D. Koutroumbas, "A novel hierarchical bayesian approach for sparse semisupervised hyperspectral unmixing," *IEEE Transactions on Signal Processing*, vol. 60, no. 2, pp. 585–599, 2012.
[16] Z. Guo, T. Wittman, and S. Osher, "L1 unmixing and its application to hyperspectral image enhancement," in *Proc. SPIE Conference on Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XV*, 2009, vol. 7334, pp. 73341M–73341M.
[17] J. M. Bioucas-Dias and A. Plaza, "Hyperspectral unmixing: Geometrical, statistical, and sparse regression-based approaches," in *Proc. SPIE Image and Signal Processing for Remote Sensing XVI*, 2010, vol. 7830, pp. 78300A1–78300A15.
[18] M. D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, pp. 2014–2039, 2010.
[19] 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.
[20] A. Halimi, Y. Altman, 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.
[21] B. Hapke, "Bidirectional reflectance spectroscopy, 1, Theory," *Journal of Geophysical Research*, vol. 86, no. B4, pp. 3039–3054, 1981.
[22] 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.
[23] J. Plaza, P. Martínez, R. Pérez, and A. Plaza, "Nonlinear neural network mixture models for fractional abundance estimation in AVIRIS hyperspectral images," in *Proc. AVIRIS workshop*, Pasadena, CA, 2004.
[24] J. Plaza, A. Plaza, R. Perez, and P. Martinez, "On the use of small training sets for neural network-based characterization of mixed pixels in remotely sensed hyperspectral images," *Pattern Recognition*, vol. 42, no. 11, pp. 3032–3045, 2009.
[25] N. Raksuntorn and Q. Du, "Nonlinear spectral mixture analysis for hyperspectral imagery in an unknown environment," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 4, pp. 836–840, 2010.
[26] J. M. P. Nascimento and J. M. Bioucas-Dias, "Nonlinear mixture model for hyperspectral unmixing," in *Proc. SPIE*, 2009, vol. 7477.
[27] J. M. P. Nascimento and J. M. Bioucas-Dias, "Unmixing hyperspectral intimate mixtures," in *Proc. SPIE*, 2010, vol. 7830.
[28] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 6, pp. 1351–1362, 2005.
[29] K. Heesung and N. M. Nasrabadi, "Kernel orthogonal subspace projection for hyperspectral signal classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 12, pp. 2952–2962, 2005.
[30] J. Broadwater, R. Chellappa, A. Banerjee, and P. Burlina, "Kernel fully constrained least squares abundance estimates," in *Proc. IEEE IGARSS*, 2007, pp. 4041–4044.
[31] J. Broadwater and A. Banerjee, "A comparison of kernel functions for intimate mixture models," in *Proc. IEEE IGARSS*, 2009, pp. 1–4.
[32] X. Wu, X. Li, and L. Zhao, "A kernel spatial complexity-based nonlinear unmixing method of hyperspectral imagery," in *Proc. LSMS/ICSEE*, 2010, pp. 451–458.
[33] B. Schölkopf, J. C. Burges, and A. J. Smola, *Advances in kernel methods*, MIT Press, Cambridge, MA, 1999.
[34] V. N. Vapnik, *The nature of statistical learning theory*, Springer, New York, NY, 1995.
[35] N. Aronszajn, "Theory of reproducing kernels," *Transactions of the American Mathematical Society*, vol. 68, 1950.
[36] J. Mercer, "Functions of positive and negative type and their connection with the theory of integral equations," *Philos. Trans. Roy. Soc. London Ser. A*, vol. 209, pp. 415–446, 1909.
[37] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle, *Least Squares Support Vector Machines*, World Scientific, Singapore, 2002.
[38] D. Haussler, "Convolution kernels on discrete structures," Tech. Rep., Computer Science Department, University of California at Santa Cruz, 1999.
[39] D. Bertsekas, *Nonlinear programming*, Athena Scientific, second edition edition, 1999.
[40] D.G. Luenberger and Y. Ye, *Linear and nonlinear programming*, Springer Verlag, 2008.
[41] A. Rakotomamonjy, F. Bach, S. Canu, and Y. Granvalet, "SimpleMKL," *Journal of Machine Learning Research*, vol. 9, pp. 2491–2521, 2008.
[42] S. Boyd and L. Vandenberghe, *Convex Optimization*, University Press, Cambridge, 2004.
[43] J. F. Bonnans and A. Shapiro, "Optimization problems with perturbations: A guided tour," *SIAM review*, vol. 40, no. 2, pp. 207–227, 1998.
[44] J. Chen, C. Richard, J.-C. M. Bermudez, and P. Honeine, "Nonnegative least-mean-square algorithm," *IEEE Transactions on Signal Processing*, vol. 59, no. 11, pp. 5225 –5235, nov. 2011.
[45] C. Jutten and J. Karhunen, "Advances in nonlinear blind source separation," in *Proc. International Symposium on Independent Component Analysis and Blind Signal Separation (ICA)*, 2003, pp. 245–256.
[46] C. Richard, J.-C. M. Bermudez, and P. Honeine, "Online prediction of time series data with kernels," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1058–1067, March 2009.
[47] 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.
[48] 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*, 2012.

APPENDIX

TABLE X
SIMULATION PARAMETERS CORRESPONDING TO SCENE 1 (CF. TABLE II)

	SNR = 30 dB			SNR = 15 dB		
	linear	bilinear	PNMM	linear	bilinear	PNMM
FCLS	$\nu = 0.001$	$\nu = 1$	$\nu = 1$	$\nu = 0.001$	$\nu = 1$	$\nu = 1$
ExtM	$\nu = 0.001$	$\nu = 0.001$	$\nu = 1$	$\nu = 0.1$	$\nu = 0.001$	$\nu = 1$
KFCLS	$\nu = 0.1$, $\sigma = 3$	$\nu = 0.01$, $\sigma = 3$	$\nu = 0.01$, $\sigma = 5$	$\nu = 0.01$, $\sigma = 5$	$\nu = 0.1$, $\sigma = 1$	$\nu = 0.01$, $\sigma = 3$
BilBay	-	-	-	-	-	-
K-Hype (G)	$\mu = 5 \cdot 10^{-3}$, $\sigma = 3$	$\mu = 10^{-1}$, $\sigma = 3$	$\mu = 5 \cdot 10^{-3}$, $\sigma = 3$	$\mu = 10^{-1}$, $\sigma = 3$	$\mu = 10^{-1}$, $\sigma = 2$	$\mu = 10^{-1}$, $\sigma = 2.5$
K-Hype (P)	$\mu = 5 \cdot 10^{-3}$	$\mu = 10^{-2}$	$\mu = 5 \cdot 10^{-3}$	$\mu = 10^{-1}$	$\mu = 10^{-1}$	$\mu = 10^{-1}$
SK-Hype (G)	$\mu = 10^{-2}$, $\sigma = 2$	$\mu = 10^{-2}$, $\sigma = 2.5$	$\mu = 5 \cdot 10^{-3}$, $\sigma = 3$	$\mu = 10^{-1}$, $\sigma = 1$	$\mu = 10^{-1}$, $\sigma = 1.5$	$\mu = 1$, $\sigma = 2.5$
SK-Hype (P)	$\mu = 5 \cdot 10^{-3}$	$\mu = 10^{-2}$	$\mu = 5 \cdot 10^{-3}$	$\mu = 10^{-1}$	$\mu = 10^{-1}$	$\mu = 10^{-1}$

TABLE XI
SIMULATION PARAMETERS CORRESPONDING TO SCENE 2 (CF. TABLE III)

	SNR = 30 dB			SNR = 15 dB		
	linear	bilinear	PNMM	linear	bilinear	PNMM
FCLS	$\nu = 0.001$	$\nu = 1$	$\nu = 1$	$\nu = 0.001$	$\nu = 1$	$\nu = 1$
ExtM	$\nu = 0.001$	$\nu = 0.1$	$\nu = 1$	$\nu = 0.001$	$\nu = 0.001$	$\nu = 1$
KFCLS	$\nu = 0.1$, $\sigma = 3$	$\nu = 1$, $\sigma = 1.5$	$\nu = 1$, $\sigma = 1$	$\nu = 0.1$, $\sigma = 3$	$\nu = 1$, $\sigma = 3$	$\nu = 1$, $\sigma = 3$
BilBay	-	-	-	-	-	-
K-Hype (G)	$\mu = 10^{-2}$, $\sigma = 3$	$\mu = 10^{-2}$, $\sigma = 1.5$	$\mu = 5 \cdot 10^{-3}$, $\sigma = 3$	$\mu = 10^{-2}$, $\sigma = 2$	$\mu = 1$, $\sigma = 1$	$\mu = 1$, $\sigma = 3$
K-Hype (P)	$\mu = 10^{-2}$	$\mu = 10^{-1}$	$\mu = 5 \cdot 10^{-3}$	$\mu = 1$	$\mu = 1$	$\mu = 1$
SK-Hype (G)	$\mu = 10^{-1}$, $\sigma = 3$	$\mu = 10^{-2}$, $\sigma = 2$	$\mu = 10^{-2}$, $\sigma = 3$	$\mu = 10^{-2}$, $\sigma = 3$	$\mu = 1$, $\sigma = 1$	$\mu = 1$, $\sigma = 1$
SK-Hype (P)	$\mu = 10^{-1}$	$\mu = 5 \cdot 10^{-3}$	$\mu = 5 \cdot 10^{-3}$	$\mu = 1$	$\mu = 1$	$\mu = 1$

TABLE XII
SIMULATION PARAMETERS CORRESPONDING TO SCENE 3 (CF. TABLE IV)

	SNR = 30 dB			SNR = 15 dB		
	linear	bilinear	PNMM	linear	bilinear	PNMM
FCLS	$\nu = 0.01$	$\nu = 1$	$\nu = 1$	$\nu = 0.01$	$\nu = 1$	$\nu = 1$
ExtM	$\nu = 0.01$	$\nu = 0.01$	$\nu = 1$	$\nu = 0.01$	$\nu = 0.01$	$\nu = 0.01$
KFCLS	$\nu = 0.1$, $\sigma = 3$	$\nu = 1$, $\sigma = 1$	$\nu = 1$, $\sigma = 1$	$\nu = 0.1$, $\sigma = 2.5$	$\nu = 1$, $\sigma = 2$	$\nu = 0.1$, $\sigma = 1.5$
BilBay	-	-	-	-	-	-
K-Hype (G)	$\mu = 10^{-2}$, $\sigma = 3$	$\mu = 10^{-1}$, $\sigma = 1.5$	$\mu = 10^{-2}$, $\sigma = 3$	$\mu = 1$, $\sigma = 1.5$	$\mu = 1$, $\sigma = 1$	$\mu = 1$, $\sigma = 1.5$
K-Hype (P)	$\mu = 10^{-2}$	$\mu = 10^{-1}$	$\mu = 5 \cdot 10^{-3}$	$\mu = 1$	$\mu = 1$	$\mu = 1$
SK-Hype (G)	$\mu = 10^{-1}$, $\sigma = 3$	$\mu = 10^{-1}$, $\sigma = 2.5$	$\mu = 10^{-1}$, $\sigma = 3$	$\mu = 1$, $\sigma = 1.5$	$\mu = 1$, $\sigma = 1.5$	$\mu = 1$, $\sigma = 1$
SK-Hype (P)	$\mu = 10^{-1}$	$\mu = 10^{-1}$	$\mu = 10^{-2}$	$\mu = 1$	$\mu = 1$	$\mu = 1$